The Treble With AI-Generated Music

Markus Kunej

Ancient Chinese philosopher Confucious once said, "music produces a kind of pleasure which human nature cannot do without". He was born in 551 BCE, however, the origins of music dates much further back. The first musical instrument, a flute made of bird bone and mammoth ivory, is over 42,000 years old [1]. Since then, music has been a part of every known society - common to all human cultures around the world. The process of creating music is seen as a "creative" task; one without a strict guideline. The best examples are often described to "convey human emotions". Today, generative adversarial networks (GAN's) are able to successfully produce other forms of art, including a printed portrait which sold for more than $430,000 [2]. Although, they haven't been able to achieve the same success in producing music. Hang Chu, a computer science PhD student at the University of Toronto who created a relatively "creepy" sounding Christmas carol using AI in 2016, stated that "composing good music is actually more complicated than we expected". Unlike images, which are often more or less the same, such as close-ups on faces, songs vary widely from each other, from melody and harmony to tempo and timing [3]. This made the training process very difficult for a GAN. At least until 2020, when research group OpenAI released their music generative system, "Jukebox". It was a breakthrough in AI music creation, as it was able to compose full-length songs, with singing, in the raw audio domain [4]. Although an impressive technical feat, the use of previous artists' work during the training of such generative music models have created a "grey area" in current copyright law [5], and require a revision.

Ethical Frameworks

To help explore the ethical implications of music generative systems such as "Jukebox", three different ethical frameworks, Consequentialist, Duty, and Virtue, will be used to provide varying viewpoints and ideologies.

The Consequentialist Framework, based on an Utilitarian approach, says to select the action that will produce the most good for people. It directs attention to the future effects of possible actions, and not necessarily the action itself. One disadvantage with this approach is that it does not specify certain actions are always wrong. Even the most evil action could result in a good outcome for some people, and therefore could be classified as ethical under this framework [6].

The Duty Framework, based on non-consequentialist theories, can be seen as the opposite to the consequentialist framework. Instead of studying the outcomes of possible actions, this framework focuses on the actions themselves. It aims to "do the right thing", by focusing on the duties and obligations one has in a given situation. It is inherently more fair than the consequentialist approach, as it creates a system of rules that have consistent expectations of all people. But, this also means that one can act ethically, even if it is known that the result will cause harm. It can also be unclear in situations with conflicting duties [6].

Lastly, a more unique ethical framework based on agent-centered theories is the Virtue Framework. It states that ethical conduct is whatever a fully virtuous person would do in the circumstances. It does this through attempting to discern character traits that might motivate

those involved in the situation, with an aim to develop one's character. However, the Virtue Framework can reinforce current cultural norms as a standard of ethical behavior, and there is often more disagreement about virtuous traits than ethical actions [6].

Generative Adversarial Networks

"Jukebox", like other music generative systems, rely on the power of GAN's to generate new examples of music. The GAN model architecture involves two sub-models: a generator model for creating new examples, and then a discriminator model, which classifies whether generated examples are real or fake (generated by the model) [Figure 1]. During the training process, the two models compete against each other, playing a zero-sum game. When the discriminator is correctly identifying real and fake samples, it is rewarded, whereas the generator is penalized with large updates to its parameters. Similarly, when the generator is able to consistently fool the discriminator, then the discriminator is model penalized and has its model parameters updated. The idealized outcome to a GAN is a generator which produces samples indistinguishable from real data, and a discriminator which cannot tell the difference between the real data and generated data (e.g. predicts 50% for real and fake everytime). At this point, the discriminator may be discarded, and you are left with a model which can successfully generate new, viable examples [7].
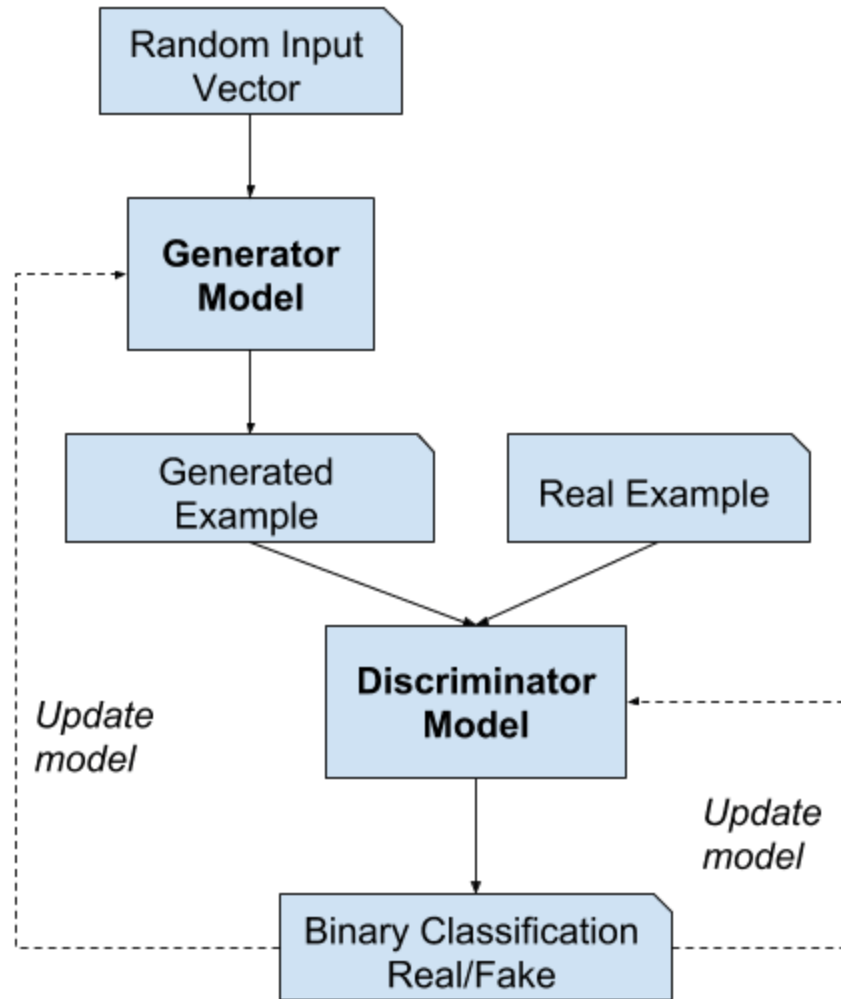
Figure 1: Generative Adversarial Network Architecture FlowChart [7]

Ethical Implications of Generative Music Systems

In terms of the music creation domain, GANs are a good choice due to their ability to model high-dimensional data (such as raw audio), and a capacity to provide multiple plausible examples for a given input, giving a user more than one song to choose from. However, like all neural network models, for a GAN to be successful it requires a very large and robust dataset to use during the training process. For example, Jukebox's training dataset consisted of 1.2 million songs [4]. A majority of these are copyrighted pieces of music and often took hundreds of hours

to produce by an artist. The ethical issue that arises from generative models such as Jukebox, is whether its output can be considered a newly created example, or rather relies too heavily on specific examples from the training dataset, and is more akin to "copying" past work. From a technical standpoint, the latter seems unlikely, due to the sheer size of training sets and the generalized structures of neurons in models. However, in Jukebox's case specifically, a user can input a specific genre of music and artist that the singing voice should replicate. This, combined with the fact that the raw audio domain is a highly dimensional domain with many intriquices, opens up the possibility that Jukebox's model is overfit to the individual artists. And so, an output would then be largely based on just one artists' past work, rather than the entire training dataset.

Consequentialist Viewpoint

Looking at this issue using the Consequentialist Framework, the best result for the most amount of people is achieved by allowing generative music systems to own the intellectual property of their outputs. The amount of new music being produced and shared with the public by these systems would increase if they had ownership of their generated songs. This is good for the common citizen, since it would mean more opportunities for them to find new music they enjoy. It could also be used as a tool by current musicians, since systems such as Jukebox can imitate their singing voice. It would allow an artist to be more efficient producing new music, making their fans (and bank accounts) happier. Some negative aspects about a flood of new AI-generated music is that it would result in a more competitive industry, especially for up-and-coming artists. It would also create plagiarism concerns, since generated music cannot be claimed under copyright by current artists, even if the song imitates their voice and style.

However, the goal of the Consequentialist Framework is to produce the most good for the most amount of people, even if it may be bad for a small group of others. And so under this framework, allowing generative music systems to own their generated outputs would be considered ethical.

Virtue Viewpoint

Alternatively, when looking at the problem under the Virtue Framework, the most ethical action would be to not allow generative music systems to own their model's outputs. With the Virtue Framework, the aim is to do whatever a virtuous person would do, and the idea of plagiarizing someone's work, especially within the music industry, is considered the opposite of that. Even if a musician is accused of being only largely influenced by a song, it is common for them to officially credit the said song, as stars Taylor Swift and Ed Sheeran have recently done [8]. The songs used during the training process of music generative systems are all meticulously put together by artists. They spend hundreds of hours writing and producing each song, and tens of years perfecting their skills as a musician. Furthermore, their livelihood depends on the income of their music. If music generative systems are allowed to own the songs they produce, they could drive many musicians out of business. Also, because systems like Jukebox can produce music mimicking current artists' voices and styles, it suggests the possibility that the model is heavily relying on the artists' past work. Although not intended, it can be considered as accidental plagiarism. As a result, the ethical decision under this framework would be to not allow the ownership of the generated songs.

Duty Viewpoint

The last ethical framework to consider this problem under, the Duty Framework, is perhaps the most fitting. This is because intellectual property disputes are most commonly dealt with through forms of copyright law. And the goal of the Duty Framework is to establish a set of consistent rules that applies to all people and then follow them, essentially the definition of laws. Before we examine the copyright laws associated with generative models, it is important to first consider the human-to-human case of plagiarism.

Currently, under the copyright law of 1976, a copyright automatically exists whenever a person writes a new song. The copyrighted elements of the song include the chord progression, lyrics, melody, and rhythm [9]. If a musician feels their copyright was infringed, they would have to prove two things: the accused had access to their composition, and the two songs have substantial similarity. Another important note is that the accused can be labelled guilty even if there was no intention to infringe the copyright, as long as the two elements above are present. [9]. For instance, George Harrison's 1970 lead single "My sweet lord" was accused of copyright infringement of the 1963 hit "He's So Fine" by the Chiffons. The judge ruled the songs were "virtually identical", however, found that the plagiarism was "subconsciously done", rather than deliberate [10]. Another example of an influential music copyright case was between Robin Thicke and Pharrell William versus Marvin Gaye. The family of late Marvin Gaye accused Robin and Pharrell's popular 2013 single "Blurred Lines" of stealing the style and "feel" of Gaye's disco hit "Got to Give It Up" from 1977. As a result, a final settlement of $5 million was paid to the Gaye family, along with a 50% share of future royalties. More importantly, judge

Jacqueline Nguyen stated the Gaye family were able "to accomplish what no one has done before: copyright a musical style" [8].

Based on these copyright laws, one could argue that the generated music from Jukebox and other similar generative music systems can be accused of copyright infringement. First, these models clearly have access to the infringed song, in the form of a raw audio file in the training dataset. Secondly, it can be argued they have substantial similarity, as Jukebox is able to closely mimic the vocal sound and style of any artist. Since it checks the two elements required for copyright infringement, and the requirements for a "similar sound" have become more relaxed in recent years, thanks to the infringement cases such as "Blurred Lines", it could easily fall under copyright law.

Though, the way a human is able to compose music is quite different to the way a generative model operates, especially in the way they "access" and interpret previous songs. Because of the differences between how humans and machines operate, laws related to Machine Learning were put into place in the European Union. Specifically, allowing the unauthorized reproduction of copyrighted works for the purpose of non-commercial Text and Data Mining (TDM). In this case, TDM is a general term covering various methods of computational analysis of information, including Machine Learning and AI. The TDM exception allows everyone to mine content which they already have access to, with research institutions (such as OpenAI) having the unlimited right to mine copyrighted data. A key rule with this law is it only applies in the non-profit case, so with it in place, current music generative systems could not commercially release the music. In the US, access to training copyrighted data is more relaxed. Unlike the EU,

US law doesn't include any specific exceptions that cover Machine Learning or AI. Instead, the Fair Use doctrine is used and has proven favourable towards technological uses of copyrighted works [11]. Fair Use is a legal doctrine that promotes freedom of expression by allowing the unlicensed use of copyright-protected works in several circumstances. Some examples include criticism, comment, teaching, and research [12]. For instance, when they publicly released Jukebox, researchers at Open AI claimed that the training of their model using copyrighted songs was under Fair Use.

So, while the use of copyrighted materials in model training is mostly ethical in accordance with current laws, what about the produced output in a generative model? This question was put to the test in April 2020, when Roc Nation LLC, rapper Jay-Z's full-service entertainment agency, filed copyright strikes against deepfaked YouTube videos which synthesized his voice. The videos were created by feeding Google's open source Tacotron 2 text-to-speech model with Jay-Z songs and lyrics, then having it read passages such as William Shakespeare's "To Be or Not To Be" soliloquy. Roc Nation stated the content "unlawfully uses an AI to impersonate our client's voice", and as a result had YouTube remove the videos, before later reinstating them, citing the DMCA takedown requests as "incomplete" [13]. This is because copyright laws pertaining to AI produced content virtually do not exist. Some US states have placed bans on specific uses of computer-generated images and video, such as in politics or revenge porn [13], but this doesn't include copyright related issues. Also, is there a difference between an AI mimicking a voice, like Tacotron 2 or Jukebox do, and a human doing an impression of a celebrity's voice? One can argue that an AI can produce a much more accurate impression than a human, but the way it achieves that goal is much more complicated than

simply copying and pasting an audio clip. Going forward, the lack of copyright laws pertaining to AI-generated content will cause many issues for artists, such as Jay-Z, as generative music systems continue to improve.

Suggested Course of Action

As part of the ethical Duty Framework, A set of rules should then be established to allow for future copyright decisions regarding AI-generated music to be made fairly and consistently. What should these rules be? Personally, I believe generative music system users should have full ownership of their model's output. However, it would also be subject to copyright claims, just like any humanly-composed song would. To account for generative models, such as Jukebox, which can purposely mimic an artist's singing, a song's "voice" should be added to the list of automatically copyrighted elements when it is produced. I believe this is the best and most ethical compromise, since it would allow for a new era of AI-generated music to be enjoyed by music lovers, while also protecting musicians from having their voice stolen. It would also allow current musicians to incorporate music generative systems as a production tool, potentially improving their own music.

Even as technology evolves around us, music continues to play an integral role in all of society. The recent success of music generative systems, such as Jukebox, hint at a future full of AI-generated music content. However, until the updates to current copyright law proposed above are put into place, it is impossible to make an ethical decision regarding the ownership of AI-generated music.

References

[1]   J. N. Wilford, "Flute's Revised Age Dates the Sound of Music Earlier," The New York Times, 29-May-2012. [Online]. Available: https://www.nytimes.com/2012/05/29/science/oldest-musical-instruments-are-even-older-than-first-thought.html?_r=0. [Accessed: 15-Mar-2021].

[2]   T. Greene, "Someone paid $432K for art generated by an open-source neural network," The Next Web, 10-Dec-2018. [Online]. Available: https://thenextweb.com/artificial-intelligence/2018/10/25/someone-paid-432k-for-art-generated-by-an-open-source-neural-network/. [Accessed: 15-Mar-2021].

[3]   K. Rogers, "Why Is AI-Generated Music Still so Bad?," VICE, 18-Dec-2020. [Online]. Available: https://www.vice.com/en/article/qvq54v/why-is-ai-generated-music-still-so-bad. [Accessed: 15-Mar-2021].

[4]   P. Dhariwal, H. Jun, C. Payne, J. W. Kim, A. Radford, and I. Sutskever, "Jukebox: A Generative Model for Music," arXiv.org, 30-Apr-2020. [Online]. Available: https://arxiv.org/abs/2005.00341. [Accessed: 07-Feb-2021].

[5]   B. L. T. Sturm, M. Iglesias, O. Ben-Tal, M. Miron, and E. Gómez, "Artificial Intelligence and Music: Open Questions of Copyright Law and Engineering Praxis," MDPI, 06-Sep-2019. [Online]. Available: https://www.mdpi.com/2076-0752/8/3/115#cite. [Accessed: 07-Feb-2021].

[6]   S. Bonde and P. Firenze, "Brown University," A Framework for Making Ethical Decisions | Science and Technology Studies, 2013. [Online]. Available: https://www.brown.edu/academics/science-and-technology-studies/framework-making-ethical-decisions. [Accessed: 15-Mar-2021].

[7]   J. Brownlee, "A Gentle Introduction to Generative Adversarial Networks (GANs)," Machine Learning Mastery, 19-Jul-2019. [Online]. Available: https://machinelearningmastery.com/what-are-generative-adversarial-networks-gans/. [Accessed: 16-Mar-2021].

[8]   E. Mullan, "Nine most notorious copyright cases in music history," BBC Culture, 05-Jun-2019. [Online]. Available: https://www.bbc.com/culture/article/20190605-nine-most-notorious-copyright-cases-in-music-history. [Accessed: 16-Mar-2021].

[9]   A. Sluzas, "Copyright Laws for Music: What Constitutes Music Plagiarism?," Philadelphia's Most Accomplished IP Law Firm, 25-Jun-2018. [Online]. Available: https://www.paulandpaul.com/copyright-laws-for-music-what-constitutes-music-plagiarism. [Accessed: 16-Mar-2021].

[10]   F. Mastropolo, "40 Years Ago: George Harrison Found Guilty of 'My Sweet Lord' Plagiarism," Ultimate Classic Rock, 31-Aug-2016. [Online]. Available:

https://ultimateclassicrock.com/george-harrison-my-sweet-lord-plagiarism/. [Accessed: 17-Mar-2021].

[11]   "EU/US Copyright Law and Implications on ML Training Data," Valohai MLOps Platform -
Train, Evaluate, Deploy, Repeat. [Online]. Available:
https://valohai.com/blog/copyright-laws-and-machine-learning/. [Accessed: 08-Feb-2021]

[12]   "More Information on Fair Use," More Information on Fair Use | U.S. Copyright Office. [Online]. Available: https://www.copyright.gov/fair-use/more-info.html. [Accessed: 17-Mar-2021].

[13]   N. Statt, "Jay Z tries to use copyright strikes to remove deepfaked audio of himself from YouTube," The Verge, 28-Apr-2020. [Online]. Available:
https://www.theverge.com/2020/4/28/21240488/jay-z-deepfakes-roc-nation-youtube-removed-ai-copyright-impersonation. [Accessed: 07-Feb-2021].