On the Ethics of Music Generation

Jack Mao, SeungU Lyu

Introduction

Art has been an important aspect of humanity throughout history - early forms of art such as murals appeared more than 30,000 years ago. As humans developed their civilization, art has developed and diverged into many different categories such as drawing, literature, and music. It is impossible to write down all the different forms of art in this document, but they all share one characteristic - the fact that they were created through the efforts of human ingenuity.

This all changed with the emergence of Machine Learning (ML) in the generation of artworks. There have been many different approaches where people used computers as a tool to create art such as "Digital Arts". However, that was a level where people were assisted by the computational power of the computer, and the computer was not the decision-maker during the process. The author would use the tool with specific intent, and will at least have some idea how their end product would look like. Machine Learning is different in that the creator builds the algorithm that will lead to creating new art, but the important decisions are made by the machine itself. Now we can say that the tool does not assist you, but you assist the tool in the creation process.

The creation of art at professional levels has been limited to people with years of training. Now with the use of ML technology, we are given the opportunity to expand art into something more innovative with relative ease. Even if one knows nothing about drawing, one can create a competent piece of drawing in less than several minutes if one finds the right algorithm to use. However, this invokes a moral dilemma: can we say works created with ML actually are original? Of course, the art is something that has never existed before and will look different from any preexisting art piece. However, we know that Machine Learning is done by training a dataset with an algorithm and recreating the pattern in a different format. One such example that clearly illustrates this dilemma is shown in the following picture.

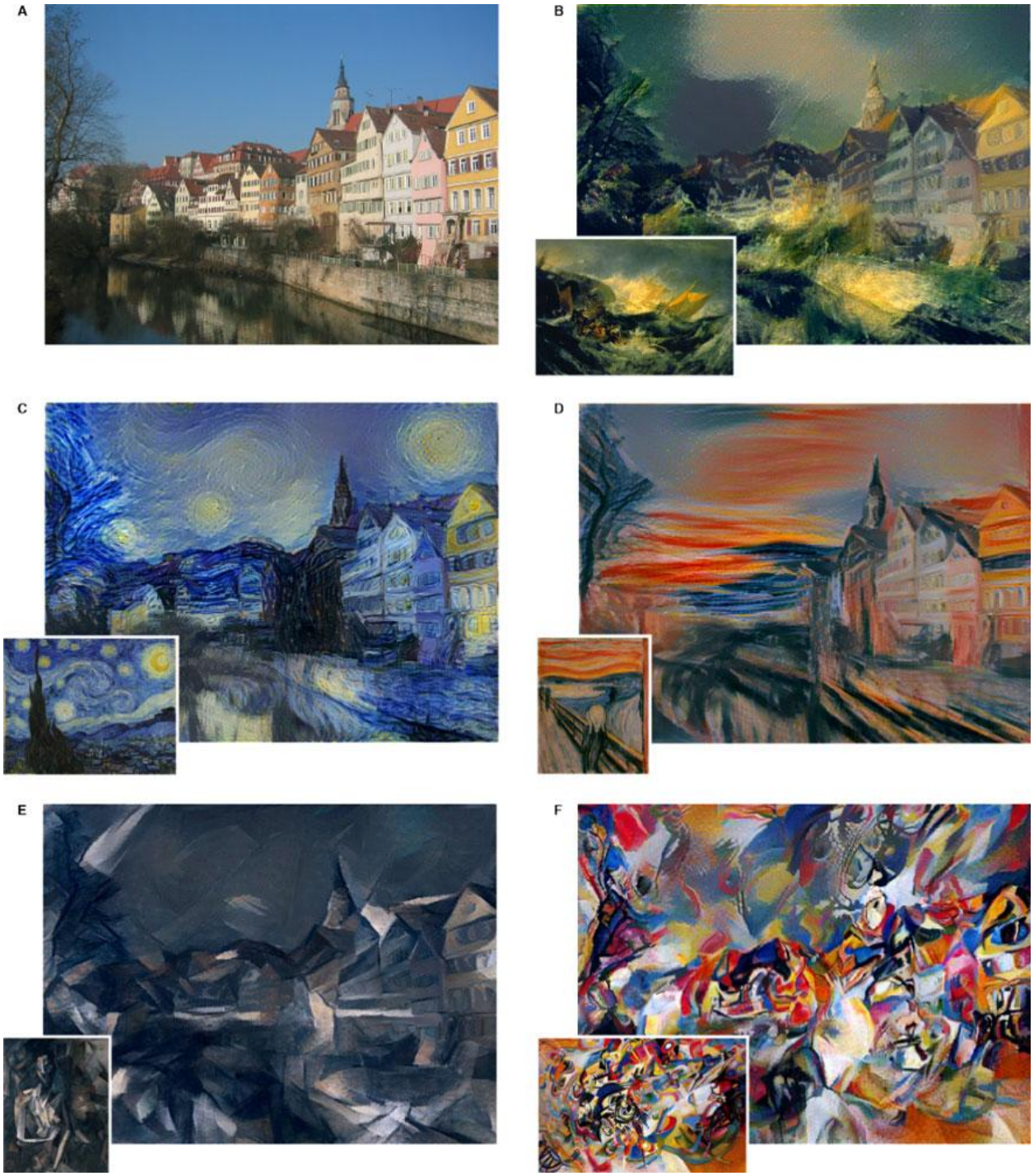

Figure '*Gatys, L. A., Ecker, A. S., & Bethge, M. (2015). A Neural Algorithm of Artistic Style. arXiv [cs.CV]. Opgehaald van http://arxiv.org/abs/1508.06576*'

In this picture, we see art generated from famous paintings. For example, model C used Van Gogh's famous "The Starry Night" to create an image that resembles his drawing style. For ML algorithms to create a new piece of art, they often have to utilize existing pieces of art. So when

an ML algorithm creates a model, there are three different groups involved in this creation: the user who used the model to create the art, the developer who made the algorithm/model, and the artists who created the original art that was used for training. All the different groups have claims to make when we think about where the image belongs. The following charts reflect the claims of ownership for each distinct group.

User	Since they are the ones who designed which dataset to use and what form of art will be created by the model, the combination is unique to the user and thus art belongs to them.
Developer	The way the model is built is designed by the developer, and an essential aspect that creates the art is embedded inside the model and decides how the final product will look like.
Artists	The art style is original to the artists and cannot easily be created by other people or by machine. Without the training data, the model can not perform anything and thus artists are an important part of the process.

The danger of machine-generated art is that the rights stated above might not be considered. It is hard to know how the art was created since the production process is done under the hood through a series of complex mathematical calculations. As machine-generated art becomes more and more adopted, it can have unexpected effects on the art community. If the process of music generation becomes streamlined and simple, there might be a probability that we will see a decrease in human originality and creativity. Instead of spending thousands of hours learning the craft, people may choose to use ML models instead. This development could make musicianship becoming a more exclusive and unprofitable business than it already is, potentially deterring many would-be musicians from ever taking a step into the field.

In this paper, we explore the various aspects of music generation. This includes the application of machine learning in music generation, a brief overview of the theory behind music generation, an explanation of our implementation of music generation, and an in-depth analysis of the ethics revolving around music generation.

Application of Machine Learning in Music Generation

Machine Learning in music generation or AI Music (AIM) is one of the biggest fields that are close to being used for commercial purposes. Some large companies are actively moving to

commercialize this technology due to the high profit they expect from the market. Some examples are listed in the chart below.

Company	Relation to AI Music
ByteDance	The parent company of TikTok, ByteDance, has acquired Jukedeck in 2019. Before its acquisition, Jukedeck was an online music generation site that offered uniquely generated music at low prices, marketed for Youtube content creators looking for appropriate background music.
Warner Music	Warner Music acquired Endel, an application that tries to create a personalized sound space based on your personal data, environment, and mood.
Tencent	Tencent partnered with Amper, a company that creates music based on the selection of mood and style.
Microsoft	Microsoft organized a research team called Muzic in 2019, which has already published 9 papers.

As more research is conducted on AI music, the difference between man-made and machine-generated music becomes almost negligible. In addition, this technology is more accessible than ever before, and will only become more accessible in the future. To prove the robustness of AI music, the following links refer to pieces of music created by Jukebox from OpenAI.

- Classic Pop, in the style of Frank Sinatra
- Country, in the style of Alan Jackson
- Rock, in the style of Elvis Presley
- Pop, in the style of Katy Perry
- Blues Rock, in the style of Joe Bonamassa
- Heavy Metal, in the style of Rage

It is amazing to know that Machine Learning can be used to not just create the melody, but also the lyric and the singing voice. Currently, machine-generated music can sound unnatural and awkward at times, making them unable to compete with professional artists. However, it would not be long before machine-generated music becomes robust enough to compete with professional musical artists. This is the reason why ethics is important in AI music. OpenAI used

1.2 million songs with corresponding lyrics as a dataset to create a tool that was used to generate the above songs. There's no way OpenAI was authorized to use all these songs from individual artists, so that's the reason why the tool is only for non-commercial use. However, the current copyright law does not protect artists from somebody creating a song that is based on their work. Rather, it is very usual how a "style" of an artist is influenced by another artist, as far as the work is not a direct sampling of an existing song. So even if two songs sound similar, there is no problem as long as there are no direct copies. Also, even if the song seems like a particular artist's style, it is really hard to know how much out of millions of songs contributed to the generation of the music, which makes the case more complicated. The part that can possibly have issues is the part where the computer imports the music and uses them to develop the model. Does purchasing music also give the right not only to listen but to use them as ML training data? There isn't a clear answer to this question yet.

For the next part of the project, we will be creating musical pieces with help from different tools to show and explain briefly how music generation is done with ML. From this, we want to show that the music generation has become something that is accessible to the public if they have some amount of programming knowledge. Also, we want to make a further discussion about the ethics of music generation once we have successfully generated our sample music.

Implementation Goals

The goal of our implementation is to show how easy it is for machine learning models to create a new piece of music. If we could generate a piece of music using AI that sounds sufficiently good in the time span that we have for the final project, we can sufficiently claim that AI models are accessible enough for people with some computer science background to generate music regardless of their musical knowledge. In the interest of not being hypocritical, we do not intend to use our created scripts for any commercial use. Our scripts are made for **research purposes only** and should not be used for profit.

The theory behind Music Generation

Automatic music generation is not a new concept. In fact, it has already been explored half a century ago. One prominent way of making music is symbolic music generation. A list of papers that use this approach can be found here. When explained in the simplest terms, symbolic

music generation represents music in the form of a piano roll, which specifies the pitch, velocity, and instrument of each note. This approach has produced many results such as the Bach chorales, polyphonic music (music with independent melodies played at the same time), and minute-long compositions.

Although symbolic music generators have achieved incredible feats, there are marked limitations. For instance, they cannot capture human voices and have trouble imitating the subtle timbres, dynamics, and expressivity that are present in most music. Thus, a better way of representing data is as raw audio. Raw audio is encoded using convolutional neural networks (CNNs). This produces compressed audio, which will be passed into a trained transformer that creates a novel compressed audio (this is where the magic happens that makes the ‘new‘ music). The final product is upsampled using transformers and decoded using CNNs.

Due to the time span of our project and the lack of computational power, we do not have nearly enough time to create our own machine learning model. Instead, we used Jukebox, a neural net that has already been trained on 1.2 million songs and their corresponding lyrics. Jukebox generates music at the audio level, which makes it sound professional. However, an important drawback of generating music at the audio level is that the sequences are very long. According to OpenAI, a 4-minute song at CD quality (44 kHz, 16 bit) has over 10 million timesteps. To give a perspective of how large this is, GPT-2 (a natural language processing model that was also developed by OpenAI) had 1000 timesteps.

One approach that OpenAI used in JukeBox is the autoencoder, which compresses audio files to a lower-dimensional space by discarding perceptually irrelevant information. The audio is trained in the compressed space and then upsampled back to the raw audio space. The autoencoder model used in JukeBox is called VQ-VAE (Vector-Quantized Variational Autoencoders). Autoencoder is an unsupervised learning technique that uses neural networks to find latent spaces (hidden learning patterns in the data). Variational autoencoders (VAE) differ from traditional autoencoders in that VAE’s put semantically similar data points closely and semantically dissimilar points far apart. Finally, the difference between VAE and VQ-VAE is that VAE learns a continuous latent representation, while VQ-VAE learns in a discrete way. The reason Jukebox chooses to use a VQ-VAE is because most real-world phenomena are better represented as discrete patterns. Music, in this aspect, fits perfectly in this category.

Jukebox is not omnipotent. Although generated songs show musical coherence in small segments, it fails to take into consideration the overall structure of the song. For instance, Jukebox cannot create a chorus (repeated lyrics). Another problem is that during the downsampling and upsampling process, additional noise is introduced to our final product. Finally, the performance speed of the Jukebox is slow. It takes about 9 hours to fully render 1 minute of audio through the model with the best GPUs (P100 or V100 which are often provided by Colab) and will take much longer with less computational power, making it an impractical tool to generate large amounts of music.

Our Implementation of Music Generation

We used the JukeBox library and created our own song on Google Colab. The lyrics are written by us, but the melody and the vocals are all generated by Jukebox. We trained our transformers in the style of Frank Sinatra. A sample of our final output can be found here with the following lyrics:

"As we finish machine learning,
we hope you are filled with wisdom.
Gradient descent accompany you everywhere you go.
Telling confounding variables at a glance.
Nothing stops you from modeling,
like the Monte Carlo simulation.
Enjoy the break and happy holidays"

Ethical Considerations

In our short jingle, we have created a song in the style of Frank Sinatra. As we listen to the music, we can hear a clear melody that is reminiscent of 1980's music. The audio is distinct and comprehensible. If not for the fact that neither of us is talented lyricists and that the music was about machine learning, it would be quite difficult for us to distinguish the product from a real song. There are no significant noises that we could definitely call unnatural. The quality is good enough such that we could say that it was a practice piece from Frank Sinatra back in the day. However, to whom or what does the right of our product go?

In the case of our generated music, there are several arguments that could be made. There are two distinctly identifiable groups that hold a stake in this music, the ML programmers, the Universal Music Group that owns the rights to Frank Sinatra's music, and the artist Frank Sinatra himself. OpenAi's Jukebox should be used only for non-commercial purposes, but let's assume that the song was made for profit during this short discussion.

On the side of ML programmers, generating music requires time and effort. During our process, the algorithm does not simply generate music with a press of a button. Instead, there are many steps involved and music is generated at 4-second intervals. Every time different samples were created, we had to choose the next sequence carefully to guide the song in the direction we wanted. Most samples had unrecognizable lyrics, which means that without our consideration the music would not sound as it is now, or even sound like complete noise. The lyrics are also created by ourselves. The only process that has been automated was the melody and the singing. We can argue that just as many artists take inspiration from each other, our model is heavily inspired by the music of Frank Sinatra. One interesting case that supports our claim is this video called the Pachelbel Rant. In this video, we see the same chords from Pachelbel's Canon in D (1680) omnipresent in many of the songs we hear today. However, this does not mean that everyone is stealing from music that is over 300 years old.

On the side of Universal Music Group, which is the company that owns the rights to Frank Sinatra Enterprises. Clearly, there is a profit problem that the Universal Music Group should care about. If we were to release our music for profit, then we would be using music without the consent of the people who own the music. If we can continue producing music in the style of Frank Sinatra and license it for a much cheaper price, then consumers will see little to no reason to spend money to acquire the right to use Frank Sinatra's original pieces.

Finally, on a more personal level, we have the artist himself. Because ML engineers do not have any formal agreement with the artist, there is a layer of detachment when using the work of other people. In our case, we wrote a song that is quite silly and wholesome. Even though we wrote the song with no evil intentions, we do not know what Frank Sinatra would feel about us using his voice to sing this silly song. We will never know what Frank Sinatra feels about this since he is dead, but his family members might object to us creating music using his voice.

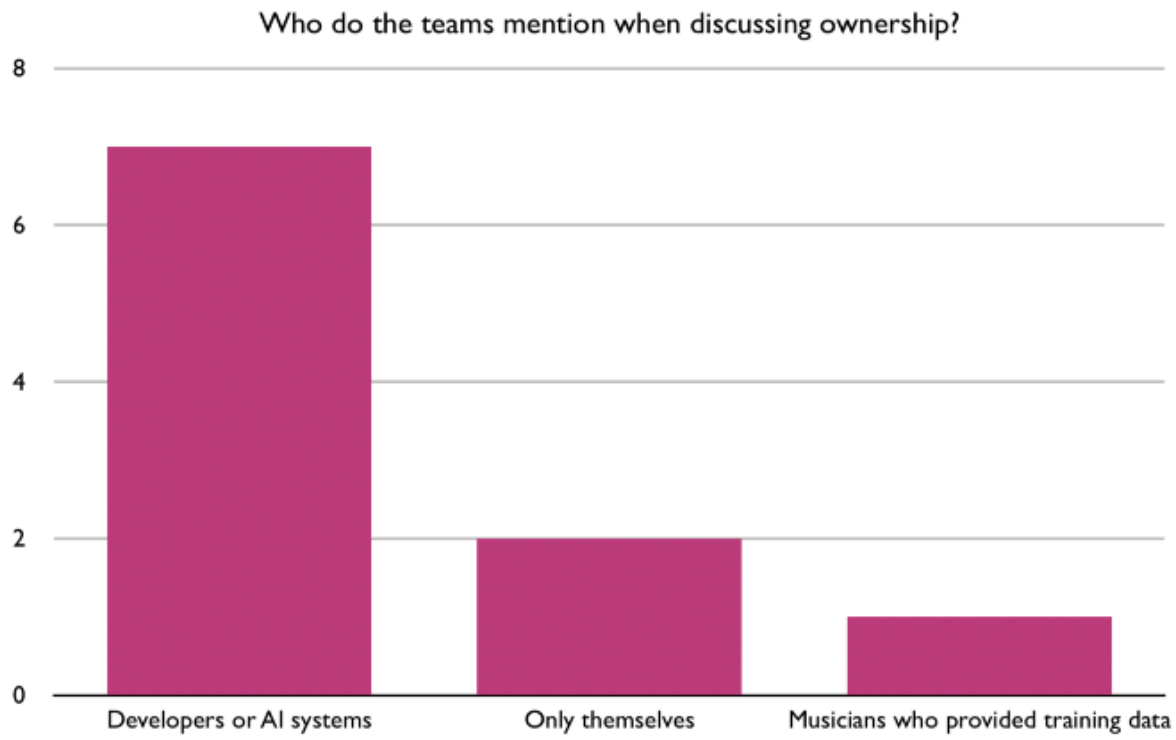

Figure from “The Ethics of the AI Song Contest” by Ed Newton-Rex and Hendrik Vincent Koops, https://medium.com/creative-ai/the-ethics-of-the-ai-song-contest-184840c3bbcc

As machine-generated music becomes more robust, greater consideration needs to be placed on the ethics surrounding generating the music. In 2020, the first international AI song contest was organized by Dutch broadcaster VPRO. The song contest focused on the collaboration between humans and robots. An integral part of the competition revolved around the ethical considerations in ownership and authorship of the song relative to the AI. Most teams only considered themselves and the AI they used had ownership over the music. However, barely anyone mentioned the musicians who provided the dataset. In the case of this competition, the musicians’ ownership is greatly under-represented.

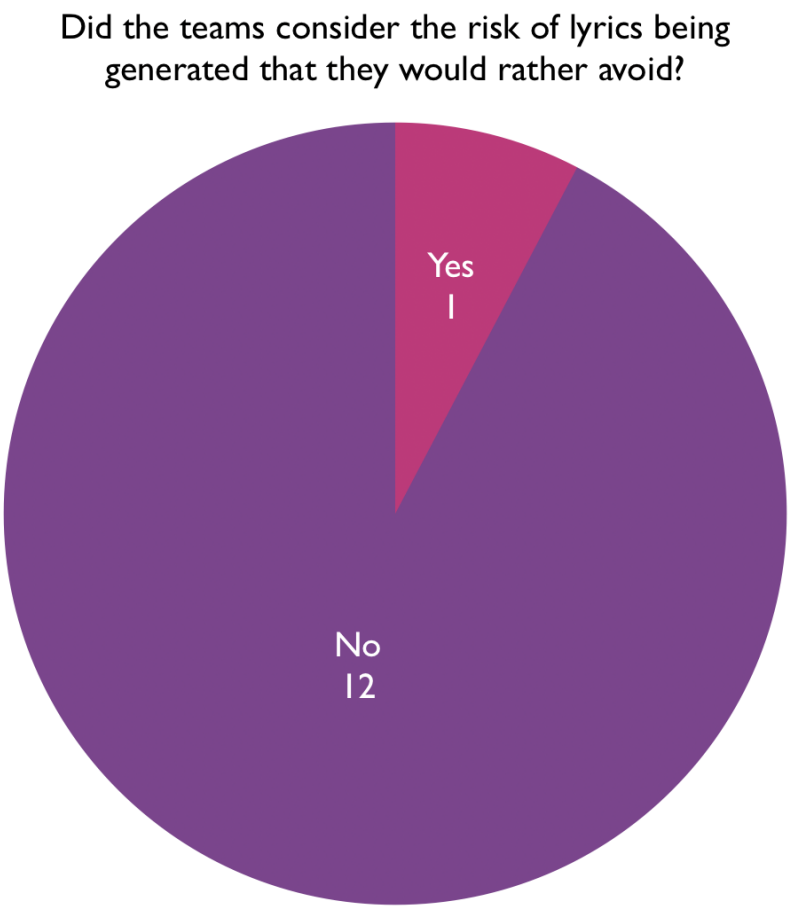

Figure from “The Ethics of the AI Song Contest” by Ed Newton-Rex and Hendrik Vincent Koops, https://medium.com/creative-ai/the-ethics-of-the-ai-song-contest-184840c3bbcc

Another important aspect that came up during this competition was the problem with the need to filter the output. During the competition, one of the generated songs included a phrase saying “kill the government.” If we scrape music indiscriminately from the web to create a dataset, we are bound to find songs that have controversial phrases that make sense in the context. For example, the song “Pumped Up Kicks” by Foster the People talks about dealing with mental health problems. However, the model will not understand the real-world context and accidentally sing about school shootings not in the way it was used inside the original song. Therefore, it would be important for music generators to check their output for racism, gender bias, etc.

In the context of commercial AI music (AIM), innovations in AI have been criticized heavily for their potential threat of human redundancy. Fabio Morreale explains in his paper that we need to take into account the socio-economic context that surrounds the music industry. Currently, there

is already a human redundancy problem in the music industry. The sales and popularity of music are concentrated on the elite few, making the possibility of AIM becoming popular exceedingly low.

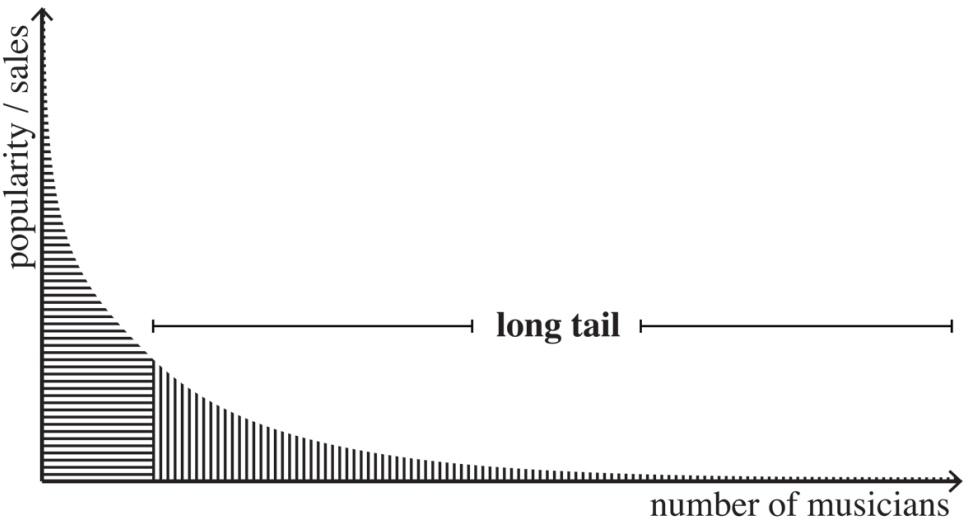

Figure from “Where Does the Buck Stop? Ethical and Political Issues with AI in Music Creation” by Fabio Morreale, https://transactions.ismir.net/articles/10.5334/tismir.86/

The length of the long tail will only get longer, as we are living in an age of attention economy. In 2017, Spotify included fake artists in their curated playlists, taking up attention slots and royalty revenues from real musicians. In Spotify’s ongoing efforts to reduce artists’ royalties, AIM becomes a particularly effective solution. AIM is low-cost and cannot organize unions to protest against unfair treatment. If unchallenged, the innocent aim of creating AI that can make music will inevitably increase competition, making the lives of struggling artists who are trying to establish themselves even harder.

As if to add insult to injury, commercial AI will most likely use large amounts of data from the trailing end of the artists to create a large enough dataset. AIM is increasing competition with other musicians while taking music from its competitors to make itself a better AI. Many of these works are likely unlicensed nor acknowledged, but there is no way to check. Therefore, artists are exploited for their innovation and creativity. In a way, this is digital colonialism, whereby human efforts are exploited as data for commercial gain.

Despite all the potential issues that revolve around AIM, there are some use cases that are certainly positive. For instance, a renowned South Korean singer Turtleman died in 2008 due to a heart attack. To honor his music, his family requested an AI rendered song to be made. This was a heartwarming and wholesome event that shows that with consent and approval, we could revive musical geniuses through technology. Instead of exploiting AIM for profit, there could be potential for humanity to use this technology for good.

Conclusion

Clearly, ethics need to be considered when conducting AIM research. Therefore, everyone involved in this process needs to step up. If unrestrained, AIM could irrevocably change the landscape of the musical communities. We can think of three big directions we see that will make AI-generated music more ethical and create positive social impacts.

First of all, AI researchers need to hold their sponsors or companies responsible for the ethical use of data. The biggest concern with engineers in the field is that they lack awareness of ethics and the social impact of their actions. At Olin, we are often driven by passion and curiosity, without knowing the details about the stakeholders or possible harm we can cause. Even for us, ethics has been a topic of interest only considered after we were prompted with this idea. Without being exposed to ethics in machine learning, we could have been one of those people who exploit work from artists without concern! Growing the awareness can be done by educating the engineers and providing an environment where ethical considerations and social impact are valued. It will not be easy to change the mindset of the current community of engineers, but we believe that education on ethics will change the mindset of future generations to come.

Companies need to be regulated from unfair business practices that intentionally take attention away from musicians of the trailing end of the spectrum. If regulations are hard due to political pushbacks, at least a guideline is needed to make the situation better. Even though people warned about the possibilities of the current situation with AIM decades ago, the regulation for use of musicians' work for the purpose of ML is still vague. This provides a lot of holes that will lead to the exploitation of works without consent by companies. On the companies' side, it is often inevitable that the company should aim for the maximum profit. Even if they know what they are doing is unethical, their action might not stop as far as what they are doing is profitable.

There should be a decision made to clarify this vague situation with copyright and ownership before the AIM industry can move on to the next step without harming anyone.

Musicians need to also be educated to be aware that their work may be exploited for unfair advantages. One direction we anticipate is where both AIM and musicians can have positive synergy with each other. AIM can greatly benefit musicians if they are used to support the creation process. For the entire paper, we've been worrying about the possibilities that AIM outcompetes the professional music industry, but we were neglecting the fact that it can aid musicians in innovative ways. Also, if AIM provides ways that musicians are compensated fairly for the use of their work on the creation of AIM, this can open a totally new revenue for musicians. This will greatly depend on what stances these two parties will take in the future application.

Music is a medium that evolves with human ingenuity. It would be a shame if human musicians were ousted by the fruits of their own labor. If we want to keep music as exciting, unpredictable, and diverse as it is today, steps need to be made so that AIM does not become a disruptive force in this community. We believe that if every stakeholder plays a part in this process, then we might be able to step into a future where AIM can be integrated into the world of music in an equitable way.

Resources

Gatys, L. A., Ecker, A. S., & Bethge, M. (2015). *A Neural Algorithm of Artistic Style*. *CoRR, abs/1508.06576*. Opgehaal van http://arxiv.org/abs/1508.06576

OpenAI. (2020, April 30). *Jukebox*. OpenAI. Retrieved from https://openai.com/blog/jukebox/

OpenAI. (n.d.). *Code for the paper "Jukebox: A generative model for music"*. GitHub. Retrieved from https://github.com/openai/jukebox

OpenAI. (n.d.). *Interacting with Jukebox*. Google Colab. Retrieved from https://colab.research.google.com/github/openai/jukebox/blob/master/jukebox/Interacting_with_Jukebox.ipynb#scrollTo=eNwKyqYraTR9

Snell, C. (2021, February 9). *Understanding VQ-vae (dall-e explained pt. 1)*. ML@B Blog. Retrieved from https://ml.berkeley.edu/blog/posts/vq-vae/

Newton-Rex, E., & Koops, H. V. (2020, October 18). *The ethics of the AI song contest*. Medium. Retrieved from https://medium.com/creative-ai/the-ethics-of-the-ai-song-contest-184840c3bbcc

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), pp.105–113. DOI: http://doi.org/10.5334/tismir.86

Deahl, D. (2019, April 17). *We've been warned about AI and music for over 50 years, but no one's prepared*. The Verge. Retrieved from https://www.theverge.com/2019/4/17/18299563/ai-algorithm-music-law-copyright-human

Mackenzie, S. (2021, March 24). *Artificial Intelligence is shaping the future of music - but at what cost?* CBC Music. Retrieved from https://www.cbc.ca/music/artificial-intelligence-is-shaping-the-future-of-music-but-at-what-cost-1.5952955

Universal Music Group and Frank Sinatra Enterprises Announce Global Licensing Initiative for the legendary entertainer's reprise catalog. Frank Sinatra. (n.d.). Retrieved from http://www.sinatra.com/universal-music-group-and-frank-sinatra-enterprises-announce-global-licensing-initiative-for-the-legendary-entertainers-reprise-catalog/

Holzapfel, Andre & Sturm, Bob & Coeckelbergh, Mark. (2018). *Ethical Dimensions of Music Information Retrieval Technology. Transactions of the International Society for Music Information* Retrieval. 1. 44-55. 10.5334/tismir.13.

Mnet TV. (2020, December 19). *One More Time EP.1*. Youtube. Retrieved from https://youtu.be/Jm0s0CEEd3Q

RobPRocks. (2019, November 5). *Pachelbel Rant from 2006*. Youtube. Retrieved from https://youtu.be/uxC1fPE1QEE

Music generation. Papers With Code. (n.d.). Retrieved from https://paperswithcode.com/task/music-generation