Cultural and Economic Impacts of Music Synthesis Systems

A Research Paper submitted to the Department of Engineering and Society

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

Liam Brennan

Spring 2023

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Advisor

MC Forelle, Department of Engineering and Society

1. Introduction

Throughout history, technology has helped facilitate the ability of humans to express themselves through music. From the first flutes over 40,000 years ago to modern music creation software, technology continuously redefines and revolutionizes how people worldwide create and enjoy music. However, with the advent of highly complex and powerful generative models capable of creating human-like art and music, we are rapidly approaching an inflection point that will change the relationship between art, music, and technology. The perceptions of art and music AI systems are commonly negative, and often apocalyptic. Some news outlets are claiming that "AI could take the jobs of as many as one billion people globally and make 375 million jobs obsolete over the next decade" (Zippia, 2023) or that art AI could lead to "entire swaths of the creative workforce evaporating. 'Concept artists, character designers, backgrounds, all that stuff is gone.'" (Salkowits, 2022). The societal and economic impact these models might have in the future remains a big question mark. This work will specifically focus on the future implications of AI music synthesis models, programs that can create complete songs with any human input.

AI has exploded over the past decade, becoming a potent economic force. In 2022, the global AI industry was estimated at over $100 billion and expected to grow by almost 20x to nearly 2 Trillion by 2030 (Technologies, 2022). Most of the AI in use today is based on deep learning, an AI technique characterized by its complexity and need for immense amounts of training data (examples used to "teach" the model) and computing power. With the rise of the internet, engineers can gather and share large swaths of data to feed into these deep learning Models. Combined with the rapid increase in computing power that allows for the speedy

training of large models, engineers can create incredibly powerful AI programs. A subset of deep learning is generative deep learning, characterized by models that can create text, art, and audio from scratch. Recent examples of these models, including Jukebox (music), Dalle-2 (art), and most recently, ChatGPT (text), have brought Generative Models to the forefront of AI discussion. Unlike more traditional types of machine learning problems, like predicting the stock market or the weather, the performance of these models is purely subjective.

In this work I will show that AI music Synthesis will help supplement and grow the creative process instead of replacing it; however, it will leave many artists and enjoyers of Non-Western music behind. This paper is organized as follows. I begin by briefly summarizing relevant literature in section 2, where I will discuss some of the most recent Generative Models and their use in the music industry, current commentaries on A.I, and the Social Construction of Technology (SCOT) framework. In section 3, I discuss the primary and secondary sources used and how I apply them to SCOT. In section 4, I first show how the developers of music AI mostly cater towards Western, white music, and the current trajectory of music AI development will continue to ignore most Non-Western groups. Second, I will show that music AI will directly supplement the music-creating process through automation and inspiration but may render many jobs in the music industry obsolete. Finally, in section 5, I will conclude the paper by discussing the implications of the design of current music AI systems, directions for improvement of these systems, and suggesting future directions for research.

2. Literature Review

Music AI development has come a long way since Hiller and Isaacson's 1958 pioneering work, in which they created programs that wrote simple classical melodies and harmonies (de Mantaras et al., 2002, pg 44). Over the past few years, with the explosion of deep learning, developers of music AI systems have begun to have the capability of creating full-length songs. MuseNET by OpenAI in 2020 was a huge step forward in music AI, leveraging deep learning to develop full-length songs by combining multiple MIDI tracks, i.e., multiple sets of piano rolls with different instruments stacked on top of each other (Pal et al., 2020). In another massive step in music AI, the developers of Jukebox in 2021 created full-length songs of raw audio instead of MIDI (Dhariwhal et al., 2021). Using raw audio instead of MIDI allowed for more complex audio effects, instruments, and singing. With the rapid improvement in other generative models, such as Dalle-2 and ChatGPT, it is reasonable to assume that full-length music Generation AI will continue to improve. While these generative models are not quite at the level of human music generation, some music AI tools are already used in the music industry, such as LANDR AI, a mastering tool (LANDR, 2023).

There has already been commentary on creative AI and its impacts on the economics of creative industries. So far, creative AI has been shown to be poor at appealing to human emotions through advertising (Bakpayev et al., 2022), and it has been argued that AI systems are currently unable to grasp an understanding of human emotion and creativity (Holford, 2019). Furthermore, it has been argued that the poor performance of AI on creative tasks will lead to more jobs focusing on empathetic and emotional dimensions through the "Feeling Economy" (Huang et al., 2019). Some articles have claimed that AI will replace entire creative economy

sectors (DelSignore, 2022). However, there has been little to no academic work on the effect of music AI technologies on the music industry and its potential use across different cultures: this is the hole I will fill in this work.

In this work, I will also leverage the Social Construction of Technology (SCOT) framework, first proposed by Pinch and Bijker (Bijker et al., 1987). SCOT posits that technology is a product of societal values and actions, instead of society being shaped by technology. Advocates of SCOT, called social constructivists, argue that technology is only adopted in the proper social context. Because of this, understanding societal context is necessary to analyze technology.

3. Methods

In order to provide an analysis of the future implications of the development of music AI systems, I will research the technicalities of AI technology, Music Psychology, Non-Western music, and the music industry. I will cover various primary and secondary sources, including academic publications, technical and analytical, economic reports, and books on music psychology and diversity worldwide. I will then combine my findings with the SCOT framework to perform my analysis. Since music is an inherently subjective and social construct, SCOT will allow me to unravel the relationship between different social groups, especially Non-Western peoples and workers in the music industry, and emerging music AI technologies.

4. Analysis

In this section, I will discuss my analysis and provide my arguments. I will split this section into two subsections. First, I will discuss the difference between Non-Western and Western music and how the current music AI systems leave Non-Western music behind. For my purposes, I will define Western music as music derived from traditional Western European music theory, including American Pop, Rock, and Classical. Second, I will discuss the effect of music AI on the Western music industry: how it will affect artists, musicians, and other groups.

4.1. Western vs. Non-Western Music

In the West, we often need to recognize the differences between music worldwide. From the beginning of human history, music has been a common factor among all civilizations and groups of people. Music reflects culture and society, and as culture is unique and diverse around the globe, different peoples have different social constructions of music. Not only does music have a diverse sound, but how music is defined, used, and listened to is diverse.

From Merriam Webster, the English definition of music is "vocal, instrumental, or mechanical sounds having rhythm, melody, or harmony." Put another way; the English-speaking world defines music solely by its musicality: the musical sound it produces. This focus on musicality is a uniquely Western obsession, and the isolation of musical sound from other arts is a Western abstraction (DjeDje, 1987). As we look around the glove, music often means more than just the musical sound associated with it. In many cultures, music is commonly intertwined with dancing, religion, or other social activities, often to the point that there is no distinct term

for what the West defines as "music." Instead, music is defined based on the social activity associated with it (DjeDje, 1987). For example, in the study of Candomble music, religious music based on African heritage in Brazil, the music "is so ubiquitous that the same term encompasses what we recognize as music and religion" (Trehub, 2015, pg 2). Many Non-Western cultures often make no distinction between musical sound and dance, in styles such as Sangita from India, Wai Khruu from Thailand, and Nkwa from Nigeria (Trehub, 2015, pp. 2-3). In these cases, developers cannot make all-encompassing music AI systems, as music is about social interaction, and these systems can only synthesize musical sound.

Beyond developers of music AI systems failing to encompass the social aspect of Non-Western music, the musicality itself has been Western-focused ever since the inception of music synthesis. One of the earliest and best-known pioneering works in music AI, by Hiller and Isaacson in 1958, involved building programs that created string quartets, classical harmonies, and classical melodies (de Mantaras et al., 2002, pg 44). In 1972, Moorer worked on creating Western-style melodies with Western chord progressions (de Mantaras et al., 2002, pg 44). In 1993, one of the first attempts at using Neural Networks, the younger sibling of deep learning, by Bharucha, focused on creating classical-style harmonies and melodies in the style of Bach, one of the most well-known classical composers (de Mantaras et al., 2002, pg 45).

Diving into some of the technicalities of the most recent music AI systems, we can also see how models are designed around Western music theory. For example, MIDI is a popular method for encoding music, commonly used for models such as MuseNET. MIDI represents music as a string of notes at specific points in time. The MIDI is then played through instruments (whether virtual or physical), which produce the actual sound. A major fallback of MIDI is that it is designed around Western music principles and cannot encode many standard Non-Western

music techniques, such as sliding tones, microtones, and improvisation. In addition, MIDI does not support Non-Western notes or scales (Standard Midi File Format, 2023). Despite being 40 years old, MIDI is still the standard for music creation (Gaston, 2019). Intentional or not, the wide use of MIDI as the backbone of music AI systems hinders these model's ability to synthesize Non-Western music.

Another significant issue is the lack of Non-Western data in the training models; most developers of music AI systems train on Western music. For example, MuseNET was trained almost entirely in classical and Western music (Pal et al., 2020). An AI model reflects the data used to train it. If most of the music used to train the models is Western, then the music model will produce Western music. We can see this in Jukebox, another system trained on almost only classical and Western music. Of the over 7,000 samples released by OpenAI, none are from African or Asian styles, and less than 100 examples are from Non-white, Western genres, such as Reggae or Samba (Dhariwhal et al., 2020).

Developers building music AI systems for strictly Western music has been the standard since its inception 65 years ago. From the perspective of SCOT, this makes sense. Over the past 65 years, many consequential music AI systems developers – such as Hiller and Isaacson, Bharucha, and OpenAI – have designed their systems in Western, English-speaking countries. Consequently, these models reflect the popular music enjoyed by a large portion of these populations, broadly appealing to the user groups most likely to be affected by their work. Additionally, MIDI's position as the standard music format in the Western music industry motivates the use of MIDI by developers. Developers of modern music AI systems continue to emphasize Western music, with all indications that this trend will continue.

4.2. Impacts on the Music Industry

While music AI can replicate the musicality of human music, there will always be a need for artists due to the social aspects of music. As mentioned previously, music's consumption, use, and importance worldwide is incredibly diverse. Developers of music AI systems may be able to build systems that can create impressive musical compositions, but they cannot build systems that can recreate the music's context. In many situations, the music's context is much more critical than the musicality itself. In these cases, the need for music AI systems comes into question.

One good example of this is religious music worldwide. As religion is often tied to the human condition, AI-generated religious music defeats the purpose. This is more than just conjecture - by studying the history of religious music around the globe, we can see a pattern of religious music being slow to adopt new music technology or being against adopting new technology at all. Some extreme examples of this can be seen in some Islamic religious texts. In the Sama, a rulebook for music historically used in some parts of Islam, secular music is forbidden. "The concept of sami'is contrasted with ghini' (cantu s) which by extension designates music, music making and performance, associated mainly with secular art music; this is normally banished by most of the authors who deal with the sami" (Shiloah, 1997, pg 1). The lack of technology used defines acceptable music; "...Koran cantillation, the singing of unaccompanied hymns, old bedouin songs and the simple functional folk tunes marking events in the life of individual and community in the category of permissible forms of sami" (Shiloah, 1997, pg 1). The main idea behind this text is that secular forms of music and unnecessary musicality distract from the primary religious goal of the music.

Another excellent example of context trumping musicality, and an example more directly related to the music industry, is live performances of recorded artists. Objectively speaking, music from a live performance can be considered inferior to the recorded version, as the recorded version is a product of the time and energy poured into recording and production. Live music is unproduced, raw, and imperfect. However, the social aspect of live performances makes them appetizing to so many people. The live performance industry is not just huge but growing every year. Taken from an economic report in 2019, the live performance industry currently supports 913,000 jobs, with a nationwide impact of 132 Billion dollars; in total, the industry directly generated 17.5 Billion dollars in 2019 (Sacks, 2019, pp. 5-6). Other than thematically showing how AI cannot replace the emotional and social aspects of music, the live performance industry will continue to demand a wide variety of jobs; instrumentalists to play the instruments, managers to manage the performances, and workers that maintain the venues.

One of the more apocalyptic perceptions of music AI is that music synthesis will replace songwriters or other people in the industry producing the music (Zippia, 2023; Salkowits, 2022). If music AI synthesis could theoretically produce full-length songs at professional quality, what is the purpose of humans in the music creation process? Instead of replacing humans in music creation, music AI will supplement artists. While music may provide inspiration for, or speed up, the creation of musical sound, it will only partially provide a replacement for the music-creation process.

In a recent overview of generative technologies in art industries, Anastrasiricha claims "AI tools will not be used in isolation as a simple black box solution. Instead, they must be designed as part of the associated workflow and incorporate a feedback framework with the human in the loop" (Anantrasiricha, 2021, pp 48-49). In other domains, such as visual art

creation, we can already see many artists leveraging AI tools, and artists such as Patrick Clair predict that “rather than replacing concept artists or putting Hollywood special effects wizards out of a job, AI image generators would simply become part of every filmmaker's tool kit" (Roose, 2022). While AI generative tools may be able to create impressive images in isolation and lower the barrier to creation, there must be a marriage between the technology and the artist. In stark contrast to the black-box view of art generation, leveraging generative models intelligently is already proving to be an important artistic skill, exemplified by the growing AI Art NFT business and burgeoning AI art competitions (Kelly, 2022).

Like AI-generated visual art, I expect AI-generated music to supplement artists in a similar way. Currently, the demand for musical artists is high and growing every year (MusicBrainz). From a social constructionist’s view, it is difficult to imagine the widespread adoption of technologies aimed at making musical artists obsolete. As musical artists adopt more music AI tools, developers will likely begin catering their tools more towards these artists, creating more possibilities for artistic expression. Indeed, we can already see music AI companies that market themselves as ways to supplement the creative process, such as Beatoven.ai and Orb Producer, begin to emerge (Beatoven AI; Ditto Music).

While music AI will supplement artists instead of replacing them, some groups of people in the industry will be hit. The demand for instrumentalists - those who make a living playing instruments - will lower as developers of music AI can already produce high-quality MIDI instrumentals, as shown by MuseNET. Developers will even be able to recreate instrument sounds, already found in systems such as Jukebox. I expect some jobs involved in the production and audio engineering to be lost, as we already see AI tools such as Magenta that help fastrack mixing and mastering of songs (Ditto Music). Additionally, some forms of mass-produced music

– such as soundtracks for ads and video games – have already entered the music creation scene (Ditto Music). As this trend continues, there will be less of a need for artists that work on creating these types of music.

5. Conclusion

The development of music AI is exciting. By speeding up the creative process and providing inspiration for new sounds, music AI will supplement artists and free them to create more complex works. Much like how basic instruments, electric guitars, and music software have shaped musical sound with time, music AI will allow artists to create art in new and imaginative ways. By removing some of the barriers to music creation, such as instrumentalism, more people can express themselves through music. However, with the importance of authenticity and context, music AI, even given the ability to recreate full songs at a human level, will not replace artists.

Unfortunately, this change in the musical landscape is poised to primarily benefit Western music. Music AI synthesis models are built for a Western audience, using almost exclusively Western music for training. Additionally, these models focus exclusively on musicality, ignoring the social element of music in many Non-Western cultures. The music industry in Non-Western parts of the world is rapidly growing (IFPI), and as the user base of some of these models may expand towards Non-Western audiences, many aspects of these models will be inaccessible.

Analyzing the potential implications of current music AI development is essential. As AI inevitably starts to create and produce music at a human level, the relationship between people, music, and technology will change. Focusing on how that relationship will change, especially who will benefit and be left out, is a critical discussion.

Engineers, data scientists, and other people working on these models need to keep the implications of their work in mind. From the perspective of SCOT, developers will first need to change the designs of their models to cater more towards a Non-Western audience, not just using Non-Western training data but also building the infrastructure of these models to be more accessible. Furthermore, the demand for supplement AI tools, instead of generative tools, will increase as artists begin adopting AI into their creative process.

This work is also essential for people working directly in the music industry. As the development of music AI is poised to alter the industry landscape, most people working in the industry will be affected somehow. Artists must look into how their creative process may change and how they can leverage emerging AI technology to improve their art as much as possible. Other people working in the industry, such as instrumentalists, must look into their jobs' futures and how they can adapt as AI improves. Finally, I also hope that general music enjoyers who want to learn about the future direction of music technology will find this work valuable.

There are many areas of research related to this work that could be explored. This work focuses on music AI; however, generative models are diverse and can affect various industries such as filmmaking, journalism, and visual art. Each of these industries will have unique responses to many questions I ask about music AI in this paper. This paper can serve as a stepping-off point for other applications of generative models.

One crucial area of research that I did not touch on was the technicalities of the legality of music AI; many legal questions surrounding the ownership of the generated art are currently fuzzy. One of the most topical examples is Deep Fakes: pieces of music that are life-like imitations of real artists. Determining how much of a claim these artists have to these models profiting off their likeness is an open question. Other legal issues to explore are data rights and the application of copyright laws. Answering these questions will shape the use of music AI by users and the music industry.

Despite the common negative sentiment around AI, especially music AI, I think the future is bright. Music AI systems will become exceptionally powerful tools that will shape the future of music. Throughout human history, music and music technology have co-evolved to expand the capabilities of human expression, and music AI is the next step in that evolution.

6. References

Anantrasirichai, N., & Bull, D. Artificial intelligence in the creative industries: a review. *Artificial Intelligence Review*, pages 1–68, 2021

Bakpayev, M., Baek, T. H., van Esch, P., & Yoon, S. (2022). Programmatic creative: AI can think but it cannot feel. *Australasian Marketing Journal*, 30(1), 90–95. https://doi.org/10.1016/j.ausmj.2020.04.002

Beatoven AI. (n.d.). *Royalty free music by Beatoven*. beatoven.ai. Retrieved May 4, 2023, from https://www.beatoven.ai/

Bijker, W. E., Hughes, T. P., & Pinch, T. (2012). *The social construction of technological systems: New Directions in the sociology and history of technology*. Cambridge, MA: MIT Press.

Creative tools for musicians. LANDR. (n.d.). Retrieved March 14, 2023, from https://www.landr.com/

de Mantaras, R. L., & Arcos, J. L. (2002). AI and Music: From Composition to Expressive Performance. *AI Magazine*, 23(3), 43. https://doi.org/10.1609/aimag.v23i3.1656

DelSignore, P. (2022, October 1). The following visual art jobs will be replaced by ai. *Medium*. Retrieved March 14, 2023, from https://medium.com/mlearning-ai/the-following-visual-art-jobs-will-be-replaced-by-ai-942f8558963e

Dhariwal, P., Jun, H., Payne, C., Kim, J. W., Radford, A., & Sutskever, I. (2020). Jukebox: A generative model for music. *arXiv preprint arXiv:2005.00341*

Ditto Music. (2023, May 2). *AI for Music Production: 10 Tools to produce like a pro*. Ditto Music Distribution. Retrieved May 4, 2023, from https://dittomusic.com/en/blog/ai-for-music-production-tools-for-musicians/

DjeDje, J. C., & Stone, R. M. (1983). Let the inside be sweet: The interpretation of music event among the Kpelle of Liberia. *Ethnomusicology*, *27*(3), 544. https://doi.org/10.2307/850662

Gaston, N. R. + R. (2019, November 8). *A brief history of Midi*. Retrieved May 4, 2023, from https://www.perfectcircuit.com/signal/history-of-midi

Holford, D. W. (2019). The future of human creative knowledge work within the digital economy. *Futures*, 105, 143–154. https://doi.org/https://www.sciencedirect.com/science/article/pii/S0016328717302598

Huang, M.-H., Rust, R., & Maksimovic, V. (2019). The Feeling Economy: Managing in the Next Generation of Artificial Intelligence (AI). *California Management Review*, 61(4), 43–65. https://doi.org/10.1177/0008125619863436

IFPI. (n.d.). *IFPI Global Music Report 2023*. IFPI GLOBAL MUSIC REPORT 2023. Retrieved May 4, 2023, from https://globalmusicreport.ifpi.org/

Kelly, K. (2022, November 17). *What ai-generated art really means for human creativity*. Retrieved May 4, 2023, from https://www.wired.com/story/picture-limitless-creativity-ai-image-generators/

MusicBrainz. (n.d.). *Database statistics*. MusicBrainz. Retrieved May 4, 2023, from https://musicbrainz.org/statistics

Pal, G., Saha S., and Anita. Musenet : Music generation using abstractive and generative methods. *International Journal of Innovative Technology and Exploring Engineering*, 9(6):784–788, 2020.

Prafulla D, H Jun, Christine Payne, Jong Wook Kim, Alec Radford, and Ilya Sutskever. Jukebox: A generative model for Music, 2020.

Roose, K. (2022, October 21). AI-generated art is already transforming creative work. *The New York Times*. Retrieved March 7, 2023, from https://www.nytimes.com/2022/10/21/technology/ai-generated-art-jobs-dall-e-2.html

Sacks, A., & Mariano, M. (2019). The Concerts and Live Entertainment Industry: A Significant Economic Engine . *Oxford Economics*. Retrieved March 5, 2023, from https://www.oxfordeconomics.com/resource/livemusic/

Sacks, A., & Mariano, M. (2019). The Concerts and Live Entertainment Industry: A Significant Economic Engine . *Oxford Economics*. Retrieved March 5, 2023, from https://www.oxfordeconomics.com/resource/livemusic/

Sandra E T, Judith B, and Iain M. Cross-cultural perspectives on Music and musicality. Philosophical Transactions of the Royal Society B: *Biological Sciences*, 370(1664):20140096, 2015.

Shiloah, A. (1997). Music and Religion in Islam. *Acta Musicologica*, 69(2), 143–155. https://doi.org/10.2307/932653

Standard MIDI-file format spec. 1.1, updated. Standard MIDI file format, updated. (n.d.). Retrieved March 4, 2023, from http://www.music.mcgill.ca/~ich/classes/mumt306/StandardMIDIfileformat.html

Technologies, N. G. (2022). Artificial intelligence market size report, 2022-2030.

Zippia. "23+ Artificial Intelligence And Job Loss Statistics [2023]: How Job Automation Impacts the Workforce" *Zippia.com*. Feb. 7, 2023, https://www.zippia.com/advice/ai-job-loss-statistics/