The Practicality and Future of Artificial Musicians

Kaeden Harrell
School of Design and Creative Technology
The University of Texas at Austin
Austin, TX. United States of America
kaedenharrell@gmail.com

Abstract - Machine learning and robotics are continually contributing to the advances of technology in many fields. Music is no exception: Artificial intelligence learns how to imitate and create new works using musicians' performances and movements. Robots are created to be able to play and perform on traditional instruments to imitate human melodic expression. This paper explores the methods and advancements that allow these artificial musicians to exist. Also, it will explore the uses of these technologies and how they will affect the future of music both practically and ethically.

Keywords - machine learning, artificial intelligence, music composition, robotics, expressive music performance, music, expression, ethics, AI, technology, musician, musical expression

I. INTRODUCTION

A. Description

Musical instruments have been found at archeological dig sites for decades across the world. From very primitive percussion instruments to flutes made from bone, there is evidence to suggest that music has been around almost if early humans have been showing signs of developing culture. One might conclude that music is a uniquely human or archaic human form of expression that is created by humans for humans and only performed by humans. This idea has remained unchallenged for centuries. However, with the progression of technology throughout the 20th and 21st centuries, this assumption and status quo are being challenged by the creation of artificial performers and advancements in the development of artificial intelligence enhancing and creating music[fig.2]. This article will be following developments and the applications of robotic and artificial intelligence using a review of academic applications. The practicality and ethics of such technologies will be discussed as well as the future of the field.

(Fig.1) A Neanderthal flute is made out of bone and is estimated to be at least 60,000 years old. This makes it the oldest instrument found. It was discovered in the Divje babe cave near Cerkno.

B. Background

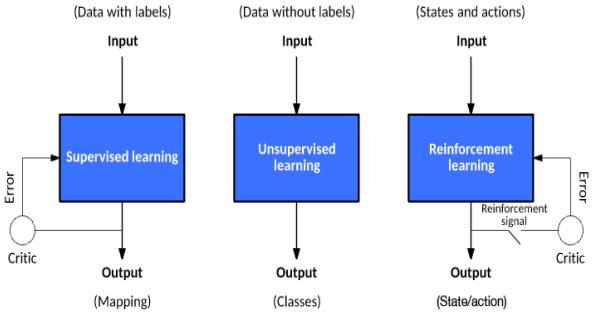
The idea of artificial musicians and intelligence has been a concept that has always had my interest. The first time I encountered this concept was when I was visiting the Toyota Kaikan Museum in Japan in 2014. There I saw the Partner Robots performing "When You Wish Upon a Star" on the violin and trumpet. I was blown away by

the fact that a robot could play the violin, much less the trumpet, which requires very precise lip and air control. By human standards, the performance was nothing special, but to witness this happening with no knowledge that this could be done was extraordinary, to say the least. It made me excited about the possibilities the combination of music and technology could bring to the world. These robots were not the first or the last artificial musicians to be shown. They have only become more common and complex over time.

II. ARTIFICIAL INTELLIGENCE ENHANCEMENT OF MUSIC

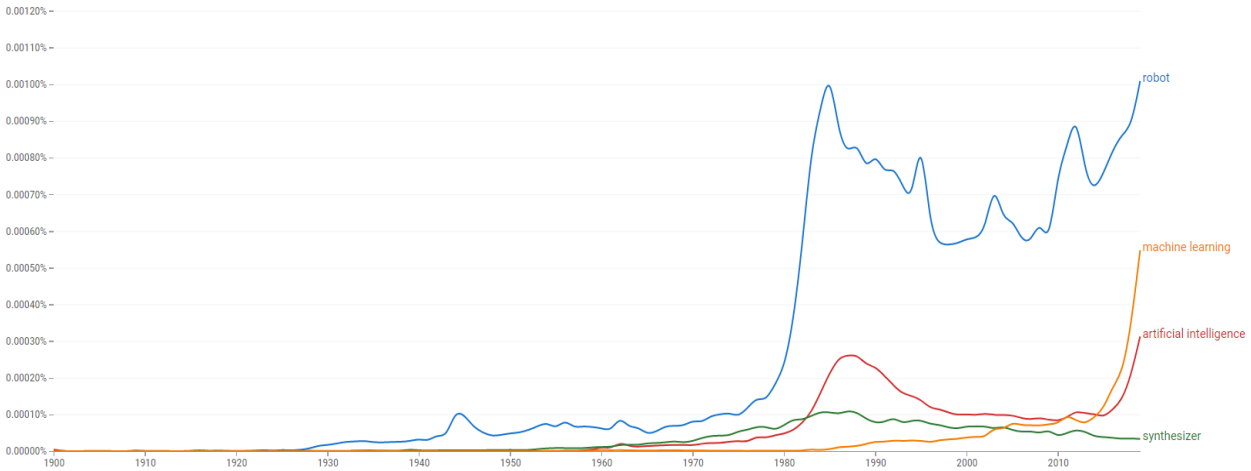
A. Deep Learning, Machine Learning, and Artificial Intelligence

It is a fact that artificial intelligence is the basis of both machine learning and deep learning. The basis of artificial intelligence is software that can complete an objective that is typically associated with or requires intelligence comparable to a human. Then using a predetermined amount of data or instructions to complete the task with little to no human input. The next step in the evolution of the complexity of artificial intelligence would be machine learning. One could define machine learning as the training of software with a limited number of datasets to predict or emulate a pre-decided desired outcome. Basic artificial intelligence is mostly rule-based. This can be something like a chess bot that goes against players using the pieces on the board to calculate the best move to increase its chance of winning or a system to help doctors diagnose their patients by checking off their symptoms to narrow down options. While one can have artificial intelligence on its own, one cannot have machine learning without artificial intelligence. Thus, there is an important distinction to make: While the two are similar, can complete comparable tasks, and are commonly confused, they are different in how they operate, process data, and create solutions.

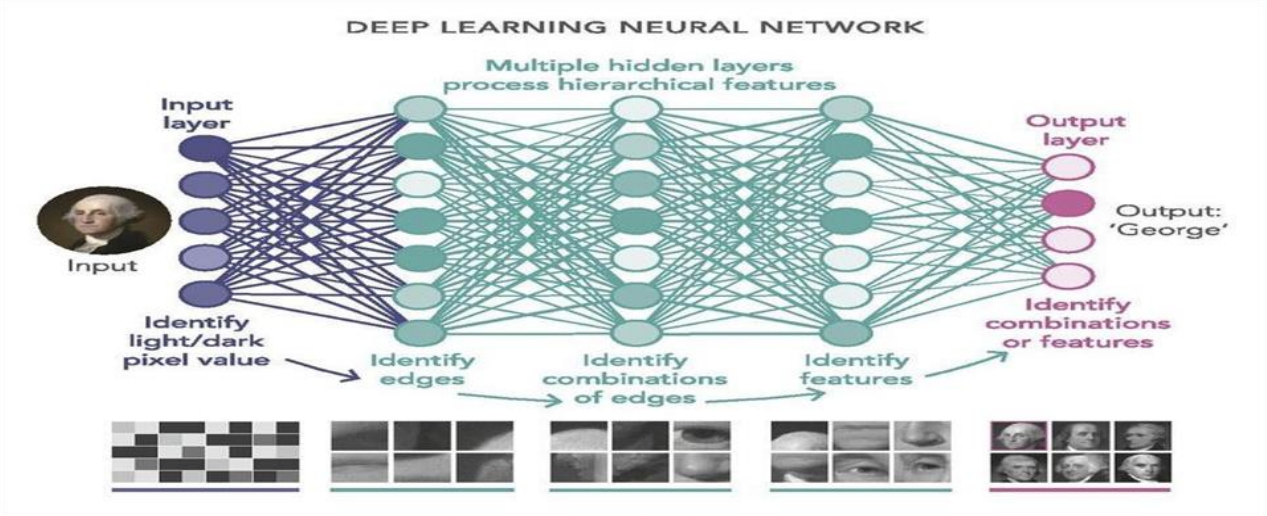


(Fig.2) Three ways machine learning can be accomplished. Two of the three require human intervention to correct or reinforce learning by the algorithm.

Machine learning, by principle, mainly learns from pattern recognition, which is generally curated by the creator(s) of the software which is one of the more commonly used forms of artificial intelligence by the public. Software that takes new imputed information and identifies patterns from learned data to make predictions. This software is generally trained by the creator who takes labeled training datasets and gives it to the program to learn to



(Fig.3) Using Google Books Ngram Viewer, this figure shows the usage of four terms and their appearances in published books from 1900 to 2019.



(Fig.4) The objective of this neural network is to determine the correct founding father based on image input. It goes through several stages analyzing the features of the photo to output the correct answer cross-comparing information based on its' learning.

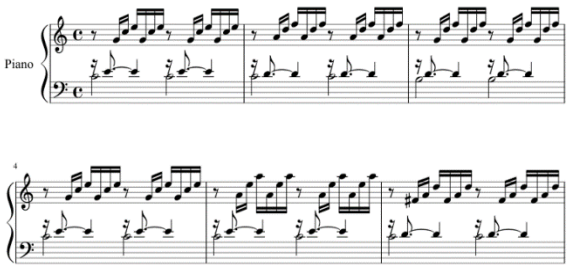
accomplish its predesigned goal. It learns from the data by making predictions and lines of learning if it is right or wrong. Over time, it will make its own decisions based on what it has learned and its originally predesigned goal that its creator set out to make.

Deep learning is a subset of machine learning. It follows the same relationship that machine learning has with artificial intelligence. It is debated whether deep learning is its' own topic or just a more focused version of machine learning. Deep learning was created from the inspiration of the human brain's process of learning and creating solutions. It is built upon multiple artificial neural networks that require little to no human correction to learn from as it goes through its learning process

Deep learning can also be thought of as multiple machine learning algorithms working together to solve more complex demands. This technology is seen in things such as self-driving cars and facial recognition systems. These systems take in new data and respond in real-time to new situations based on their learned experiences and training. It is what allows the car to recognize a stop sign or a pedestrian it has never seen before and come to a complete stop. The levels of complexity and variety of applications are still being explored to be implemented in virtually every field, including the fine arts such as music.

B. Schenkerian Analysis Using Machine Learning

The idea of implementing any kind of artificial intelligence into the world of music can seem strange to a world that has existed acoustically for so long; however, this process can lead to information that one can only find through analysis of music using machine learning. Such a use could be Schenkerian analysis, which is based on Heinrich Schenker's theories of tonal music. In its most simple form, it is used to dive deeper into music than just its written form. This is traditionally practiced by looking at the larger form in phrasing and the observation of the pattern of its notes. The main idea is to take the complexity of the piece and simplify it into a guess of the perceived goal of the composer. Following Phillip B. Kirlina and Jason Yust, we can see how they attempted to bring a more quantifiable approach to a study that generally advises against such practices [8].



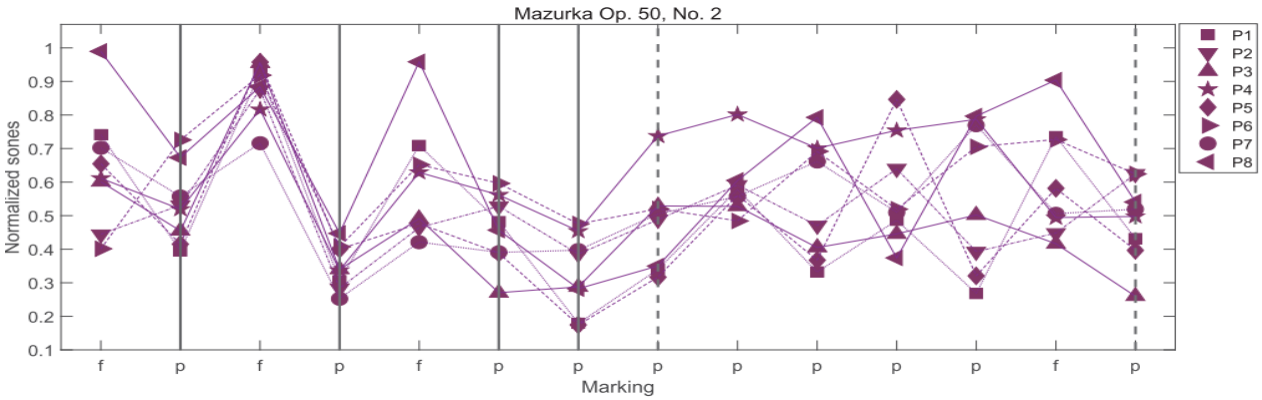
(Fig.5) This is the first few measures of Bach's Prelude in C. The first four measures make a chord progression of I to ii to V to I when looking at the melodic context. However, when looking at the function of the four measures and the harmonic context, it functions as a I chord. The goal of stability reveals itself as the progression does not go anywhere musically.

Their study explores how using a deep learning algorithm could accurately analyze the data given to it and reveal what is the most important information in the analysis. They broke the factors of Schenkerian analysis into melodic, harmonic, and metrical features. Melodic context can be defined as how the melody of a piece can show or give a suggestion of how the overall portion functions. Harmonic context typically follows chord progressions of a piece but can also be something as simple as a counter melody. The final context is metrical or rhythmic factors. Metrical components are the notation of a piece such as a tempo, time signature, rhythms, and rests which can give the context of how phrases are meant to be grouped.

Following these rules allowed them to test their machine learning on each of the features to see what factor contributed the most to the success of an accurate analysis. Training their algorithm by data from Schenkerian analysis text and examples done by humans, experiments showed low accuracy. By omitting certain variables across experiments, the algorithm was able to become more accurate. While metrical information was found to be the least important, harmonic information was found to be the most requisite in analysis. This was because harmonic context is crucial to the conditioning of predicting patterns of the features and identification of non-chord tones. It shows how harmonic context is integral to the process of the analysis. Following the interval patterns and scale progression of the melodic information was also essential and aided in the accuracy of the algorithm when added to the factors of harmonic context. Since harmony and melody are two sides of the same coin, it is often assumed that this is the case but because the variation of Schenkerian analysis changes from the scale of the sample; it gives evidence that the "correct" answer can change when looking at a pattern from note to note instead of from measure to measure.

C. Study and Enhancement of Notation Using Machine Learning

The notation of music is only written for the performers typically referring to the notes on the page. Dynamics are another key factor in the notation of music. Pieces will indicate the level at which notes are to be performed. This does not mean that it is a universal standard, however. Fortissimo is considered to mean very loud by musicians, but exactly how loud does that mean? Who decides the definition and pinnacle example that all fortissimos are based on? One's perception and senses are what it is based on. One performer may play two pieces that have dynamic markings that are used but play them at different levels. Another scenario could be two performers playing the same piece, but their performed dynamics differ. Basing the study of 44 recordings of Chopin's Mazurkas, a deep



(Fig. 6) This graph displays the loudness values for Mazurka Op. 50 No. 2. The solid vertical lines show that all eight recordings are classified correctly. Dashed vertical lines that seven out of eight recordings were marked correctly. P1-P8 indicate each of the different pianist tracked.[5]

learning algorithm was created to study the correlation between marked dynamics and performed dynamics from 8 pianists [5].

The methods of analyzing and predicting the dynamics of the performances followed different machine learning approaches. Decision trees, support vector machines, artificial neural networks, and k-nearest neighbor algorithms were all used and applied to the recordings to varying successes. The results show that in the same piece, the algorithm can track the levels and markings with fair accuracy but has less success when following different pieces by the same pianist.

The conclusion seems to be that using such methods to render musical expression into data to predict other pieces and performers' renditions is quite sensitive and not quite as exact as one might hope. The suggestion was that a 44-fold validation experiment might not have been enough data for the model to have been successfully trained. While being able to quantify a piece in data does provide insight into how they are performed on average hypothetical performance, ultimately performers will continue to make choices to change the performance for themselves and the audience.

III. ARTIFICIAL PERFORMATIVE MUSICIANS

A. Robotic Technology

The idea of an automaton or a robot is nothing new, while the technology to sustain such creations is relatively new compared to the concept. There are many types of robots, it can be understood that a robot is a machine that can perform or replicate tasks that are typically associated with or for humans. This can range from something that looks like a human, an animal, a plant, or just a normal machine. The limits of what can be done have not fully been explored yet. This variety of form factors adds to the ability to create artificial musicians that can perform in a way that normally could never be done by normal humans. However, some wish to replicate actions or abilities performed by humans in their robotics. Music and instruments are not an exception to this exploration of what can be achieved.

B. *Artificial Percussion Players*

Percussion is one of the oldest forms of music that humans have created, and its variety can be seen throughout the world. The simplicity of making sound through the striking of an object makes it a prime candidate for making artificial performers. Michael Krzyzaniak's djembe robot, Kiki, shows us how such things are possible, using the combination of robotics and machine learning to create an artificial musician [6].

(Fig.7) This is the final version of Kiki developed by Michael Krzyzaniak. It features a frame to secure the djimbe and three robotic arms that can move to strike for the different strokes required to perform. Here are examples of Kiki demonstrating three strokes of the djembe.[6]

Timbre is the perceived character of sound that is produced from a musical instrument or voice. There are many ways to describe different timbres, but their use and variety help contribute to the diversity of musical expression. Timbre is a French word that is derived from the Greek word that means drum. Most drums cannot change pitch easily but rely on the manipulation of timbre for expression. The goal for Kiki was to be able to produce human-like timbres in the form of the three djembe strokes of bass, tone, and slap in a dynamically expressive matter. The bass stroke is made by striking the drum in the center with an open hand. The tone is created by the palmar crease striking the rim of the drum so that four ridged fingers strike the edge of the drumhead. Slap tone is made in a similar matter as tone, but the fingers are looser and curved so that only the fingertips make a connection.

The hands were constructed with an aluminum and spring-steel skeleton for flexible and lightweight movement. The 'muscles' were made from a vinyl core and silicone. The final layer of 'skin' was made from latex. This combination was the closest to accomplishing the practicality and sound of a human hand. Three arms were created because of how different timbres can be created based on the location of the drum hit and the angle of impact. Krzyzaniak asked musicians who claimed percussion as their main instrument, "Do you have any remarks about Kiki's use of timbre?" after giving a demonstration of its abilities. The overall comments seemed to say that although not perfect compared to a human, Kiki was able to perform a variety of timbres and strikes that most would not associate with being produced by a robot.

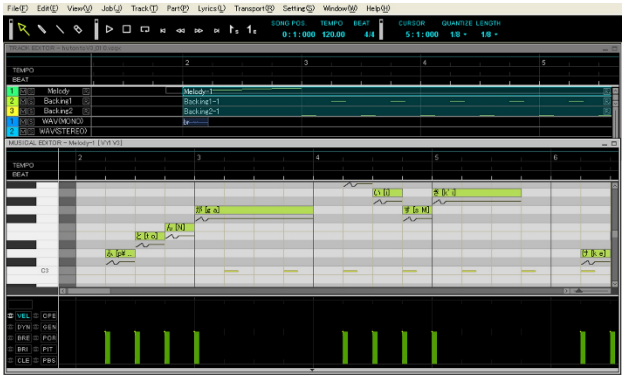
This artificial musician is a fantastic case study of what robotics can do in the realm of percussion. It makes us evaluate how we as humans create sounds with our hands and how our hands can create unique properties that other materials cannot. The djimbe is a particularly hard instrument to perform because of the complex rhythms and strokes that are played. This makes the ability of Kiki to perform even more impressive.

C. *Artificial Vocalists*

The human voice is the most accessible and oldest instrument known. Every person capable of speaking can sing with enough practice. It makes sense that this logic could apply to machines as well, with enough development any machine could sing. With the boom of computers and musical synthesis in the mid-twentieth century, the combination of the two into speech synthesis was only a matter of time. The first computer to ever sing was an IBM 7094 at Bell Labs in 1961. The song performed was 'Daisy Bell'. Listening to it now, it seems very odd, strangely creepy, and very inhuman. However, at the time it was revolutionary and inspiring for many. Famous Sci-Fi writer Arthur C. Clarke had the chance to witness one of these performances. It was so

revolutionary and left such an impression on him that he decided to use it as part of the climax for his screenplay, *2001: A Space Odyssey.* Now we have technology that far surpasses that of the IBM computer such as the Yamaha Vocaloid technology

The Vocaloid is different from anything discussed before because this is a commercially available product. Anyone from beginner to professional can buy the software online and start writing songs with an artificial singer. The software takes all the possible combinations of syllables, pitches, and dynamics so that it can output the desired results based on user input. This software plugs in directly to most Digital Audio Workstations and works from midi. The truth behind the scenes technology of Vocaloids has not been seen because of this commercial aspect of it but it has not stopped it from becoming a huge success.



(Fig.8) This is an example of Vocaloid 3 from their website. Notice the syllables are written where the notes are to indicate what is being sung where.

The popularity of this software can be seen in the artificial Japanese Idol, Hatsune Miku. Hatsune Miku is a persona of one of the available voices. Her voice is originally sampled from the voice actress Saki Fujita. The Vocaloid technology uses Miku as their poster girl. She has an anime design and can be seen on everything from shirts to cell phones to figures. Here is a live performance of one of her most popular songs in a sold-out arena. She is a very large success, to say the least.

The main difference between Miku and a human singer is her accessibility and consistency [7]. She will sound the same for every recording or concert if the creator so chooses. The human voice can get sick, age, and tire which can prevent singers from performing the same way every time. Advancements and the use of post-production on voices cast an unrealistic portrayal of singers in recordings that can be hard to perform live without unfailing audio technology. The perfect consistency that Miku displays is something human singers could never accomplish. This helped Miku become popular as fans or official creators could make their own songs that use her professional-sounding voice that never fails. Plus, having fans create songs is a way for fans to be more engaged, free advertising, and sells more of the product for Yamaha. Currently, Vocaloid is the most popular voice synthesis software on the market. It might be a while till truly original computer vocal singing synthesis is in a state where it is performable. Until then, Vocaloids like Hatsune Miku can provide us with songs and shows to help further all artificial musicians around the world.

IV. Practicality and Ethics

A. Practicality of Artificial Musicians

When looking at the practical use of artificial musicians one must be open to the form of what the artificial musician is. As outlined before, these musicians can be tools to be used by humans as an aid or something that can perform by itself. Robotics is the closest thing to physical performers that one could see but a program run on a laptop could also qualify as a performer. As with human musicians, the quality of these artificial musicians varies with both the time that has been put into the development of their skills and the costs that it takes for them to be used. With human musicians, they spend their time practicing and widening their abilities. This is done over a long time with improvement generally in a linear path rather than a timeline of a project.

Almost every musician comes with a price for their performances and that price depends on their quality, demand for their style, and popularity. Artificial musicians are similar in many of the same ways. It takes time to develop the technology to be able for it to become operable and produced. The overall quality of artificial musicians is not comparable to the quality of seasoned professionals. Currently, one could go to any college, bar, or even high school to find a musician that is better at performing than any artificial musician. For this reason alone, one could consider the concept of artificial musicians impractical but that is not even where it ends. The price that it takes to develop and make these artificial musicians is drastically different than if you were to hire human musicians. Our current technology cannot sustain such things now because of the practicality, cost-effectiveness, and convenience that human musicians provide.

The definition of what is practical when it comes to having artificial musicians perform boils down to if one is willing to deal with the very niche abilities one can perform, the time it takes to develop, the size/transportability of the artificial musicians, and the quality at which it can perform. At this moment in time, no artificial musician can outperform human musicians in any of these categories or do so in a way not normally capable of human abilities.

B. Ethics of Artificial Musicians

Fear of machines, robots, and artificial intelligence replacing human jobs has been a worry for years in every job field. This worry is not unwarranted since it has been seen in fields such as factory work where workers are replaced with cheaper and more efficient robots. Many have assumed that the fine arts will be the last bastion of automated jobs in the future, especially music. The goal of most artificial intelligence and robots is to automate and make jobs easier. Music's goal is for the entertainment of oneself or others. These two things are contradictory in the eyes of many. This is for good reason because music is an expression of human emotion by humans for humans.

When used as a tool by a human musician to enhance either their or other musicians' abilities, there are little to no ethical concerns. Choosing to play on a piano or a synthesizer does not mean that a musician is not playing a piece. The tools of using robotics and artificial intelligence are the most common and helpful to musicians currently. It is as if using an algorithm to organize and curate a search result or a playlist. Some people get paid to develop the technology, some people pay to use it, and people who use the technology to make money in a way that they use it uniquely for themselves. Technology is meant for the betterment of our lives. When technology takes away the income or job of an individual is when there are ethical concerns.

C. Future of Artificial Musicians

The hypothetical future where artificial musician technology is advanced enough to be comparable to humans is where most discussions of what the future will be like are set. There are countless possibilities both positive and negative for the future. Venues could choose to permanently house a robotic band that can perform various genres rather than hiring bands, this leaves many individuals without a place to perform and make money. Why hire a composer for your film project when you can tell an artificial intelligence exactly what you need, and it can have it ready in minutes at a fraction of the cost? This might seem like a reach too many to replace artists with artificial ones because why would someone want something that is emulating human music?

Legal questions such as 'If a machine creates a new piece of work, then who owns the Copywrite to it?' will be asked in the future. Celebrities such as Jay-Z have already tried to take down videos that have been created with deep-faked audio and music made by artificial intelligence that was trained to mimic his voice. This is already an issue that is being debated so one can imagine that the waters will become murkier as artificial

musicians have more independence from the input of humans.

Artificial performers, such as Hatsune Miku, can sell out arenas and many songs to human audiences. She has reached millions of people with her songs and concerts and sells various forms of merchandise all over the world. However, the only thing that is artificial in her success is the vocal synthesis technology behind her voice. Her marketing, songwriting, performances, and public image have all been things created by humans to make money from other humans. This will more than likely be the case for most artificial musicians in the future, a novel thing that is interesting and entertaining because it goes against the norm of what a musician is. However, ideas and concepts of what is real or normal change over time. Music is no exception to this. The novelty of this experience might not even register with some, and they purely enjoy the music for what it is.

In the end, music is a subjective experience like any art. There is the experience of the creator and the consumer. Art can exist solely for the fulfillment of the creator. There is something about it that seems to be intertwined with human nature. Even if artificial musicians were to take every job available, humans would still seek to play if not only for their own being. Just as Kiki and its creator can play together to make music, the combination of artificial and organic musicians will create new and interesting possibilities for music in the years to come in ways that no person can predict.

V. Conclusion

A. Summary

In this article, we have explored what makes an artificial musician, the technologies that allow for their creations, the use of that technology, and its future in our world. They are more commonly used as a way of research rather than practical use currently. The insight and tools that artificial musicians can provide information that innovates and improves human musicians. Given the limitations of materials, technology, and limited research on the topic compared to others of today, the conclusion one can draw is that artificial musicians are not practical in their current forms. The development of artificial musicians will grow and become more complex as technology improves. A future for these artificial musicians is not guaranteed but as with every musician, it takes time to develop the skills to be able to perform.

B. Acknowledgments

I would like to firstly thank my mother for always telling me that she is proud of my work. Thank you to my thesis advisor, Maura Murnane, and former advisor, MJ Johns, for helping focus my ideas on this thesis and giving fantastic feedback from their perspectives. My peers and roommates played a very important role in helping me finish this thesis whether it be helping me think out concepts or giving me moral support. Last but certainly not least, thank you to my faculty advisor Chris Ozley. He has been a great mentor for this thesis and throughout my collegiate career. I am forever grateful for all the advice he has given me over the years.

REFERENCES

[1] J. Solis, K. Ozawa, M. Takeuchi, T. Kusano, S. Ishikawa, K. Petersen, and A. Takanishi, "Biologically-Inspired Control Architecture for Musical Performance Robots," International Journal of Advanced Robotic Systems, vol. 11, no. 10, p. 172, 2014.

[2] J.-P. Briot and F. Pachet, "Deep learning for music generation: challenges and directions," Neural Computing and Applications, vol. 32, no. 4, pp. 981–993, 2018.

[3] J.-P. Briot, "From artificial neural networks to deep learning for music generation: History, concepts, and trends," Neural Computing and Applications, vol. 33, no. 1, pp. 39–65, 2020

[4] J.-Y. Lin, M. Kawai, Y. Nishio, S. Cosentino, and A. Takanishi, "Development of Performance System With Musical Dynamics Expression on Humanoid Saxophonist Robot," IEEE Robotics and Automation Letters, vol. 4, no. 2, pp. 1684–1690, 2019.

[5] K. Kosta, R. Ramírez-Melendez, O. F. Bandtlow,

and E. Chew, "Mapping between dynamic markings and performed loudness: A machine learning approach," Machine Learning and Music Generation, pp. 63–86, 2018.

[6] M. J. Krzyzaniak, "Timbral Learning for Musical Robots," dissertation, 2016.

[7] N. Prior, "On vocal assemblages: From Edison to Miku," Contemporary Music Review, vol. 37, no. 5-6, pp. 488–506, 2017.

[8] P. B. Kirlin and J. Yust, "Analysis of analysis: Using machine learning to evaluate the importance of music parameters for Schenkerian analysis," Machine Learning and Music Generation, pp. 41–62, 2018.

[9] P. J. Ponce de León, J. M. Iñesta, J. Calvo-Zaragoza, and D. Rizo, "Data-based melody generation through multi-objective evolutionary computation," Machine Learning and Music Generation, pp. 87–106, 2018.