Senior Project Proposal

Edward Wang

January 12, 2024

I Title of Project

Humans vs. AI: Can we tell the difference musically?

II Contact Information

- My Phone Number: (520) 491-0213
- BASIS Advisors' Names: Dr. Jeffrey Winslow, Dr. Christian Martin
- Internship Location: N/A - Independent Research
- External Advisor Information:
  - Name: Dr. Rita Borden
  - Email: [redacted]
  - Phone Number: [redacted]

III Statement of Purpose

With the emergence of self-driving cars, extensive text-based AI models, and self-service restaurants, AI has become a hot topic in our modern world. Despite the benefits of this technological growth, AI poses a threat towards jobs that require lower skills, such as serving food, cleaning streets, and driving vehicles. AI can drive much better than humans can, AI can write much better than humans can, and AI can interact with humans much better than humans can. The creative sector has always touted itself as being immune to AI, but now AI can also create paintings, compose music, and write literature. So how much of a threat does AI pose to musical composers? Is there a way to distinguish AI-generated music from human-made music? And if not, what implications will that have on the music industry?

IV Background

Growing up, I have always loved coding and music, and I knew starting from middle school that I wanted to end up going into an electrical engineering or computer science field while keeping music close by my side. I loved going to coding camps and writing small JavaScript programs that would walk a character across the screen or make them do a little dance. Going into high school, my eyes were set on taking AP Computer Science A, and taking that class really helped me to understand how computers interpret and interact with the code I write and led me to branching out into languages like Java, Python, and Ruby.

Going into Senior Projects, I knew I wanted to do something relating to coding and music and undertake a larger coding project unlike anything I had ever done before. Despite my inexperience, I spent a lot of time reading into coding languages like Python, which is used a lot in AI and machine learning. I really leaned into the musical aspect of the project when I saw MusicLM, a music-based model that takes text prompts and writes music accordingly. The piano pieces felt surprisingly real, like something that I could play on my own piano, and the rhythms and harmonies and melodies were consistent throughout the piece, almost to the point that besides a robotic twang in the background, the piece was near indistinguishable from human-written music. I knew I wanted to look into making AI make piano pieces and comparing them to my own compositions, and I fell in love with the idea of having a more independent research project to work on creating music and assessing the ability of humans to pick out AI from among their own.

V Prior Research

The Western definition of art splits the term into four distinct categories: visual arts, performing arts, literary arts, and interactive media. This project will focus mainly on artificial intelligence within performing arts. Although there have been several musical systems throughout human history, the start of what we consider "Western music" stems from Ancient Greece, with the rise of whole number ratios in tuning. This eventually develops into to a seven-note scale throughout Medieval Europe which marked the start of the five main periods of Western music: Renaissance, Baroque, Classical, Romantic, and Contemporary (Lucas, 1893). This project aims to gauge how well artificial intelligence will be able to replicate specific challenges presented in the music of these periods.

The Renaissance and Baroque eras were marked by polyphony and ornamentation and decoration, with lavish harmonies and grandiose music embracing small ornamentations in the music (Mackay 2016). The Classical period is defined by "good" harmony and structure that sought to replicate Ancient Greek musical systems (Blume, 1970). The Romantic period was marked by individuality and emotion; composers conveyed personal feelings through music that focused on harmonies that used more complex chords and chord progressions, diverging from the simpler, "good" harmony of the Classical period (Truscott, 1961). The Contemporary period is marked by a separation between music and structure, focusing on abstract musical ideas such as polytonality and odd time signatures, and a rebellion from the harmonic doctrines of previous periods (Schwartz, 1993). This project will mainly focus on if artificial intelligence will be successful in replicating these various styles of music.

AI also has several hurdles to jump over concerning musical theory concepts, such as pitch, time, melody, intervals, and harmony. While humans are hardwired to process music—neurons all around our brains fire when we hear music, and we are predisposed to recognizing it over other sounds, similar to how we are predisposed to recognizing human voices—artificial intelligence is unable to recognize why a musical piece "sounds good," it only has the ability to essentially throw a bunch of musical bits and pieces from songs it is trained on at the wall and hope something sticks.

Musical artificial intelligence uses ANNs (artificial neural networks) to replicate musical sounds from pieces to use in its own compositions. The ANN is composed of layers with connections between nodes of different layers. The first layer receives inputs (here it would be the different musical pieces) and processes it into data for the next layer, with weights determining how the data is manipulated from layer to layer, and this continues through each layer until the AI outputs a "final product." To improve the AI's final product, a backpropagation algorithm updates the weights, and the AI makes another output. The backpropagation algorithm updates the weights again, and this process of constant error correction continues until the AI is decently suited to musical composition (Hardesty 2017).

This project revolves around the MuseNet learning model created by OpenAI. This model is a sort of "umbrella model" in that it was designed to focus on many genres of music. The model has been used to combine works by classical composers like Mozart and Rachmaninoff with modern songwriters and musicians like Lady Gaga and Adele, unlike other models that may be more focused on a specific genre of music. The model is also capable of creating standalone works based on the compositions of a specific composer, which will be utilized to create audio files for the project (OpenAI 2019).

More information about prior research is included in the addendum of this proposal.

VI Significance

The most important part of this project is whether human-made and artificial intelligence are indistinguishable. With the advent of new artificial intelligence technologies in creative sectors such as literary, visual, and performing arts—the most prominent of these being the new progresses made by the GPT-4 language-based models—the idea of humans being unique in our creative endeavors is being challenged. For art produced by an artificial intelligence such as DALL·E 3, it is becoming harder for the average viewer of visual art to see a clear difference between the two (Rogers, 2023). You might be able to see a problem with the hands not having enough fingers or the eyes being wonky or the brushstrokes going in the wrong direction, but even these problems are key determiners that art is AI-generated. Many of the artists in Rogers's article state that "there's no logic behind it" and "it lacks soul," but these are not quantifiable factors that can determine if the art is created by a human being or an AI. Because of this, there have been several cases of human-made art being flagged as AI art across forums online.

This situation in the visual arts community is almost perfectly reflected in the performing arts community. For the average listener, there is not a precise way to understand the difference between human-made music and AI music. A large portion of music created today uses predictable patterns in their chord progressions or musical form which can be replicated very well by artificial intelligence. Musicians can say that the pieces "lack form" or "sound wrong," but similarly to AI in visual art, these factors are not quantifiable and are dependent on the musical experience of the person listening to these two pieces (Mahendra, 2023). A gut feeling is almost nearly all that humans can rely on in telling the difference between human-made and AI art, so it is important to find a "marker" of some sorts that signifies the use of artificial intelligence.

The inherent repetitive nature of music means that it is also extremely easy for artificial intelligence to rip off copyrighted music. There have already been feuds between human creators over music sounding the same, whether it be Katy Perry's song "Dark Horse" having a similar ostinato to Flame's song "Joyful Noise" or Dua Lipa's song "Levitating" sounding like Artikal Sound System's song "Live Your Life" (Neely, 2019, 2022). Both have led to large lawsuits that have set precedents of "similar but not the same" in the legal system for musical copyright, but especially since artificial intelligence must train off databases, it takes other artists' works to make its own. Because of this, there is a moral dilemma of whether artists' songs are stolen to feed an artificial intelligence that replicates their music without any features to distinguish real music from the artist to fake music generated by the artificial intelligence and if this music deserves to be copyrighted.

VII Description

This project will largely consist of internet research revolving around the topic of setting up the MuseNet music model to produce musical pieces. These pieces will be used in a survey that will gauge how well participants are able to reliably identify AI music from manmade music. Finally, a binomial distribution will be used to analyze the data from the survey and make conclusions about the human ability to differentiate AI music from manmade music aurally. The final product will consist of three pieces created by me and three pieces created by the AI which will be used in a piano performance following the project's conclusion.

VIII Methodology

To research if people can tell the difference between music made by humans and music made by artificial intelligence, I will be using OpenAI's MuseNet neural network to generate ten fifteen-second musical clips from different eras and genres, each of which I will complement with my own fifteen-second compositions and two unique compositions from two different composers from that specific era and genre. I will have to pick more obscure compositions to prevent errors originating from people recognizing a tune as having been made by a human.

Although I will be largely doing my research independently, I will be consulting with Mrs. Rita Borden about the form of my compositions and which pieces I should pick to compare between my compositions and MuseNet's compositions. I will also be consulting with Dr. Jeffrey Winslow with help regarding how to set up MuseSet, train it on a database of music that originates from a specific era using large MIDI databases of pieces such as IMSLP, and create its own compositions.

From there, I will transcribe all the music into MuseScore 4 and re-render them. This way, errors that could originate from different sound fonts can be removed from the project. This process will produce a variety of mp3 files that I will use in my survey.

I will compile the audio files in a survey that randomizes the order that the files come in and provide the genre that the music is from. In addition to these questions, I will also be asking the person taking the survey about their experience in music, more specifically how many years of experience the person taking the survey has in music (lessons, programs, etc.), which instruments the person taking the survey can play, and the skill level the person taking the survey believes their musical experience to be at.

In the end, I will look at the results to see if people were able to successfully narrow down which audio recording was made by the artificial intelligence by using a binomial distribution, assuming that a random guesser would get the correct answer only 25% of the time for each question, to see if there is a statistically significant difference between the results of the survey and random guessing.

In the survey, I also plan to ask participants to rate the emotions that each piece can make them feel on a scale from one to five (one being "does not evoke emotion" and five being "evokes a lot of emotion"). Because of this, I will pick pieces that exemplify a specific emotional category (happy, sad, angry) and ask participants how much each piece is able to effectively evoke the specified emotion. Doing this, I aim to see if artificial intelligence can evoke emotions as well as manmade music.

At the end of the two months allotted for this project, I aim to gain a better understanding of whether music made by artificial intelligence can accurately be differentiated from music made by humans to make a realistic judgement on whether artificial intelligence has the potential to take over aspects of the music industry and the performing arts community such as songwriting or instrumentation.

IX Problems

The most notable problem that I face in this project is that I must rely on a survey to obtain the results I want. Because of this, I am constricted largely to the population within Flagstaff as it would be much harder to spread my survey to a wider population outside of the city, which means that it is much more difficult for me to generalize my findings to the general populace. There is a greater chance for human error to exist within the survey, whether it be difficulty following the instructions in the survey or problems with the audio files. It is especially difficult to set up forms with audio files as the answers, so it is very likely that I will have to host the audio files on a website and number them to have people match them up on Google Forms. This may cause some errors in asking participants to match up each audio file with each question in the Google Form.

The length of the survey may also turn out to be a problem, as people may not complete the survey. Although the survey should take only ten minutes, it is worth going through a test run for the surveys to see if some of the audio samples need to be removed to allow the survey to take up less time and retain more respondents to the end of the form.

Another problem that I must consider is that OpenAI's Jukebox uses up a large amount of the GPU, meaning even a fifteen-second render will likely take many hours, especially considering that my computers are relatively older and slower compared to the supercomputers that were used by OpenAI's research team. This means that I must begin rendering approximately a month before I release the survey to give myself enough time to effectively render the audio files for the survey. To optimize this process, I will be consulting with Dr. Jeffrey Winslow and online sources. With luck, a shorter rendering time will mean that I can have longer audio files for the final survey, but as the situation is now, the process is too slow to allow for longer or higher quality results. Because of this, I am largely constricted to shorter compositions with lower quality audio, but I believe that by running each audio file through MuseScore 4, the music created by the AI will be standardized within a specific sound font, which may smooth out errors within the AI-generated pieces.

X Budget

This project will ideally include no budget, as MuseNet is open source, and much of the equipment for the project is already in my possession. The highest the budget may be is $10, should I need to buy a website to host the audio files for the project, but it may not even be necessary, as the files could be stored within a GitHub repository instead. Should any costs arise, I will be able to pay for them.

XI Annotated Bibliography

Blume, F. (1970). *Classic and romantic music; a comprehensive survey.* United Kingdom: W. W. Norton.

Hardesty, L. (2017, April 14).*Explained: Neural networks.* MIT News. https://news.mit.edu/2017/e neural-networks-deep-learning-0414

Mackay, A., & Romanec, C. (2016, December 7). *Baroque Guide.* Tafelmusik. https://web.archive.org/web/20161207213948/http://www.tafelmusik.org/sites/all/themes/tafelmusik

Mahendra, S. (2023, June 13). *Can People Tell the Difference Between Music Created by AI and People?.* Artificial Intelligence +. https://www.aiplusinfo.com/blog/can-people-tell-the-difference-between-music-created-by-ai-and-people/

Neely, A. (2019, August 2). *Why the Katy Perry/Flame lawsuit makes no sense.* YouTube. https://youtu.be/0ytoUuO-qvg?si=EgSqwQ_jIHXd4lCZ

Neely, A. (2022, March 6). *Did Dua Lipa ACTUALLY Plagiarize Levitating?.* YouTube. https://youtu.be/HnA1QmZvSNs?si=0hDQfh-5YyBZN3bk

OpenAI. (2019, April 25). *MuseNet.* OpenAI Research. https://openai.com/research/musenet

Proceedings of the Musical Association. (1893). United Kingdom: Stanley Lucas, Weber & Company.

Rogers, R. (2023, January 21). *How to Spot AI-Generated Art, According to Artists.* Wired. https://www.wired.com/story/how-to-spot-generative-ai-art-according-to-artists/

Schwartz, E., Godfrey, D. (1993). *Music since 1945: issues, materials, and literature.* United Kingdom: Schirmer Books.

Truscott, H. (1961). *Form in Romantic Music.* In Studies in Romanticism (Vol. I, pp. 29–39). essay, Boston University Graduate School.

XII Addendum

XII.I Extended Prior Research

XII.I.I Abstract

For the last forty years or so, artificial intelligence in the music sector has been incomparable to human composition, but in the modern world, it is becoming harder and harder to effectively tell the two apart. Artificial intelligence faces problems in pitch space, musical time, melody and harmony synthesis, and replication of different styles of music. The improvement of neural networks through the number of connections and nodes has allowed artificial intelligence to bridge this gap in projects such as MuseNet and Jukebox, which has made distinguishing between AI and human music much more difficult. This project aims to see if humans can tell AI music from manmade music and gauge if AI music can convey emotions through music as well as or even better than manmade music.

XII.I.II Article

For thousands of years, we humans have seen human creativity and art as irreplicable, a diamond that is inherent to our nature yet unique in that it separated us from other animals. Although we different cultures have ascribed very different definitions of art throughout human history (Davies, 1991), the Western definition of art splits the term into four distinct subcategories: visual arts, performing arts, literary arts, and interactive media. Of these four, performing arts and visual arts are seen by anthropologists as the oldest of the two arts. We might never know which one is older; Paleolithic and Neolithic peoples did not build art to last, nor did they write down their music or words, but we have rough estimates of the origins of these art forms.

The size of the human brain has changed very little from our *Homo erectus* and *Homo neanderthalensis* ancestors (McDermott, 2021), and this has led many anthropologists to believe that we have engaged in visual arts and performing arts since the birth of humanity in the savannahs and plains of Africa. The first evidence we have of humans engaging in any form of art is the Divje Babe flute, also known as a tidldibab. The flute was carved by either Neanderthals or Cro-Magnons[1] from the femur of a cave bear and was found in a cave in Slovenia (Rizzi, 2023). The flute has been carbon-dated to be over 50,000 years old, originating from the Paleolithic era, and its sound is reminiscent of panpipes or traditional Native American flutes.

But what tunes would have been played on this flute? Could it have been strictly liturgical, used in the animist religions of the Paleolithic peoples? Or could it have been cultural songs featuring core cultural beliefs that Paleolithic peoples held? Or perhaps the flute was used purely for signaling? The fact of the matter is that we will likely never know. Even writing did not exist at that point in time, let alone musical notation, so the songs that were played on that flute are likely unreconstructible.

[1] Cro-Magnons are also known as Early European modern humans (EEMH). These people were distinct from the Neanderthals that lived in Europe prior to their arrival, as these Cro-Magnons were *Homo sapiens*, not *Homo neanderthalensis*. The Cro-Magnons are believed to have interacted and interbred with the Neanderthal population of the area, leading to the extinction of the Neanderthals and the establishment of *Homo sapiens* as the dominant species in Europe.

The oldest music notation that humans have discovered is the Hurrian Hymn to Nikkal, dated to approximately 1400 BCE (Hochelaga, 2020). The hymn was found in Ugarit, consisting of cuneiform carved into clay tablets. Written for a nine-stringed *sammûm* and a vocal, the hymn is dedicated to the Semitic goddess of the moon, Nikkal, but some problems lie in the fact that the notation it stems from does not describe the lengths of the notes, only the notes' pitches, thus the true melody of the Hurrian Hymn is likely lost to time. This has not stopped several reconstructions of the original hymn from arising in academic groups, but if you listen to the Hurrian Hymn, it might sound very unfamiliar and detached from music that we listen to today. This stems from the fact that our ears are tuned specifically to Western music system in general, which can be very different from the various music systems that can be found all around the world.

The music that we are familiar with today stems from Ancient Greece, with the rise of whole number ratios in musical tuning. This tuning system spread to the rest of Europe through the Romans, and eventually, following the fall of the Roman Empire, different areas bordering the Mediterranean Sea began to utilize this tuning system. Although this tuning system spanned a large portion of Europe and Northern Africa, the concept of specifically Western music did not take hold until the Renaissance era. Within the Christian monasteries spread across Medieval Europe, a system of six notes arose: *ut*, *re*, *mi*, *fa*, *sol*, and *la*. These notes came from a Gregorian chant concerning the worship of Saint John the Baptist (Lucas, 1893). This marked the birth of the modern Western music system.

In modern music history, there are five main periods of Western music: Renaissance, Baroque, Classical, Romantic, and Contemporary. Below I will be giving a brief overview of each to give us a sense of 1) what each musical techniques were in place during each period, 2) what an AI should be expected to mimic if it is trained on a database of music from only one of these eras, and 3) the extent of the history of Western musical theory.

The Renaissance period was marked by a newfound musical diversity. Rather than the Gregorian chants that were all sung in unison[2] , new experiments with harmony and chamber music were emerging. The musical period ranges from 1300 to 1600 and is characterized by new advances in ensemble music and fugal composition. The music moved away from the previous secular constraints to give composers a higher degree of control over their compositions' harmonies. Even with this shift away from the medieval constraints, counterpoint became more constrained, particularly how dissonances were treated. Even so, composers were scratching the surface of promoting expression and emotion in their compositions and shifting away from the unity that had marked sacred music in Medieval Europe (Arkenberg, 2002). Composers like Palestrina focused on fugal writing, which became the basis for Johann Joseph Fux's treatise on fugal counterpoint, *Gradus ad Parnassum* (Fux, *Præfatio ad Lectorem* 1725).

The Baroque period was marked by grandiose music that was filled with ornamentation and decoration. The Baroque period especially marked the height of fugal writing and advances in chamber music. The musical period ranges from 1600 to 1750 and was marked by the creation of common-practice tonality, which was an approach to writing music in a particular key, a type of harmony which has been cemented within the Western style of music. Improvisation was very common, with pieces like Bach's Prelude in B-flat

[2]It is believed that the Gregorian chants were sung in unison to represent the doctrine of unity in Catholic Christianity.

major in Book I of the Well-Tempered Clavier featuring quick runs and grand, rolled chords in harpsichord music. Fugal counterpoint was also very common, and the era was notable for progressing new ideas in counterpoint and establishing rules that still define contrapuntal writing to this very day. The era saw lavish harmonies and grandiose music that involved new instruments such as the harpsichord and the viola da gamba, and the era somewhat embraced small blemishes (ornamentation) in the music, as suggested by the name "baroque," which stems from the Portuguese word *barroco*, which means a "misshapen pearl" (Mackay, 2016).

The Classical period was marked by a return to structure and "good" harmony reminiscent of classical antiquity, more specifically that of Ancient Greece. Major advances were made not only in musical harmony, but also in sonata and symphonic form (Kamien, 2006). The musical period spanned 1750 to 1825 and was characterized by homophony, a shift from the polytonal fugues of the Renaissance and Baroque periods, but counterpoint was by no means forgotten, as it was used often by composers like Mozart or Beethoven. Another important style that emerged from the Classical period was the style galant that emphasized elegance in simplicity whereas the Baroque had emphasized seriousness and grandeur. Other styles of music also included Germany's *Empfindsamkeit*, literally sensibility, which aimed to create music that contrasted moods and sought to project more emotion through the music within a piece. The size of the orchestra would also increase, and instruments would become better at creating and projecting sounds. The Classical period was defined by structure, and it would be these concepts that helped to influence the Romantic era that was to come (Blume, 1970).

The Romantic period was marked by a divergence from the previous structure of the Classical period into more individualistic and emotional music. Harmonies became richer and more unique compared to the more stratified harmonies of the classical era (Truscott, 1961). The musical period spanned 1825 to 1925 and was characterized by individualism; composers composed what they wanted to compose and sought to evoke non-musical stimuli. Dissonance built up tension and unrest in the music that would be released in dramatic waves, and the concept of *Sturm und Drang* (literally "storm and stress") arose in Germany as conveyance of extreme emotions to the audience, such as fear or unrest, through foreboding minor keys. Impressionistic qualities also led composers to bend the color, timbre, and tones of instruments to mirror the abstract qualities that could be found in visual art at the time, giving rise to unique textures, ambiguous tonality, extensive harmonies, and the use of exotic scales and modes. Because of all these new advances, the Romantic era is arguably the most important to the foundations of Western music, and many of the pieces composed in this era have become commonplace among the modern populace.

Finally, we arrive at the Contemporary era. This musical period is marked by a separation between music and structure. Pieces are made with largely abstract musical ideas; pitch and melody are less important, and individuality is stressed. Dating from 1925 to the present, contemporary music has always been more controversial, mirroring new attitudes towards visual art which believed that anything could be art (Schwartz, 1993). Composers experimented with dissonant pitch language, creating atonal pieces that lacked a key center. There was also the development of new techniques that allowed composers to have greater levels of control in the compositional process, such as twelve-tone rows, which reordered the twelve notes in a scale to occur in a specific order within music that was dictated by the composer. Composers embraced chaos as well, making sporadic and unpredictable notes appear throughout their music. Despite these shifts to less structured music, movements like neoclassicism sought to reestablish older musical ideas like counterpoint or consonant harmonies in newer composition.

With this new era, however, there has been quite a bit of pushback. Recent music has become formless and unappealing to modern audiences. There is very much an elitist atmosphere that surrounds contemporary classical music, even as it departs from the usual musical expectations of modern audiences, and this is slowly disengaging people from attending contemporary classical music concerts. This is best summed up in this quote: "Music that sounds good is no longer taken seriously, and music that is taken seriously no longer sounds good" (Pleasants, 1955). There is an atmosphere of enlightenment that is cast down upon people who do not enjoy contemporary classical music, and this has slowly led the public away from modern classical music.

So how does this musical history overview coincide with having artificial intelligence create music? The main reason that this overview is required is that we need to know what we should expect from this artificial intelligence. We can ask artificial intelligence to create music willy-nilly at its leisure, but without any frame of reference to compare it to, our work is nullified, unable to procure music that is pleasing to the ear or that replicates a certain era. Artificial intelligence is centered around the concept of "large-scale models" and training networks of digital neurons to work in a similar function to our own brains, so perhaps we could have an artificial intelligence create random noises akin to the piano works of modern classical music, but if we desire a more enriching experience with pleasantly structured music, we must train artificial intelligence to recognize patterns in music and replicate them within its own works.

Humans pride themselves on their irreplicable creativity, but artificial intelligence has begun to get quite good at replicating music. Compared to artificial intelligences from 2020, it has become significantly easier to replicate others' music, but we still live in an era where artificial intelligence is in a grey area; sometimes we can tell that it has been used, but other times, it is harder to spot the difference (Beaumont-Thomas, 2023). So, what is keeping musical artificial intelligence from advancing further in the industry when visual art artificial intelligence and mechanical artificial intelligences have found so much success within our human society? To understand this, we need to understand the hurdles that a musical artificial intelligence would have to jump over to effectively create music that pleases the ears of the general populace. We need to dive into what makes music "pleasant" to us and confront each problem and how artificial intelligence will be able to overcome it.

Music theory is an especially hard topic to confront, because although people are likely familiar with music, music theory is a whole beast to dissect. We are mandated to build up the foundations of Western music from the ground up.

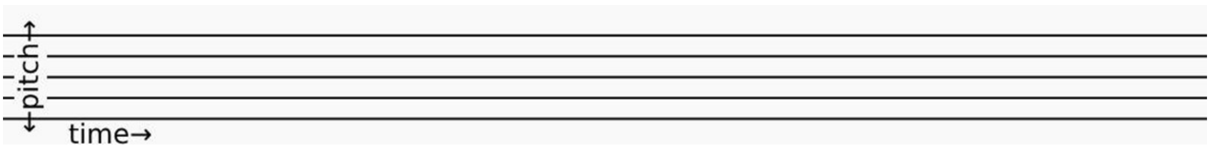

Figure 1: The two dimensions of music that are notated in sheet music.

Above is a musical staff. If we go up and down the staff, the pitch will get higher and lower respectively. If we move right on the staff, we move through time, in other words, the music plays forward.

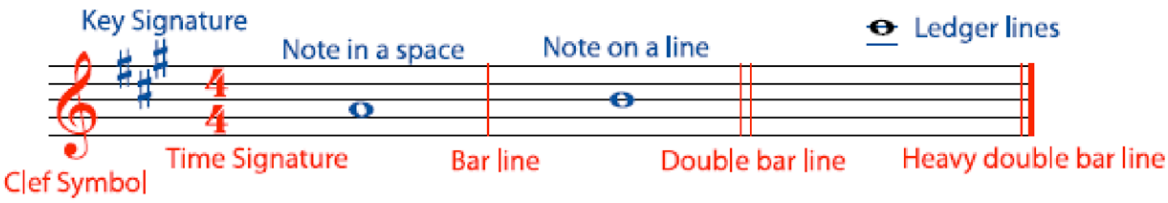

Figure 2: A display of some important symbols in common notation. Image credit: (Schmidt-Jones, 2015)

Above are some basic symbols in Western common notation. A clef symbol acts as a reference point from which we can decipher the notes on the staff and what we need to play. The key signature denotes which keys must be perpetually altered throughout a piece of music, whether they should be lowered by a half-tone or raised by a half-tone. A time signature denotes the beats in the music, a bar line denotes that the main beat, known as the downbeat, will follow it, a double bar line represents a sectional transition in a piece of music, although it is used leniently throughout Western music tradition, and a heavy double bar line denotes the end of a piece of music. We use ledger lines to demonstrate that a note is too high or too low to fit on the pitch space denoted by the staff. Concerning these elements, artificial intelligence has little trouble understanding these concepts, since there is no inherent creativity within these concepts.

The first roadblock that artificial intelligence has is with musical notes and intervals. Although music is hardwired into our brains, artificial intelligence has no idea of what sounds good and what does not. We may hear a pleasant melody that we can hum along, but to an artificial intelligence, the melody is a jumble of symbols denoting what note to play and when to play it.

Remember those notes *ut*, *re*, *mi*, *fa*, *sol*, and *la*? *Ut* slowly gets dropped in Western music solfège for *do*, partially because *ut* does not start with a consonant, rather it ends in a consonant. Meanwhile, a new note *si* was also added to the six notes. *Si* was originally ignored in Western music as it created a dissonant interval with *do*, but starting in the Baroque period, in the early 1700s, *si* was grouped in with the rest of the notes due to its strong tendency to resolve to *do*. Because it was easier to sing *ti* than *si*, *si* was replaced with *ti*, and we are presented with our modern scale consisting of the seven notes that we are largely familiar with.

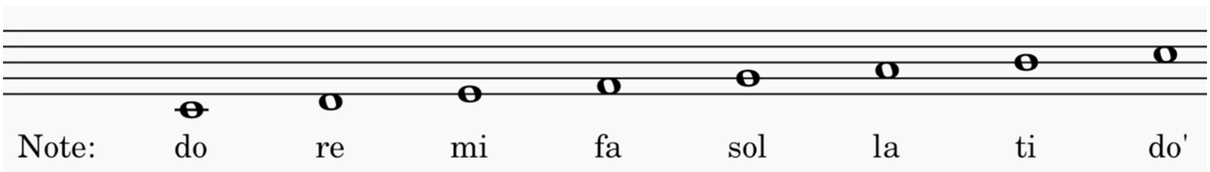

Figure 3: A display of the seven notes of the western musical scale and the *do* note an octave above middle C.)

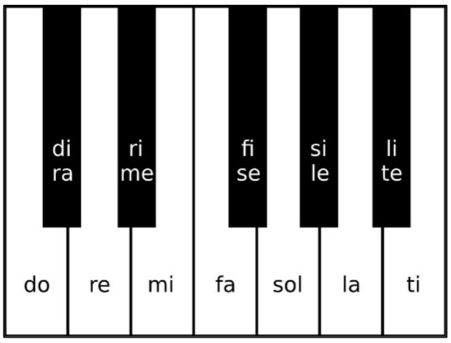

Figure 4: A display of the twelve notes in an octave of the western musical scale with notes names provided in solfège.

We also need to realize that there are some missing notes between the notes we are familiar with that we forget to mention. These are deviations from the "diatonic" system of music—the system being the seven notes we have discussed above—that are referred to in music as "chromatic." A chromatic scale features twelve notes rather than seven and is comprised of half steps—the space between a white key and a neighboring black key[3]—while a diatonic scale is comprised of a combination of half steps and whole steps—the space between a white key and a neighboring white key[4].

In order to more efficiently cover this topic, we will switch from the solfège that many are familiar with (*do*, *re*, *mi*, etc.) to letter names, which are more commonly used in musical pedagogy and performance. A corresponds to *la*, B corresponds to *ti*, C corresponds to *do*, and so on so forth.

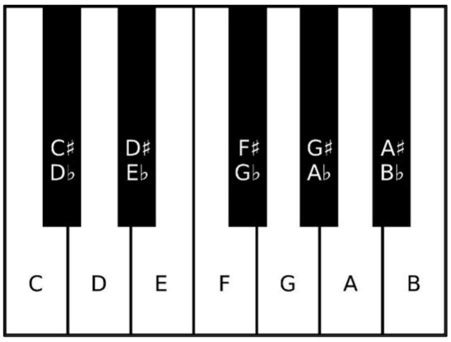

Figure 5: A display of the twelve notes in an octave of the western musical scale with notes names provided in letters.

Within a chromatic scale, we will find alterations to our seven-note scale to access the black keys. Going a half step in music is known as a "sharp," while going a half step down in music is known as a "flat." These are called "accidentals", and we use the symbols ♯ and ♭ to denote these two respectively. So C♯ represents the black key that is a half step above C, and E♭ represents the black key that is a half step below E. Looking at the diagram, you might be wondering why each black key has two distinct names. If we go up a half step from C, we get C♯, but we can also recognize that if we go down a half step from D, we get D♭, which lands us on the same note. In music, we call these two notes "enharmonic." Although they produce the same sound when played as they correspond to the same key, they are pedagogically different within music theory.

[3] I gloss over a bit of the complexity here for simplicity's sake, but half steps represent neighboring notes in the Western scale while whole steps represent notes that are two notes apart. What do I mean by this? If we arrange a keyboard so that all its notes—both white and black keys—are in order like so: C C♯ D D♯ E F F♯ G G♯ A A♯ B C, then a movement that I describe as a "half step" can be shown as a movement from a white key to a neighboring black key, a black key to a neighboring white key, or in the case of notes B, C, E, and F, a shift to their neighboring white key. The gap between B and C is a half step, and the gap between E and F is a half step, but the gap between C and D is not a half step, as there is a black key between the two notes.

[4] For a little bit of context to this annotation, you may preview [3] first. If we arrange a keyboard so that all its notes—both white and black keys—are in order like so: C C♯ D D♯ E F F♯ G G♯ A A♯ B C, then a movement that I describe as a "whole step" can be shown as a movement from a white key to a neighboring white key, a black to a neighboring black key, or in the case of notes B, C, E, and F, a shift to a non-neighboring black key. The gap between C and D is a whole step, as is the gap between C♯ and D♯. Another example of a whole step would be the gap between B♭ and C or E and F♯, but the gap between E and F is a half step, as the two notes are directly adjacent with no black key separating the two. E♭ and F♯ is also not a whole step, as the two notes have two keys between them rather than one, even though they may seem to be "neighboring black keys."

Despite how different C♯ might feel from D♭ or how differently the two notes might be treated in musical settings, in the eyes of artificial intelligence, the notes are the same. There is no difference to an artificial intelligence as to which one is used, so the two are merged within the pattern recognition algorithms of the artificial intelligence. This same phenomenon occurs with all the other black keys.

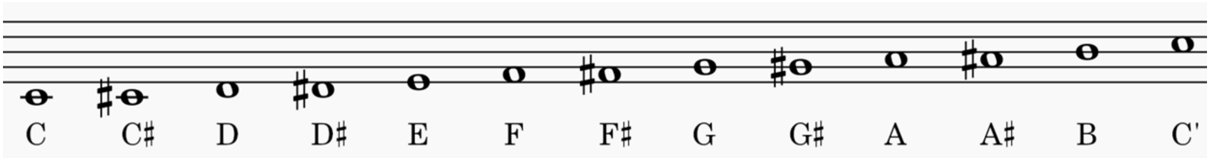

Figure 6: An ascending chromatic scale. These are the twelve notes of the Western musical scale with the addition of C', which represents the note an octave above middle C.

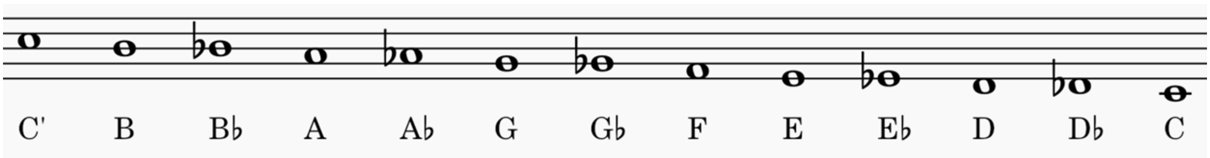

Figure 7: A descending chromatic scale. These are the twelve notes of the Western musical scale with the addition of C', which represents the note an octave above middle C. Note that the black keys in the ascending scale have been replaced with their enharmonic equivalent.

So we have a small foothold in the large world that is music theory, but all of this does not seem very difficult for an artificial intelligence to understand, so why is it so difficult for artificial intelligence to replicate human music well, and why has progress been made only so recently?

The main reason for this is that it is hard for artificial intelligence to compose music well. An artificial intelligence can sift through collections of music theory books, but when it comes time to put pen to paper, the artificial intelligence has a lot more important topics to consider, such as melody, harmony, rhythm, beat, and musical progression.

For this section, I will be using the children's song "Long, Long Ago" as it is featured in Book 1 of the Suzuki Method for Violinists. I will transpose the piece down to C major to avoid accidentals within the music.

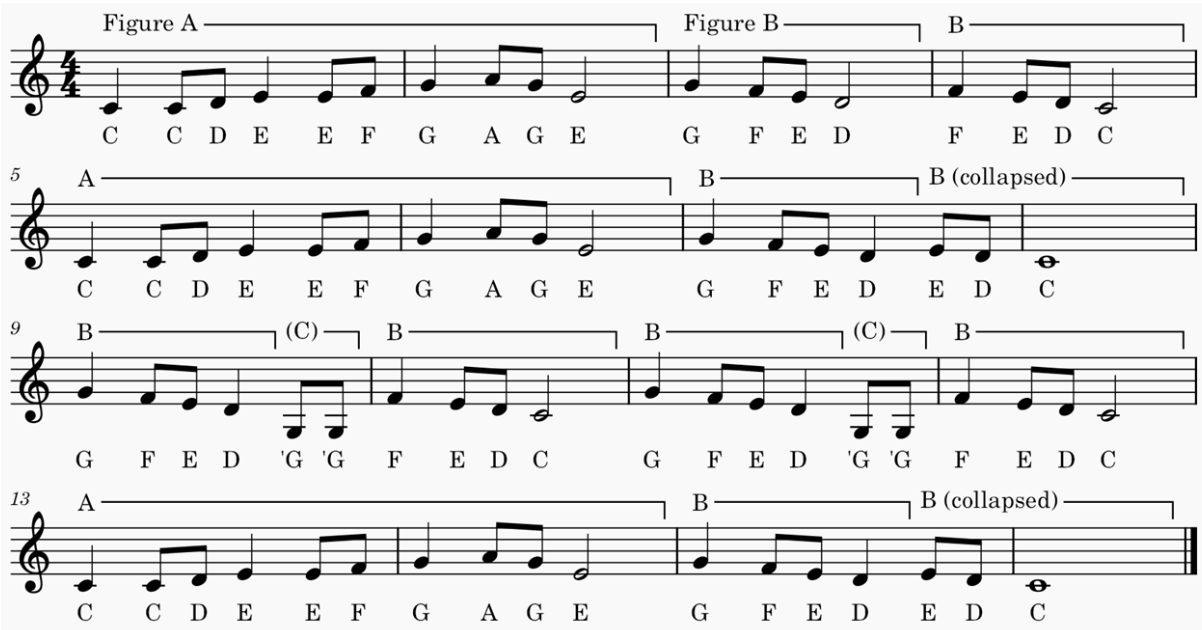

Figure 8: "Long, Long Ago" adapted from Book 1 of the Suzuki Method for Violinists (Suzuki, 2007).

As you can see from the music, we have three—and arguably even two—motifs that occur throughout the music. Motif A is the main theme of the melody. It starts with C, which we call the "tonic," as it is this note that music will revolve around. Motif A moves upwards to G, which we call the "dominant." The dominant note is always five notes above the tonic, and one very important thing to keep in mind is that while the dominant represents tension within music, the tonic represents release. Motif A continues to move up to A before it descends to E, which is a note that sits between the tension of the dominant and the resolution of the tonic. This E at the end of the phrase signifies to the listener that the music will continue.

Motif B consists of four notes played in descending order. The first time we hear Motif B, it starts on G, our dominant, and it descends to D, which is a note that represents tension that will be resolved in the next phrase. The second time we hear Motif B, the motif is shifted down by one note, and we start on F instead G. The motif descends to C, which represents the first time that the music resolves, but even as we have returned to the tonic, we still feel a sense that the music is not resolved, as the phrase starts on F, which also represents tension within our music, albeit less tension than the dominant. We will refer to F in the music as the "subdominant," a note which is four notes above the tonic and one note below the dominant.

The next line features a similar exchange between Motif A and Motif B, but this time, rather than the second occurrence of Motif B starting on the note F, which is our subdominant, it starts on the note E, which carries less tension compared to the subdominant. Motif B is clipped to only three notes this time, and this motif ends on C, our tonic and our resolution. Although we have returned to our tonic, we still feel that music must continue, and this is because of the form of music. Our brains love music that follows a ternary form, that is, ABA, and this is because we feel like we are following a kind of small story. A represents the exposition, B represents the journey, and A represents a return home. This is almost a musical hero's journey of sorts, and we can find this in many common songs like "Twinkle Twinkle Little Star" or "Für Elise" or "Mary Had A Little Lamb."

In the next line, we enter our "B section," this time starting with Motif B rather than the Motif A that we usually start with. The motif begins on the dominant G and ends on D, signifying tension, and this tension is increased with Motif C[5], which contains two strikes of the dominant G in the lower register. We have another occurrence of Motif B that starts on our subdominant F, so again, although we resolve to the tonic C, we still have a feeling of continuation and tension that pushes us into a repetition of the same two bars. The same BCB motif occurs once again, and now that tension has been built up in this B section, it can be released with a return of our main theme.

The main theme returns in the next line, with our previous Motif A arising again to let us know we are in the A section of the piece now. We have essentially returned to where we started, and we can see that our fourth line is merely a copy of the second line. This time, as the ABA ternary pattern is fulfilled, the music feels fully resolved, and the piece ends.

[5]Despite my treating it as a motif, Motif C is extremely simple and could merely be treated as a transition between two occurrences of Motif B. This is why there are arguably only two motifs in this piece and not three.

This analysis of a piece is very natural for humans because we innately feel the tension rising during the B section and the expectation of the music's continuation, but artificial intelligence does not have this feeling, and this has been one of the biggest hurdles in the progression of artificial intelligence with regards to music: artificial intelligence has a hard time with concepts in music that we innately process and understand. We are so familiar with music in our day-to-day lives that it is nearly impossible for us to articulate to a machine how to create music that sounds pleasant and resolved. This melody feels "right" to us, while other melodies feel "wrong," and although we can analyze why we feel this way, artificial intelligence has a much harder time with these more subjective categories of "right music" and "wrong music," and thus it has a much harder time creating a feasible melody to the music.

After the melody comes harmony. We have long strayed from the unison Gregorian chants of the Pre-Renaissance movement, so we ascribe harmonies to our melodies to provide more texture to the music and cement the concepts of tension and release within our music.

Although there are many ways that we could harmonize any melody, as harmony is largely a subjective field, I will be sticking to a strict harmonization based on classical European musical standards. I could in theory harmonize the melody to give it a dark foreboding sound by accompanying the melody with minor chords or give it a jazzy sound by accompanying the melody with more extended and harmonically rich chords, but for the purpose of this exercise, it is more important that we understand the difficulties that harmonization poses to artificial intelligence. Demonstrated right here in this paragraph is one problem that artificial intelligence is forced to tackle: there are several different ways to harmonize a melody that are all acceptable, representative of the "pleasing to the ear" sounds that we enjoy in our music.

Figure 9: "Long, Long Ago" adapted with a simple harmonization from Book 1 of the Suzuki Method for Violinists (Suzuki, 2007).

Here we have a basic harmonization of "Long, Long Ago." We can see that there are only two chords present within the music, a C major chord—denoted with the symbol C and consisting of the notes C, E, and G—and a G dominant seventh chord—denoted with the symbol G7 and consisting of the notes G, B, D, and F. Anything we notice about C and G? They represent the tonic and the dominant of the piece respectively. When we harmonize using a C tonic chord, we signal a release of tension in the music, whereas when we harmonize using a G dominant chord, we signal a build-up of tension in the music.

We can see that our piece starts with a C major chord. This establishes the key of the piece, more basically, the note that the music will be revolving around for the piece. We have two measures of the C major chord before we get our first G dominant chord. The chord signals a new tension in the music that wants to resolve, which we get in the next measure.

The same chord progression is echoed in the next line, with two C major chords followed by a G dominant chord followed by a C major chord, but the third line is where we see more harmonic development occurring. Coming from a C major chord, we have a feeling of resolve, so when we are met with a new G dominant chord, tension is introduced to the piece. Even as this G dominant chord resolves to a C major chord, we do not feel a true resolution, as the E in the C major chord clashes dissonantly with the F in the melody, creating a minor ninth, an interval that is displeasing to the ear[6]. This same progression is repeated, with the tension of the G dominant chord resolving to the C major chord and another clash between the E and the F arising again.

We finally escape the tension in the last line, which follows the same chord progression of the first two lines.

[6]This stems from the just intonation tuning systems establish by Pythagoras. Pythagoras theorized that all notes could be tuned through pure whole number ratios, and generally, the smaller the whole numbers involved in the ratio of an interval (the distance between two notes), the nicer the interval will sound to us. We love intervals like the octave (ratio of 2:1), the perfect fifth (ratio of 3:2), and the perfect fourth (ratio of 4:3) because of their simple ratios, but some intervals sound "worse" to us subjectively because of their more complicated ratios, such as the minor second (ratio of 16:15), the tritone (ratio of 64:45), and the minor ninth (ratio of 32:15).

Figure 10: "Long, Long Ago" adapted with an alternative harmonization from Book 1 of the Suzuki Method for Violinists (Suzuki, 2007).

Harmony is difficult even for human brains to grasp, since for any given melody, there are almost an infinite number of different choices that a composer could make with regards to what chord comes next, and although there is some structure to harmony, it is a choice that is up to composer, which is difficult for artificial intelligence that needs to be told what to do and cannot decide its next move creatively.

To the left is an image of an alternative way that I could have harmonized "Long, Long Ago" that utilizes more complicated chords. Although the piece features the same melody, the mood is completely changed by introduction of different chords, leaving a cheerful children's song sounding darker and more introspective. This exemplifies how difficult harmonization can be, especially for artificial intelligence, since it cannot ascribe meaning to these chords and understand the way we do that a chord progression sounds good or not.

Until now, we have only seen complex analysis of different pieces, but how does artificial intelligence and machine learning actually fare in musical composition? Luckily for us, there are already some projects that aimed to achieve just that.

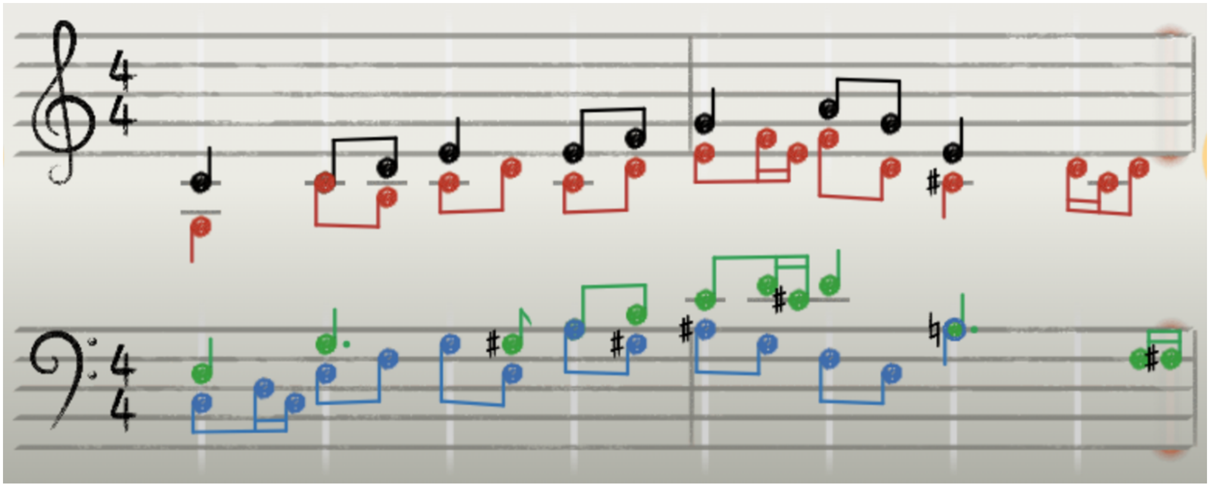

Figure 11: A result from a 2019 Google Doodle that featured Google's Magenta music model being trained off of Johann Sebastian Bach's pieces to create fugal works.

In 2019, Google released its first ever Google Doodle that was powered by artificial intelligence. The user was prompted to write a soprano line for a four-part fugue, and then an artificial intelligence would take that soprano and attempt to harmonize it based on its training with a massive catalogue of Bach's pieces (Hannah-Murphy, 2019). Above is the result I got when I put the melody for "Long, Long Ago" into the Doodle. Below I will convert this into a cleaner engraving, and we will analyze what the artificial intelligence did right and what it did wrong.

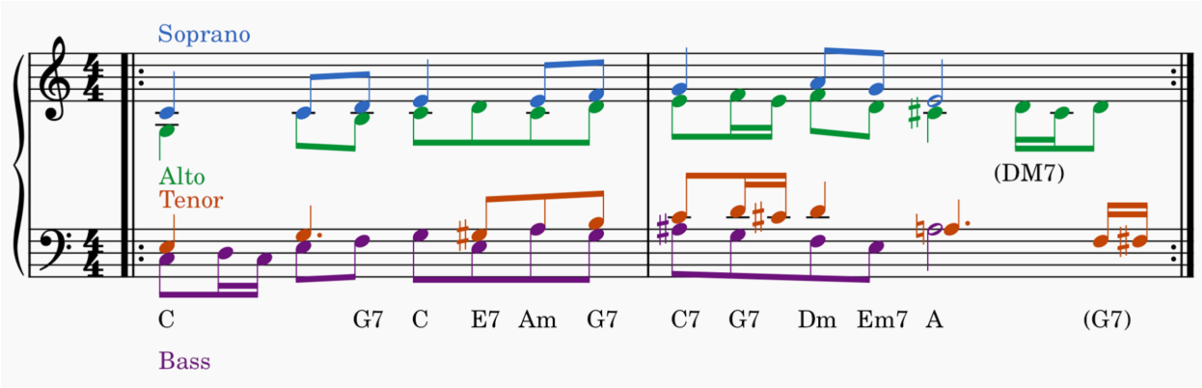

Figure 12: "Long, Long Ago" adapted by Google's Johann Sebastian Bach Google Doodle. Each color corresponds with a voice in the counter point, and the key is shown below: ■ Soprano ■ Alto ■ Tenor ■ Bass

This is a footnote regarding the image: [7].

The progression of the fugal writing is not as bad as one might expect, but all the progression is bunched up together, which makes the progression of the fugue hard to follow as a listener. One important cornerstone of fugal writing is the concept of "separate but whole": each voice is independent from the others, but the voices come together to create a whole musical work. The individual voices are mediocre. Although each voice exhibits patterns reminiscent of a melodic line, there is not a particular structure for the voices, and they do not sound particularly pleasing individually. As a whole, the fugue does not flow very well either. Because of the rapid changes in chords, the work makes it difficult for a listener to develop a deeper melodical understanding through only the ear.

Something interesting to note is the implied DM^7 and G^7 chords at the end. In music theory, this chord progression is referred to as a "two-five-one," written in Roman numerals as ii-V-I[8]. It is not immediately clear why the artificial intelligence chose to add these chords into the piece, but it seems to be due to the fact that the Google Doodle repeats the two bars of music in perpetuity, meaning that the chords have to return to the C major chord that start the fugue, but in order to effectively achieve this, the Doodle is forced to add in two out-of-place chords in order to facilitate this transition. This seems to be one problem that exacerbated the flaws with the work.

This Google Doodle was released in 2019, only about four to five years ago at the time of my writing, so what has changed? How has artificial intelligence become advanced enough that it threatens to take down the music industry by generating personalized

[7] For the C7 chord at the start of the second measure, I chose to interpret the A♯ as its enharmonic equivalent of B♭. In this way we can ascribe a label of C7 to the chord, which is much easier to work with compared to the alternative of A♯dim7♭♭3.

[8] ii-V-I represents the chords of the progression, which we can derive from their symbols. Since we are in the key of C, C represents I, D represents II, etc. Starting with ii, we count to the second degree of the scale, which is D, and we use a minor chord, which is represented by the lowercase Roman numeral being used in place of the uppercase Roman numeral. V represents the dominant, as it is five notes above the tonic of C. Thus, our V chord would be a G chord. The uppercase Roman numeral represents that this is a major chord. Finally, I represents the tonic, which is C. Thus, our I chord would be a C chord. The uppercase Roman numeral represents that this is a major chord. One thing we must note is that the ii chord in the Doodle progression is a II major chord, which differentiates it somewhat from the ii-V-I chord progression.

music for us?

The main reason for improvements in musical artificial intelligence is the advances of artificial neural networks, more specifically, dense neural networks.

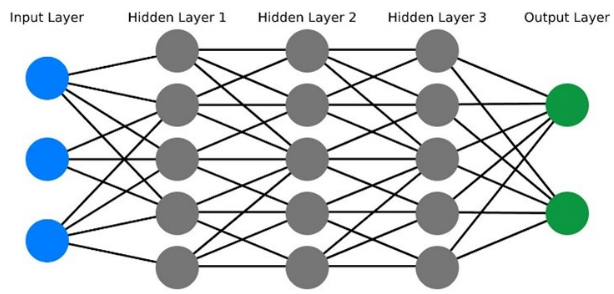

Figure 13: An artificial neural network with three input nodes, three hidden layers with five nodes each, and two output nodes.

An artificial neural network (ANN) is a branch of machine learning that uses nodes and connections between them to mimic the neural structure of animal brains. An ANN consists of an input layer, one or more hidden layers, and an output layer. For my example, I will be considering an ANN with three hidden layers. First, the inputs are given to the ANN, and they are provided to the hidden layers. The first hidden layer takes the input received and processes it into data for the next hidden layer, which it sends to the second layer for it to process. The second layer processes the inputs it receives into data for the third hidden layer, and the third hidden layer sends the data to the output layer. The ideology behind an ANN is that, similarly to an animal brain, we take in perceived stimuli, which triggers a neural pathway in our brain that allows us to act considering said stimuli (Hardesty 2017). Another important way that ANNs mimic animal brains is that each connection between each layer has a weight between zero and one, with zero representing none of the data from a node being sent through the connection, and one representing all the data from a node being sent through the connection. This weight is the key to an ANN, and similarly, our own brains have neurons that send weighted electrical signals to the next neuron. This is why ANNs have become such an important part of modern machine learning: their ability to effectively replicate the connections that we have in our own brain allow it to be trained to fulfill its goals (Hardesty 2017).

Although ANNs have all these capabilities, how do we know what weights should be applied to each connection? Which connections should be made? How do we get the results that we want? When we train ANNs, we first start off with a set of nodes with connections that are all weighted the same. The ANN creates an output, and a backpropagation training algorithm processes the output and updates the weights of the ANN. The ANN creates another output, and the backpropagation algorithm processes the output again and updates the weights of the ANN. This process continues, a back-and-forth between the ANN and the backpropagation algorithm, until the ANN can fulfill its goal decently well. This constant feedback loop is also seen in nature through different biological processes such as homeostasis and natural selection.

Usually, we can continue this cycle until ANNs with a 5% error rate or below remain, as this marks a statistically significant difference from a non-trained ANN. For more important applications, the ANN can continue to be refined to smaller and smaller error rates. We can then use these ANNs for applications ranging from spellchecking to aircraft component fault detectors to warehouse management and transport to self-driving automobile to the topic of this paper, musical generation.

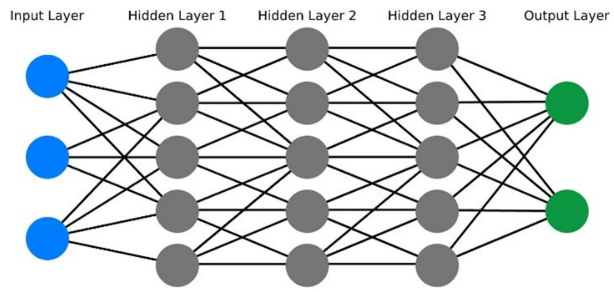

Figure 14: A dense neural network with three input nodes, three hidden layers with five nodes each, and two output nodes.

Music generation uses a subset of ANNs known as dense neural networks, or DNNs. This means that each node in one layer is connected to all the subsequent nodes in the next layer. DNNs have been used for a long time throughout machine learning work, but new advancements are allowing for more nodes, more connections, and more layers, which has allowed machine learning algorithms to improve drastically.

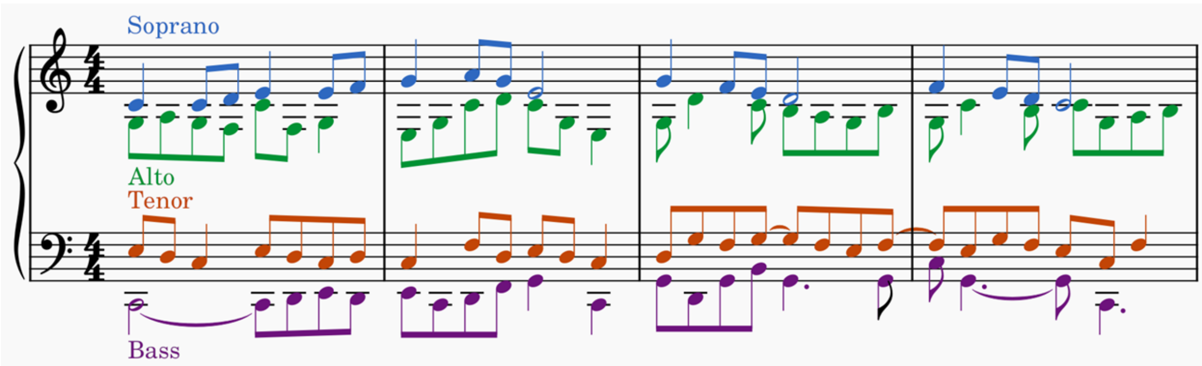

Figure 15: "Long, Long Ago" adapted by Google's Magenta Project. Each color corresponds with a voice in the counter point, and the key is shown below:
■ Soprano ■ Alto ■ Tenor ■ Bass

If we plug our melody of "Long, Long Ago" back into the Magenta Project again, we get the result shown above. This time, each melody can stand alone by itself, and harmonically, each melody is combined to make a whole. Each voice can be heard distinctly from the other voices, and the harmonic progression of the entire fugal work is much more aligned to the original chord progression that we had laid out before.

From this point in time, artificial intelligence can only get better as computers are able to process more musical inputs, push that through more layers and nodes, and refine itself through repeated error correction. It is likely that artificial intelligence will be completely indistinguishable from human-made creations, and the creativity which we believed to be so special may be rendered commonplace with the advent of creative artificial intelligences. Even the small fugal counterpoint present in the Magenta Project is hard to tell apart from an amateur contrapuntal composer, and so in some cases, the gap between artificial intelligence and humans is already closed creatively. Only time will tell us of the implications of these improvements, but as of now, our unique creativity may not be as unique as we previously believed.

XII.II Extended Bibliography

Arkenberg, R. (2002, October). *Music in the Renaissance.* The Met. https://www.metmuseum.org/

Beaumont-Thomas, B. (2023, April 19). *We soon won't tell the difference between AI and human music – so can pop survive?* The Guardian. https://www.theguardian.com/music/2023/apr human-music-pop-drake-kanye-west-the-weeknd

Blume, F. (1970). *Classic and romantic music; a comprehensive survey.* United Kingdom: W. W. Norton.

Davies, S. (1991). *Definitions of Art.* United Kingdom: Cornell University Press.

Fux, J. J. (1725). Præfatio ad Lectorem. In *Gradus ad Parnassum* (pp. 11–13). forward, Typis Joannis Petri Van Ghelen.

Hannah-Murphy, L. (2019, March 21). *Honoring J.S. Bach with our first AI-powered Doodle.* Google. https://blog.google/technology/ai/honoring-js-bach-our-first-ai-powered-doodle/

Hardesty, L. (2017, April 14). *Explained: Neural networks.* MIT News. https://news.mit.edu/2017/ neural-networks-deep-learning-0414

Hochelaga. (2020, August 14). *The Oldest (Known) Song of All Time.* YouTube. https://youtu.be/KElPnD-dbkk?si=j6UklEAgNhZ4NLpt

Kamien, R. (2006). *Music: An Appreciation.* United States: McGraw-Hill.

Mackay, A., & Romanec, C. (2016, December 7). *Baroque Guide.* Tafelmusik. https://web.archive.org/web/20161207213948/http://www.tafelmusik.org/sites/all/themes/tafelmusik

Mahendra, S. (2023, June 13). *Can People Tell the Difference Between Music Created by AI and People?* Artificial Intelligence +. https://www.aiplusinfo.com/blog/can-people-tell-the-difference-between-music-created-by-ai-and-people/

McDermott, A. (2021, October 27). *What was the first "art"? How would we know?* PNAS. https://www.pnas.org/doi/pdf/10.1073/pnas.2117561118

Neely, A. (2019, August 2). *Why the Katy Perry/Flame lawsuit makes no sense.* YouTube. https://youtu.be/0ytoUuO-qvg?si=EgSqwQ_jIHXd4lCZ

Neely, A. (2022, March 6). *Did Dua Lipa ACTUALLY Plagiarize Levitating?.* YouTube. https://youtu.be/HnA1QmZvSNs?si=0hDQfh-5YyBZN3bk

Pleasants, H. (1955). *The Agony of Modern Music.* United States: Simon and Schuster.

Proceedings of the Musical Association. (1893). United Kingdom: Stanley Lucas, Weber & Company.

Rizzi, S. (2023, June 5). *Hear the world's oldest instrument, the 50,000 year old neanderthal flute.* Classic FM. https://www.classicfm.com/discover-music/instruments/flute/worlds-oldest-instrument-neanderthal-flute/

Rogers, R. (2023, January 21). *How to Spot AI-Generated Art, According to Artists.* Wired. https://www.wired.com/story/how-to-spot-generative-ai-art-according-to-artists/

Schmidt-Jones, C. (2015, February). *Understanding Basic Music Theory.* Open Textbooks for Hong Kong. https://www.opentextbooks.org.hk/system/files/export/2/2180/pdf/Understand

Schwartz, E., Godfrey, D. (1993). *Music since 1945: issues, materials, and literature.* United Kingdom: Schirmer Books.

Suzuki, S., Preucil, W. (2007). *Suzuki Violin School: Volume 1: Violin Part.* United Kingdom: Summy-Birchard, Incorporated.

Truscott, H. (1961). Form in Romantic Music. In *Studies in Romanticism* (Vol. I, pp. 29–39). essay, Boston University Graduate School.

Tyrrell, J., & Sadie, S. (2003). United Kingdom: Groves Dictionaries, Incorporated.

Wildridge, J. (2022, October 15). *Characteristics of Romantic Era Music.* CMUSE. https://www.cmuse.org/characteristics-of-romantic-era-music/