Tunes of Innovation: A Framework for Analyzing Artistic Style in Musical Compositions

Aarav Gadkar[1], Adam J. Strawbridge[2]

[1] Carlmont High School, Belmont, California
[2] University of California, Davis, Davis, California

Student Author: Aarav Gadkar, Carlmont High School, Belmont, California

ABSTRACT

In the modern technological revolution, artificial intelligence (AI) has developed capabilities to generate new musical compositions by using algorithms trained on existing music. Despite advancements in this field, the full potential of AI in music remains unexplored. We provide a framework to analyze specific stylistic aspects of an artist's musical style including structure, instrumentation, and melodic and rhythmic style. As a case study, we used songs by the pop music star Ed Sheeran's to define a detailed profile of an Ed Sheeran composition using an artificial intelligence program and assess its ability to capture the artists' stylistic features. We use MuseNet as an example, but the framework is generalizable to any model. Popular Ed Sheeran songs were analyzed using concepts of music theory. After identifying similarities between these compositions, we compared our findings with the performance of MuseNet to determine what aspects of the style are captured. The study demonstrated that generative AI models used for music compositions are quite accurate in predicting artistic styles such as composition structure and melody. This reflected the rigorous training of these models into embeddings related to these features. For characteristics such as instrumentation, the models captured the styles accurately only if the original training-set included the instrument in question, but accuracy dropped for esoteric instruments. In conclusion, AI models were effective in emulating artistic styles, but still had room for further retention of a human touch to the music that invokes in the listener emotion and sentiment in the overall composition.

Introduction

With advancements in computer hardware and software, capabilities of artificial intelligence (AI) programs have grown at a tremendous pace in the last decade. Generative AI uses AI models on available data to create new content. One field that has greatly benefitted from generative AI is music. It has emerged as a tool for writing original lyrics and creating novel musical compositions and productions. It has transformed both the music industry and the process involved in the creation of music. Multiple AI models, such as AIVA (AVIA, n.d.), Magenta (Magenta, n.d.), and MuseNet (OpenAI, MuseNet n.d.), have been developed to analyze musical data to generate completely new and original musical compositions. These technologies have evolved over the last few years, from the expansion of digital synthesis and technological standardization with Musical Instrument Digital Interface (MIDI), to the incorporation of AI and machine learning in music composition and performance. Earlier works such as "Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys" by Krumhansl and Kessler (Krumhansl, 1982) have become key inputs to an AI model with a sense of tonality and cognition.

However, it is important to understand the capabilities as well as the current limitations of AI to generate music to improve the future of AI generated music. In this paper, we provide a framework for developing a stylistic profile of a particular artist's compositions. By observing an AI's capability to recreate features that are common in an artist's music, such as harmonies, rhythms, and instrumentation, we assess the overall capability of it to emulate

an artist's style. This framework helps us identify gaps and areas of improvement for these generative AI models to further emulate the entire spectrum of musical experience that an audience receives listening to music of human artists. Further, understanding an AI's capacity to interpret different elements of style helps with retention of genuine emotion and artistic style of this music, which can help grow appreciation for modern developments in the music industry. With this focus in mind, this work attempts to address the following specific objectives:

(1) To develop a standard framework to analyze the capabilities of AI software generated music to capture specific technical and stylistic aspects of an artist's music such as composition structure, instrumentation, and melodic and rhythmic style.
(2) Can this framework help us identify to what extent AI-generated music convey genuine human emotion and artistic styles compared to those of modern artists?

After developing a framework, we analyze the capabilities of MuseNet as the AI tool of choice to capture and reproduce the artistic style of the music compositions of Ed Sheeran. The analyses in this work is limited to MuseNet and the compositions selected in the analyses, but the conclusions can be generalized by additional research on compositions of other artists and evaluating other AI tools.

Methods

Developing the framework

The framework for developing a stylistic profile of an artist includes a detailed analysis of songs by the artist. The transcriptions for these compositions are generated manually via concepts of music theory. The core parameters of the music from these transcriptions, such as instrumentation, style, structure are obtained to serve as the input data for AI models.

The primary analysis of each song is divided into: (i) Broad analysis that includes analyzing each song's structural patterns, finding patterns in the instrumentation, extracting key melodic and harmonic schema, and identifying key rhythmic styles, and (ii) Analysis of the iterative unit where the fundamental iterative unit of the song is analyzed. It consists of transcribing the general beats of the song onto sheet music and analyzing percussive elements to indicate the songs' atmosphere. These steps are described below in detail. Following these primary analyses, a second order analysis is performed for specific patterns and features of similarity across all compositions. This gives us the general style of Ed Sheeran's music. The AI model is examined for its suitability to replicate this style, and is the focus of the results section on this paper.

Broad Analysis

Structure: For this project specifically, structure is defined as the organizational properties of a song. The most common structure in pop music is: (Verse 1, Chorus, Verse 2, Chorus, Bridge, Chorus); a technical notation used to shorten this long string of structural parts is ABABCB, with each letter representing the corresponding part. This is a lyrical structure; however, in modern pop music, it also translates into a musical structure. It is important to examine because it will allow for a coherent structure that the AI can replicate instead of it being one large composition without any set structure.
Verse (A): The verse is typically where the storytelling occurs, whether instrumental or lyrical. They usually are lyrically unique from each other and oftentimes melodically unique as well. The verse is crucial to establishing the mood and musical context of the song and leads up to the attention-catching chorus.
Chorus (B): The chorus is the segment of the song that repeats throughout the majority of the song. It is supposed to be the most memorable and intriguing portion of the song and is often used to convey the central message of the song. In modern pop music, the chorus is usually the most intense in terms of instrumentation and harmony use, giving it a distinct feeling from that of the verses.
Bridge (C): The bridge is used as a contrasting factor from the rest of the song, and introduces further variety, straying away from the instrumental and harmonic patterns of the verses or chorus. It often smoothly leads into the final chorus.

Instrumentation: The analysis of instrumentation involves determining specific instruments and vocals, and observing patterns and relationships that is drawn between them. There are a few different parts to this analysis. The first step is identifying and listing the different instruments used in a piece, including acoustic instruments (i.e. guitar, piano), electronic/digital instruments (i.e. synthesizers, electric bass), percussion (i.e. drum sets, body percussion), and vocals. We then identify the individual role of each instrument. It is an observation of how each instrument contributes to the texture of the piece. This can be in the form of rhythm, harmony, melody, dynamics, or accentuation. Finally, we identify how the overall instrumentation of the music changes throughout the piece. Some examples of this sort of analysis are the addition or removal of certain instrumental parts between different structural parts of the song, changes in instrument assignments, or changes in intensity and dynamics.

Melody and Harmony: To analyze melody and harmony, we observe the arrangement and combination of pitches at a given moment in the piece. Melodic analysis is simply addressing how the melody of the song changes throughout each section of the piece, and how that contributes to the style of the artist. For instance, one may say that a melody containing short, staccato phrases gives the piece an upbeat and rhythmic motif. Harmonic analysis includes observing the textural patterns of a given section of the piece. We observe the patterns of the harmonic layering and how each instrument contributes to the harmonic schema of the piece. We define the patterns of common chord progressions using tonic, dominant, and subdominant functions. We also include a general analysis of the common patterns of dynamic and layering changes.

Rhythm: To analyze the rhythmic style of a song, we determine key features of the general beat and tempo of the song. This includes determining patterns of rhythmic syncopation, accentuation and articulation, development of the fundamental tempo (e.g. beats per minute) throughout a piece, and instrumental contributions to the rhythmic style. Syncopation refers to the consistency of rhythms in the harmonic layering. For example, the percussion and bassline could syncopate by following the same rhythmic pattern of emphasizing the downbeats of the rhythm. Accentuation and articulation refer to which beats are emphasized at a given moment in the piece. They create rhythmic patterns by emphasizing upbeats and downbeats in the fundamental beat of the song. The development of the fundamental tempo throughout a piece simply shows the patterns of changing rhythms. Does the tempo slow down or speed up between different sections of the song? To answer this question, we observe the beats per minute of each section. In instrumental analysis of the rhythm, we observe how each instrument is contributing to the rhythm. We extract common patterns of instruments that provide the fundamental tempo (e.g. the bassline provides the main rhythm, while other instruments contribute to the rhythmic layering).

Analysis of the Fundamental Iterative Unit

The fundamental iterative unit, or "the beat", of the song is the key instrumental patterns and rhythms that build the foundation of a piece. To analyze this unit, we transcribe the pitches and rhythms of all the fundamental instruments and harmonies onto sheet music using notation software. This includes observing the key, time signature, dynamics and instrumental layering of the song.

Artificial Intelligence Model: MuseNet

MuseNet is a generative AI for original musical compositions through the analysis of Musical Instrument Digital Interface (MIDI) file patterns, which contain elements of harmony, rhythm, and style. From a collection of online datasets of MIDI files, it predicts the next note in a given sequence of notes that fits the prompt or sequence of notes given as input. Each combination of notes is given a token, a sort of placeholder that it can reuse and examine throughout its composition, forming chords, harmonies, and other stylistic elements (OpenAI, n.d. *"*Composer and instrument tokens"*) Sp*ecifically, there are multiple types of tokens that help the AI understand the general musical components of the sound that is being analyzed at a specific instance. For example, a certain pitch token will identify the pitch of a note that is being analyzed by the AI.

To simplify musical patterns, MuseNet focuses on the starting point of each note, ignoring the duration, allowing it to easily predict a subsequent note in a sequence. The system uses a process known as Pair Byte

Encoding, assigning tokens to each note's start time and identifying the patterns between the different tokens (OpenAI. n.d. "*Data set").* MuseNet has further been trained to examine the music through the beats of the music and through actual seconds. This is how it implements different time signatures and rhythms in its generated music.

During the training of MuseNet, different strategies are implemented to include more musical elements to better address prompts and generate fundamentally diverse compositions. The AI is given the ability to transpose music, giving it the ability to change the key of the song, allowing the AI to compose music in different key signatures and fit the range of the prompted instruments. It also learned how to integrate dynamic patterns and general volume schema to further enable the AI's composition to fit the prompt given. Finally, the model is able to rearrange tokens in the pattern to experiment with more complex chord progressions and harmonies.

Other structural and musical elements of the style are represented by embeddings (OpenAI. n.d. "*Embeddings"*). There are four major sorts of embeddings, all used to capture as much of the style as possible. Positional embeddings reflect a token's position relative to the rest of the composition. Timing embeddings keep track of the time to set the rhythm and tempo of the composition. Note embeddings capture the general pitches of each token. Structural embeddings divide compositions sectionally, allowing the composition to fit structural parameters of the prompt.

Results

Piece Analyses of Ed Sheeran's compositions

This work primarily focuses on technical and musical characteristics of the pieces in order to obtain a more objective generalization of an artist's music. We do not analyze lyrical patterns that are typically addressed with a different set of AI tools and may be a subject for future analysis. For example, Sheeran's verses often contain deep introspective stories and experiences, creating an emotional and expressive mood in his lyrical style. We focused on the following compositions to look at Ed Sheeran's music as a case study: *Castle on the Hill*, *Perfect*, *Overpass Graffiti*, and *Photograph*. The transcriptions for these compositions are generated manually on MuseScore (MuseScore, n.d.) by using deep understanding of music theory (Figure 1). The core parameters of the music from these transcriptions, such as instrumentation, style, structure are obtained to serve as the input data for the AI models.

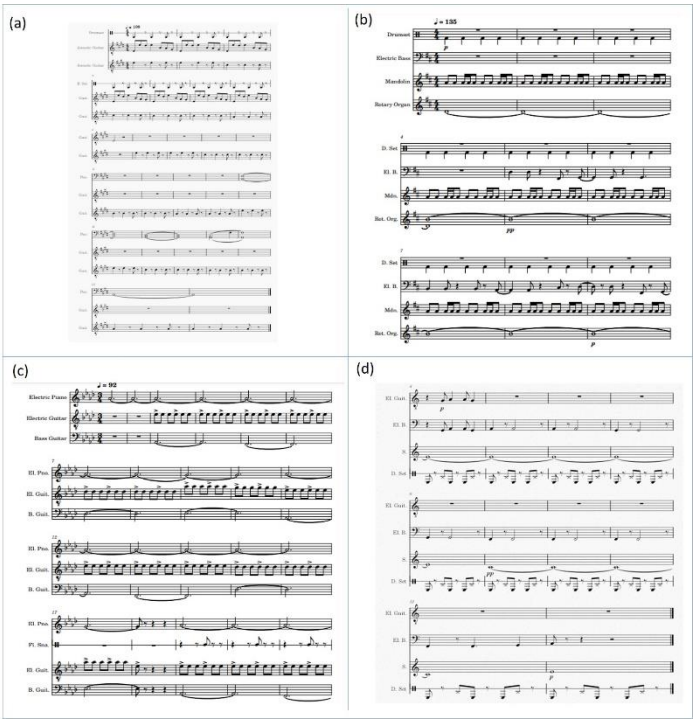

Figure 1: Transcripts of four of Ed Sheeran's compositions developed by a theoretical understanding of music compositions; the four compositions used in this work are (a) Photograph, (b) Castle on the Hill, (c) Perfect, (d) Overpass Graffiti

Structure: Sheeran's music commonly adheres to the ABABCB pattern, with very few exceptions such as *Don't* and *Sing*. He often displays a consistent style in his structural organization, with each verse and chorus being similar lengths throughout the song (no more than five seconds of variation). For example, the length of the two verses in *Castle on the Hill* measure 28 and 29 seconds. In *Overpass Graffiti*, they measure 42 and 39 seconds. Sheeran also commonly utilizes the pre-chorus to transition from verse to chorus.

Instrumentation: Sheeran also introduces variation in the instrumentation and foundations between different segments of each piece. Choruses implement more complex harmonies and layered instrumentation, and usually a louder sound, than the verse and bridge, giving it a distinct upbeat motif. Bridges often contain less harmonic and instrumental layering than the chorus, simplifying the transition between choruses. Further, the vocal melody of the piece comes in during the same section as the bass, usually being the first verse.

Melody: Specific harmonic and instrumental schema can be analyzed using the previously introduced transcriptions of the pieces. Many pedal tones usually hover around the tonic and dominant notes (often within two semitones) in the scale. These pedal tones maintain a fundamental pitch for the song throughout its period. For instance, the electric piano in *Perfect* produces a pedal tone that hovers within one semitone of the tonic (between A♭ and G in the key A♭ major) (Figure 2).

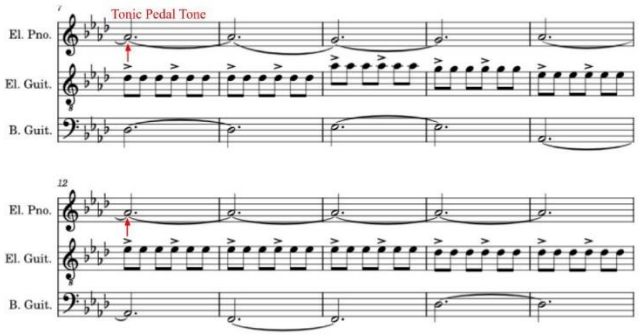

Figure 2: The instrumental pattern at the start in *Perfect* shows a Tonic Pedal tone that is captured by MuseNet based on the timing and position embeddings.

Rhythm: The elements of the beat are usually consistent throughout the entire piece, and the rhythm is usually given by a constant and repetitive acoustic pattern. In *Castle on the Hill*, the mandolin provides a fundamental rhythm that continues throughout most of the song (Figure 3).

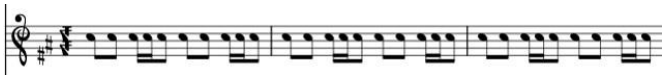

Figure 3: The Mandolin rhythm in *Castle on the Hill* that provides the fundamental tempo is captured by MuseNet based on the position and structure embeddings.

The different elements of the beats, and often the vocal melodies, are syncopated together. In *Overpass Graffiti*, the percussion, bass, and guitar are syncopated during the initial introduction of the bassline, exhibiting a downbeat, upbeat, upbeat, downbeat pattern (Figure 4).

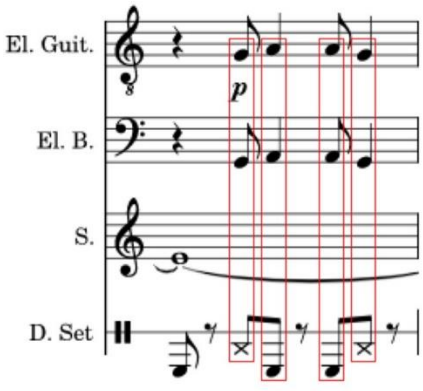

Figure 4: *Overpass Graffiti* shows the downbeat, upbeat, upbeat, downbeat pattern of percussion and harmonies that are a common feature of Ed Sheeran's compositions.

In Ed Sheeran's music, the fundamental sound of the music is set with an acoustic instrument (e.g. acoustic guitar, mandolin) and the harmonies are introduced in the bassline and other digital synthesizers. Rests are usually implemented in areas to emphasize upbeats. Such can be represented in the bass line of *Castle on the Hill* (Figure 5 Top) and the finger snaps in *Perfect* (Figure 5 Bottom).

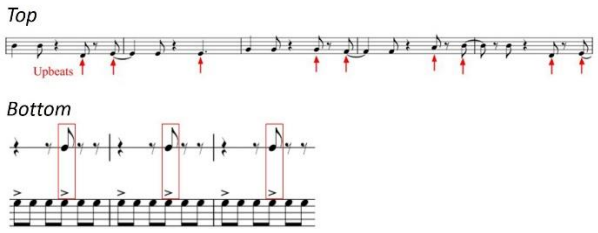

Figure 5: *Top*: Upbeats in the bassline are emphasized in *Castle on the Hill; Bottom:* Finger snaps are on the accented upbeats in *Perfect*. These features show that the usage of a strong rhythmic pulse and thoughtful rests are common in Ed Sheeran's music.

Analysis of MuseNet to capture an artist's style

We analyzed which stylistic schema observed in the second order analysis fits well into the training and functions of the AI model. Going through each of the four segments of analyses: structure, instrumentation, melody and harmony, and rhythm, we assessed how accurately the AI can mimic each one.

Structure: MuseNet can replicate the structure of most of Sheeran's music. Its ability to incorporate structural embeddings into its compositions allows it to accurately present the ABABCB structural pattern. Using the structural embedding, the AI can accurately comprehend the structure in a way that humans understand the placements of the verses, choruses, and bridges in a piece of music. The structural embeddings will also allow the AI to effectively use pre-choruses to create transitions between the different structural components. In general, structural embeddings will allow the AI to implement different musical schema in each component and keep them consistent for repetitions of verses and choruses. Therefore, MuseNet can accurately present Sheeran's structural style in its compositions.

Instrumentation: MuseNet can also identify and implement the general instrumentation of Sheeran's style. It can identify differences between acoustic and electronic sounds as a part of its training, and it can also replicate the tonality and sound of the instruments on a high level because of its training in transposition and dynamics, thus allowing the AI to compose a piece using the largely acoustic foundation of Sheeran's music with layered digital instruments and percussion. However, MuseNet's AI is not generally focused on a large instrumental database, and therefore may not be able to include the variety of instruments that Sheeran uses, such as more uncommon synthesizers and string instruments such as the mandolin. Moreover, as stated in MuseNet's limitations, the accuracy of the AI in following more restrictive prompts about instrumentation is low due to the lack of formal emphasis on instrumentation. Therefore, MuseNet's compositions will include the general instrumentation patterns in Sheeran's style but may struggle replicating the variety of instruments used by Sheeran in his music. For example, in Sheeran's song *Galway Girl*, uncommon instruments are used, such as the Bodhrán and tin whistle, instruments that MuseNet may have difficulty replicating.

Melody: Melodic and harmonic schema is also very highly developed. First of all, its fundamental function to predict the next note in a series of notes allows it to accurately learn from large databases of MIDI files that represent Sheeran's style to form a melody inspired by Sheeran's style. It comfortably learns the general melodic patterns and accurately present them in its compositions. The AI's ability to form chords and harmonies by manipulating the placements and patterns of different tokens allows it to incorporate key chord progressions and harmonic schema similar to that are used in Sheeran's style. Combined with the structural and note embeddings, as well as the ability to integrate different dynamic patterns, the AI is able to incorporate varying harmonic layers throughout the piece, mimicking Sheeran's versatility in forming harmonies. The AI's extensive understanding of musical composition also

allows it to incorporate more complex harmonic elements that Sheeran uses, such as tonic and dominant pedal tones as well as fitting basslines. Therefore, MuseNet is well versed in reproducing Sheeran's melodic and harmonic style.

Rhythm: Because Sheeran's rhythmic style is largely consistent throughout his music, the AI can easily pick up on how his rhythms generally sound. MuseNet's ability to use Pair Byte Encoding to create tokens for the start of each note can allow it to easily identify the rhythmic patterns in Sheeran's music by marking each subsequent note with the same method. This allows the AI to accurately implement rests and accents to emphasize upbeats in its compositions. Because of the method used by the AI to understand complex rhythms using intervals that are both expressed in the tempo of the music as well as time in seconds, the AI can easily produce a large range of rhythms, simple or complex. Sheeran's rhythmic style of repetition and persistence is easily understandable to the AI because of its extensive training in manipulating different intervals of time and utilizing different tempos in its music. Further, its ability to keep track of time in seconds will also allow it to keep a consistent length of the structural components, something Sheeran specializes in. Finally, positional embeddings allow it to use the same rhythmic patterns for different instruments throughout the song. This can help replicate Sheeran's signature syncopated rhythms.

Overall, MuseNet can replicate Sheeran's style quite accurately. With very little discrepancies throughout Sheeran's vastly unique style which incorporates a multitude of artistic techniques, the AI can reliably compose music with his style–these compositions would likely convince listeners that it sounds like Ed Sheeran. The compositions will retain the upbeat, melodic, and unique motifs that Sheeran implements in his music. Because the AI can map so many stylistic elements from Sheeran's music to its own compositions, ordinary listeners will feel the joyous, energized, and comforted feelings. However, given that the recreated music may not add much unique variety to already existing music, it is likely that the human touch, also referred to as the "soul" of the composition, may end up lacking in AI generated music.

Discussion

The use of artificial intelligence in creation of music has been around since the early days of AI. Back in 1981, AI researcher, Marvin Minskly explored why we like music in "Music, Mind, and Brain: The Neuropsychology of Music" (Minsky, 1981). In today's world, it is commonly accepted that production of music will use some form of AI to make the task more efficient. Experimenting with different scales, voices, rhythms is commonplace in a music studio. Academic researchers working on the intersection of AI and music have collaborated with Industry to develop interactive generative AI tools to make these tools accessible to a broader audience. For example, Chris Donahue at Carnegie Melon University is involved in several music AI projects in collaboration with Google, like SingSong (Donahue et al., 2023) and MusicFX DJ (Google, n.d. "*Music FX"*). Another AI music collaboration of artist Skygge with researchers at Sony's Flow Machines Project produced the first AI generated album Hello World (Helloworld Album, n.d.). It started as a small project with a similar aim as in our work - trying to capture and reproduce the concept of musical "style".

The framework developed in this work applies conventional music theory to popular music and assesses its utility to create new music that matches the style of a particular artist. This framework can be used as a basis to develop a musical profile of an artist that can be used as an input into a generative AI model, or simply by a "fan" to produce music similar to the artist of interest. Our case study using MuseNet as an example, suggests that generative AI is fairly accurate in emulating an artist's style based on input provided for the learning algorithm, and will likely be integrated more and more into the music production and composition industry. It provides evidence that AI can follow modern art styles and is adaptable to a wide variety of stylistic approaches, by using modern digital technology such as tokens, embeddings, and MIDI files to precisely capture musical styles. Although artistic styles are well captured, because the software is trained only on a few existing compositions, the generated music may not add much unique variety. An interesting question to explore will be whether providing a larger, diverse library for learning will enable the AI to capture the artist's style more intricately, or do the variations confuse the AI and it loses the artist's signature style.

As of December 2022, MuseNet was discontinued by OpenAI. However, several other generative models such as Google's Magenta, MusicLM, and OpenAI's new Jukebox are available. Future research should examine a larger dataset including several artists using our framework to generate an input into AI models. By using multiple

artists' styles, we can further and more accurately assess the capabilities of these AI models. More variables should be considered, such as the time period, origin, and genre of music.

As generative AI algorithms become increasingly powerful, a growing topic of debate is what makes a musical composition original and unique. In the high-profile trial Griffin et al. vs. Sheeran**,** the plaintiffs argued that Sheeran had copied elements of Marvin Gaye's 1973 song "Let's Get It On" in his 2014 hit "Thinking Out Loud." The crux of the lawsuit was that "Thinking Out Loud" used a similar chord progression to "Let's Get It On." Ed Sheeran eventually won the lawsuit; however, it highlights an important issue regarding intellectual property rights, especially in the age of AI. While a classical music piece explores a particular theme in detail and is typically unique in its exploration, popular music often follows common patterns. Most pop songs are constructed a fundamental set of common chord progressions and usually include a pattern containing the verse, chorus, and bridge. Many popular music songs from different artists may use the same chord progressions and structure. So, what makes a song original? Since AI, by definition, uses prior compositions to learn, can any music composed using AI be considered unique, or is it the process used to create the music that makes it unique? This argument has certainly been raised in several infringement lawsuits, including the one mentioned here. Another consideration is that uniqueness may stem from the physicality of the artist—the performance, energy, and emotions during the performance. As AI becomes more powerful, niche fields using cognitive psychology in music, such as cognitive musicology, will become increasingly important in creating a truly novel, human-like experience. We conclude this discussion with the hypothesis that AI will become a useful, friendly tool to assist artists in composing novel, intricate, soulful music, rather than being viewed as a competitor. Benign uses of AI to emulate an artist for learning or entertainment purposes should be welcomed and perhaps seen as a testament to the artist's popularity.

As far as this work is concerned, we believe the findings and conclusions in this work are a significant step in understanding and improving the capabilities of AI generated music, and insights from this work can influence future research in this field.

References

AIVA. (n.d.). *AIVA - Artificial Intelligence Virtual Artist*. Retrieved September 29, 2024, from https://www.aiva.ai/

Donahue, C., Caillon, A., Roberts, A., Manilow, E., & Esling, P. (2023). *Singsong: Generating musical accompaniments from singing* (arXiv:2301.12662). arXiv. https://arxiv.org/abs/2301.12662

Google. (n.d.). *Music FX*. AI Test Kitchen. Retrieved September 1, 2024, from https://aitestkitchen.withgoogle.com/tools/music-fx

Helloworld Album. (n.d.). *Helloworld album*. Retrieved September 1, 2024, from https://www.helloworldalbum.net/

Krumhansl, C. L., & Kessler, E. J. (1982). "Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys." *Psychological Review, 89*(4), 334–368. http://music.psych.cornell.edu/articles/tonality/1982K&KesslerPsychRevfromJournal.pdf

Magenta. (n.d.). *Magenta: Music and art generation with machine learning*. Retrieved March 4, 2024, from https://magenta.tensorflow.org/

Minsky, M. (1981). Music, mind, and meaning. *Computer Music Journal, 5*(3). https://web.mit.edu/6.034/www/6.s966/Minsky-MusicMindMeaning.pdf

MuseScore. (n.d.). *MuseScore: Free music composition and notation software*. Retrieved March 4, 2024, from https://musescore.org/en

OpenAI. (n.d.). *Composer and instrument tokens*. MuseNet. Retrieved March 4, 2024, from https://openai.com/research/musenet

OpenAI. (n.d.). *Data set*. MuseNet. Retrieved March 4, 2024, from https://openai.com/research/musenet

OpenAI. (n.d.). *Embeddings*. MuseNet. Retrieved March 4, 2024, from https://openai.com/research/musenet

OpenAI. (n.d.). *MuseNet: A state-of-the-art music generation model*. Retrieved March 4, 2024, from https://openai.com/index/musenet/