Fighting Time: Artificial Intelligence and the New Era of Musical Expression

Kali Anchlia

Abstract

This paper explores the dynamics of and developing interaction between human creativity and artificial intelligence (AI) in the field of music. It mainly focuses on the emotional and psychological effects of music reception regarding AI-generated music. The study explores the emotional impact of music, concentrating on jazz as a special sub-genre characterized by improvisation, which provides insight into the cognitive abilities of AI. This analysis emphasizes lacking performance ability, along with the shortcomings of AI composition systems. However, the future potential of AI is also emphasized by the evolution of AI-technology in music. With the conversation situated within the context of western philosophical viewpoints, the hypothesis posits the growing importance for musicians to embrace AI, while still addressing the significant emotional gap that remains between listeners and AI-generated music.

1. Introduction

Imagine each individual life encircled within a sphere. Every person is the center of their own sphere. The things in this sphere are only that which we allow to enter. It is rare that we open the tiny door at the edge of our sphere and let something else in. Humans after all, at our very core, are survivors. Therefore, the dream of our world is carefully constructed in order for us to do exactly that: survive. When we listen to music however, the key that we have kept hidden finds its way into our hands, and we open the door before we realize what is happening. If music is our siren, then we are the sailors at sea. However, modern life is not ruled by myth, and our siren could be either human or technologically made, or some combination of both. The question is then posed, can artificial intelligence produced music ever tempt us the way human touched music can?

This paper presents the relationship between AI and humans in the field of music, as well as the potential road to acceptance. It centers around the emotional and psychological aspects of the human mind and the place of music in this dynamic.

Our feelings no doubt play a large role in the way we receive the music that is communicated to us. Research shows that regardless of the genre, the overall effect of emotions is a powerful driving factor in how music is received by the listener. A sub-genre explored more in depth however is jazz music, as it includes an aspect not as apparent in other genres: improvisation. With an emphasis on the context of cognitive systems and the suitability of improvisation to AI, several perspectives are discussed. Along with this, the contrast of old and new AI systems and artificial intelligence's overall evolution, clarifies the hypothetical questions described above.

Based on the information that is presented, it will be clear that in order for AI-generated music to meet or surpass music created by humans, it must first evolve to develop cognitive desire similar to humans, and meet emotional standards both in compositions and in live performances. Although it is currently extremely lacking, AI has the technological potential to advance and eventually generate music barely distinguishable from humans, at which point musicians must adapt to AI's presence in the field.

2. Emotional Connection, Human vs AI Generated Music

The strength of music is often underestimated. Music has the ability to close the gap between mind and body, providing an almost psychedelic experience. The purpose of music is to make others feel, and consequently, music is absolutely reliant on how well people are able to relate to and appreciate these combinations of sound. Music is an entity where other's opinions do matter, and there is always room for growth and development. Therefore in this section, we will touch upon to what extent AI versus human generated music can– or cannot– evoke emotion.

To lay the groundwork for further discussion, we must first consider the fact that the human brain is interested in changes in sound. We often cannot connect emotionally with only one constant frequency. When addressing the difficulty of 'modeling expressiveness' in music through AI, Ramon Lopez de Mantaras and Josep Lluis Arcos develop this necessity of variability by introducing and explaining the constantly firing auditory neurons in our brain [1]. If auditory neurons repeatedly receive the same stimulus, their firing rate decreases over time: this process is known as *habituation.* In other words, without alterations in the music's pitch, rhythm, and dynamic, we perhaps unknowingly detach ourselves from the listening experience. It is simply no longer interesting enough for our mind to be actively following the music.

If avoiding such habituation is the case, then the basic assumption would be that only a distinct 'change in sound' is needed for artificial intelligence to match the emotional intelligence of human beings. However, we cannot forget to apply higher-level parameters based on music theory and machine learning, in order to prevent what is otherwise sure to be a cacophony of disruptant, horrifying notes. Although an evolution of AI systems and more specific parameters will be discussed later on, it is beneficial to introduce here the three blanketing techniques for incorporating AI algorithms into music as described by Robert Laidlow in his 2022 thesis [2]. These techniques are not regarding the systems themselves, but rather what human musicians can do with already established assistance algorithms. However, it is important to note that these methods are more related in this case to classical and ensemble music than in a jazz-style approach. (The contrast between the two genres will be considered later)

1. Interlocking: acknowledges the limitations of AI, specifically its current inability to produce coherent music for lengthy durations, by alternating between AI-generated and human-created music.
2. Collaging: involves the combination of various layers of AI-generated music. Uses solid micro-level AI music, to piece together and formulate a final cohesive sound.
3. Hidden Layers: a blueprint of the music is formed solely by AI (allows the composer more creative possibilities). However, the individual elements of AI-generated music are not directly heard in the final piece.

The obvious interference here of humans in AI's own 'creative process' shows the limitations artificial intelligence has thus far when it comes to creating music, and expands the view from considering each note separately, to a more holistic perspective of the overall piece. What is more, the three techniques described above apply to ethical questions and the overall awareness of AI, and thus will be further explored. Still, Laidlow's assertion is narrow in the sense that it is written from the point of view of a composer, rather than a machine learning engineer. However, Laidlow's evaluation does convey the understanding human musicians have of feelings and emotions. The skill of placement (again referring to the three techniques) based on one's senses is something to consider as a powerful tool in music.

While one side of the relationship is the composer, the other is naturally the listener. After understanding not only the general process of our neural system when it comes to music, and

techniques currently used by musicians to formulate music so that it fits these criteria, it is vital to consider the actual human to human connection and its relation to emotive responses. A feature by The Learning Network, prompted students from Ohio, Vancouver, Washington, Kentucky, and Washington, D.C., to submit writings on their thoughts on AI replacing artists in the music generation [3]. In an almost unanimous agreement, they found it inconceivable that AI could replace the unique charisma and talent of live, human performers. The general consensus (although we cannot ignore that the sample size of the investigation was small) was that people like to know that they are listening to someone's music— someone being the key word here— who has gone through similar pains and obstacles as them. It is no surprise that a human finds comfort, hope, and excitement in relatability. Even when ignoring AI's impact on music, one can date as far back as the popularity of Kraftwerk in the 1970s, to represent this idea that new innovation can never really surpass the familiarity of the connection between the performer and the audience.

2.1) Jazz, AI Improvisational Systems

Jazz music is known for its expressive and authentic qualities, and here we must emphasize that the function of AI in jazz versus other genres must be different in order for the points addressed in the previous segment to be successful. There are several implications of accepting AI composed jazz music in the broader jazz community. Understanding cultural and societal ramifications, along with the preservation of jazz heritage is crucial. However, what must truly be perfected is the *art* of improvisation. Improvisation is more than vital to jazz:it defines the genre. Improvisation is both a reflection of individual artistry and a collective tradition passed down since the late nineteenth century.

Therefore, along with the history of music, there has been a subsequent history of what we classify as *Improvisation Systems,* which are related to the real time creation of jazz solos [1]. The existing research synthesized by Mantaras and Arcos highlights chronologically some of these improvisation systems (which they state are one of three main subsystems of computer music in which AI techniques have been applied), along with their methods of functioning. For instance, the Flavors Band system attempted "random functions" paired with "musical constraints" in order to generate variations in jazz [1]. Its well known arrangement was an improvised solo of John Coltrane's composition 'Giant Steps.' Noticeable was its use of procedural language to define pre-set rules for improvisation.

Another example is ~GenJam~ which interestingly uses something called a genetic algorithm to improvise [1]. A genetic algorithm simulates the logic of Darwinian selection in the field of machine learning, where only the best attempts at improvisation are selected for replication and the rest are discarded. In the 1994 GenJam release, a human listener acts as a judge, and scores the improvisations of the system. Later iterations have also attempted the same method but with the use of automatic improvisation rating algorithms based on important melodic and harmonic qualities, rather than solely on the subjectivity of humans.

Employing neural networks was explored in 2001 with ~Franklin~, a system taught to improvise through transcriptions of human-solo improvisations [1]. Most notably the saxophonist Sonny Rollins contributed to Franklin, both in solo recordings and with the overall style of the system after trial and error of imitation. In all of the three above examples, different algorithms in improvisation systems are used for the same ultimate goal: for AI to improvise on the spot, make something that sounds unique, and gives people the feels.

However, some criticize, “The lack of interactivity with a human improviser… on the grounds that they remove the musician from the physical and spontaneous creation of a melody” [1].

So we have a contradictory goal here. On one hand, we expect AI to fully improvise on its own, but on the other hand, we critique an AI system that does not include a human mind in the process. This duality brings to attention the varying degrees of acceptance prevalent in society (both in the past and in modern day) of AI in music. Thom notes how these expectations were met via the popular 2001 Band-Out-Of-A-Box (BoB), a “real-time” interactive system that uses an unsupervised probabilistic clustering algorithm [1]. The BoB functions by learning all possible notes it could play at a given point in time, and then reacting to four-bar solos of human improvisers. The enhanced interaction between AI and humans did allow BoB to be a more successful machine than some previous attempts, especially because it adapted human musicians' intentions and artistic vision. These human-involved AI systems however, may defeat the point of improvisation altogether. If improvisation is defined as ‘creating and performing spontaneously and without preparation,’ then how could a system where a human composer is changing a generated solo until it becomes something that suits their style be classified as improvisation? We can also see a limit in authenticity within the jazz genre if AI is simply imitating the styles and chord progressions of human musicians it learns from.

2.1a) Memories and Rules in Live Performance

One unanswered question is rooted in the challenge of live jazz performance, and the capability of AI in navigating between structure and improvisation in these scenarios. Ramalho and Ganascia claim that “Jazz musicians' activities are supported by two main knowledge structures: memories and rules” [4]. If we take this idea as the basis for everything that follows, more light will be shed on the specific mechanisms in which AI generates improvised solos. In one important way, artificial intelligence and humans are similar: both learn from past experiences. Let’s begin with ‘memory’ and the trigger point of memory, context. Musicians cannot always justify the actions they take during performances; several of them are actually the result of unconscious memories being activated mid-event. Most of the time these memories will appear as a result of the environment and piece one is performing, and alter the method of performance through “concepts such as tension, style, swing and contrast” [4]. As jazz is an intuitive and ever-developing genre, these concepts can only be mastered by observing, listening, and *imitating*. Whether this imitation is through watching videos of jazz legends performing, or playing in a band and bouncing ideas off of fellow members, storage of long term musical memory is important. AI seems to be learning in the same way as humans then, although it may be building up memory differently. After looking at imitation from the perspective that it is simply a form of knowledge and learning, the benefits seem to outweigh the problem of authenticity posed before. Especially when we realize there are double standards when humans do the same thing. Now, this is not to say that imitation is equivalent to improvisation, and again, the ethics of taking one’s style as their own should still be debated. Regarding exactly how memory– imitation or not– is stored, we have a contrast between long term and short term memory systems. AI seems particularly adept at processing and communicating short term memory, but has difficulty with longer pieces of music. Typically its shortcomings include creating genuine emotional responses because of its limited interpretive abilities and the previously discussed repetitiveness. To combat these limits, Princeton computer science student Ji-Sung Kim developed his 2016 project ‘DeepJazz’, using a long-term short memory (LTSM)

recurrent neural network [5]. Traditional RNNs (recurrent neural networks), work with short musical phrases, but LTSM's do so with longer, more complex compositions. Ji-Sung created DeepJazz's LTSM to be trained with only one piece, Pat Metheny's "And Then I Knew.", and subsequently generated piano solos. A chief technology officer and critic Ken Weiner, although acknowledging the use of the now popular LTSM as a good method, believes the reduction of the instrumentation to only piano is extremely simplistic (Weiner 2019) [5]. He compares "fixed" notes of a piano with "malleable extended tones" of wind instruments such as trumpet, trombone, and saxophone, and believes that the complexity of vibrato and similar expressiveness is something AI cannot perform. David Borgo, a University of California, San Diego, music professor and saxophonist adds onto the concern: "Research in this area has tended to focus on getting computers to play the 'right notes,' but we are still a long way from designing systems capable of the micro and macro temporal, timbral and textural adjustments necessary to groove together and to develop high-level collective improvisation in an unscripted fashion with human musicians (rather than insisting that human musicians improvise with, or groove to, the computer)". Clearly, progress still needs to be made in AI memory, especially when Jazz music is known for its *soul* and ability to create and control the mood in a setting (when played live). Of course, feeling is only half of it; the other half is 'rules', which is much better suited to artificial intelligence in its current state. These rules consist of commonly agreed upon music theory, and some unavoidable incomplete thoughts (in which case it would rely upon memory to decide what exactly to do). Although jazz is seen as more of a free form of music, theory and understanding of the way notes act in relation to each other is extremely important. The most basic definitions of musical concepts integrated in AI would be on "pitch, duration, amplitude… temporal sequence of notes… set of simultaneous notes… Scales and rhythm… melodies and/or chords" [4]. The application of music theory can be ultimately divided into 4 categories; melodic analysis, harmonic analysis, rhythmic analysis, and structural analysis [6]. Shih-Lun Wu and Yi-Hsuan Yang present their 'Jazz Transformer' as a neural sequence model that analyzes these categories.

1. **Melodic Analysis: melodic phrases and patterns (pitch range, note durations, intervals between notes)
2. Harmonic Analysis: chord progressions and harmonic structure (jazz chords, harmonic tension/release)
3. Rhythmic Analysis: rhythmic patterns and complexities (beat patterns, syncopation, swing)
4. Structural Analysis: overall form and organization (traditional jazz song structures like AABA, blues form, other common forms)

Although different machines integrate music theory into their algorithms in different ways, they are always thoroughly encoded, because without understanding of sound, AI would not be able to generate anything of substance– memory or not. The quantitative measures of the Jazz Transformer is a perfect example of this, and the conclusion Wu and Yang found correlates to previous knowledge of AI's ability, as well as adding that it lacks new styles, techniques, and cultural influences. The challenge is giving machines intrinsic motivation and the ability to make creative decisions based on real-time feedback; this is called autonomous creativity [7].

"If we produce sounds with the intent to make music, that is music," said Shlomo Dubnov. "That means that, to become musical, the computer must have its own intent."

French scientist and composer Francois promotes PACT's as a generic framework for representing the potential actions (or intentions) that musicians may take within the context of

performance [4]. PACT's consist of a combination of low level and high level instructions. Examples of low-level PACTs include "play loud," "play this rhythm," and "play an ascending arpeggio," . A low-level PACT on all three dimensions could be, "play this lick transposed one step higher". High-level PACT's include "play syncopated" (on rhythm) and "use major scale" (on pitches) [4]. These can be activated either during a specific moment, or encapsulate the entire song; its flexibility is the reason it is considered by some, a basis of problem solving for musical creativity modeling. The AI relies on the playability of these PACT's however, and because of its varying dimensions, the less abstract a PACT is, the more it is 'playable'. For example, "play funk-rock" is less playable than "play an augmented fifth on the third beat". There are also restrictions based on combinability of PACT's; a simple example is that "play loudly" and "play quietly" cannot be combined. This potential actions structure doesn't seem to utilize a musician's (AI's) intentions described earlier, however the perception, composing, and execution modules of PACT's does lean towards it more than other systems [4]. The three elements they rely on are short term memory contents, the context and mood, and a future chord grid segment. The simplicity is ingenious; first the perception module corresponds to the context and mood, and puts these events into the AI's short-term memory (STM). Next, the composing module uses the context stored in the STM to combine playable PACT's, and sends this information to the 'future' chord grid segment (similar to lead sheets). Finally, the execution module takes the information from the chord grid and executes the PACT's at the correct time by sending information back to the perception module. In this way, the AI is able to develop some sort of musical intention, although it is still relying solely on the context fed to them. Rather than focusing on rules and memory separately, PACT's combines both in an effective way. If this idea is perfected, AI already comes a lot closer to becoming the self-thinking and feeling improvisation systems we doubt they can be.

The Defense Advanced Research Projects Agency (DARPA) hopes to advance the progress of AI in music even faster, and are conducting research on jazz-playing robots that can interact and improvise with human musicians [5]. DARPA hopes to blur the lines between man and machine, by studying and implementing nuances of jazz music. If they succeed, we truly won't be able to differentiate much between artificial intelligence and humans (at least in terms of live performance); however the physical aspect of this is very interesting. In 1970, a Japanese roboticist Masahiro Mori first proposed the *Uncanny Valley Theory* [8]. Based on collected data, Mori reached a conclusion that as robots appear more human-like, they become more appealing to humans— that is until this likeness reaches a figurative 'uncanny valley'. At a certain point, the feelings of commonality and emotional connection devolve into a sense of unease. It's when the robot becomes too unnaturally human-like, that we begin to feel frightened. The uncanny valley brings to attention the difficulty of accepting AI if it even reaches a point where it can look, act, and play live music exactly the same as humans.

2.1b) Cognitive Systems and Philosophies

To add upon jazz and improvisation in live settings, we must explore cognitive systems and the psychology of the human mind, along with its application to artificial intelligence. Francisco Varela and Umberto Manturana propose an *autopoietic perspective,* which centers around the idea that living organisms are self-creating and self-sustaining. In the previous segment, much of the 'memory' based discussion focused on external adaptation and remembrance depending on various contexts. However, the autopoietic perspective shifts the focus to internal coherence, and emphasizes the continuous self-renewal and unity of organisms [6]. With this shift in focus

there is a whole new dimension of thinking. Previously we could only look at the surface level impact of memory, rules, and context, but now we delve into mind-body connection and awareness. Steve Torrance and Frank Schumann expand upon the autopoietic perspective, and umbrella into the idea of Enactivism, which is defined by two key concepts. The first is "Laying down a path in walking", initially introduced by Varela in the conclusion of her 1991 work [6]. She proposes that organisms create their world as they interact with it, rather than passively mirroring it. To relate this concept to music, musicians generate music in the moment, and cannot take back the actions they have already performed. Varela says this is akin to laying down a path without the ability to correct; in other words, AI generates its own world with its perceptions and history of learnings, rather than just completely copying others. Therefore when AI makes a musical mistake, it is proof that it is attempting at being the 'self sustaining' entity that is the very definition of the autopoietic perspective. Varela quoted a poem in her work by Antonio Machado, a Spanish poet, and the last line according to Torrance and Schumann captures the essence of jazz improvisation and performance in general, "when turning around you see the road you'll never step on again". Following this, the second concept is "Sense-making", which was also first written in a paper by Varela in the early 1980s. "Sense-making" suggests that the nervous system creates meaning rather than processes information. She applies this to the interaction between an organism and its environment, relating it to musicians generating meaningful musical responses in real-time [6]. If the nervous system creates meaning, then AI has to be able to observe the environment it is performing in, rather than only focus on its own system and musical interests. Selfishness usually never makes a good performance; knowing what one wants to make the audience feel however, certainly gives intention to your instrument. Intention can also be communicated between fellow band members. Enactivism is a valuable theory, because we are able to attempt to look at things from the perspective of artificial intelligence, and see how it may cross the line between technology and humans.

Since jazz involves temporal awareness, it often lends itself to be an extremely physical and emotionally charged performance. It's this energy and spontaneity that allows the audience to forget about effort required to get to a point where improvisation comes to the tips of their fingers. Improvise originates from the latin 'improvisus' meaning 'unforeseen' [6]. Although the music should be unforeseen to the audience listening to the music for the first time, it is very rarely completely unforeseen to the performer. Based on evidence, we already know that experience of some sort is required for musicians and artificial intelligence alike, but we haven't yet truly discussed practice. Even a highly skilled musician— or rather *especially* a highly skilled musician– will tell you that practice is key; jazz or classical, it doesn't matter.

"... it is a mistake to link the "improvised" too closely with the "unprepared." In fact a considerable amount of improvised performance is extensively planned in advance—and much jazz performance includes, or is indeed mainly comprised of, playing from pre-written sheet music. Even in most contemporary jazz performance, where musicians may improvise singly or collectively over extended stretches, there will usually be pre-arranged structural constraints" [6].

In jazz, the line between preparation and performance is blurred, because the musician always learns something new about the way to approach a piece when performing. A dialogue between Herbie Hancock and Wayne Shorter (well known jazz- musicians) actually underscores the notion that improvisation in jazz requires "constant, unseen effort" to achieve true brilliance [6]. In this case, AI's failures and attempts to improvise are only part of its "constant, unseen

effort", and will eventually (on its journey to development) lay the correct path and become a successful self-creating 'organism'.

If, when applying the previously described 'sense-making' AI is creating meaning in real time, is it doing so consciously? If so, then to what extent of consciousness is required for AI to be on par with human musicians? We now discuss mindlessness versus mindfulness in order to understand AI's (and human musicians) expertise development. Looking at American philosopher Hubert Dreyfus's texts as a general starting point allows for an introduction to these questions, and acts as a basis for establishing complex views. Typically, there are two cognitive systems (with no designated names) which philosophers commonly debate. The first corresponds to Dreyfus's notion of "immediate coping" where he argues that organisms operate automatically and pre-consciously [6]. In other words, he says that the expertise development of a musician is a state of unconscious thought, because in his words, "thinking disrupts smooth coping," [9]. He explains the idea with his famous interpretation of why Baseballer Chuck Knoblauch kept throwing errors [9]. The second system emphasizes the early stages of development, where organisms act slower, more reflective, and rely on method-based thinking. With this, rational mindedness combats the supposed effortlessness of the actions described in System 1. Tensions arise because of the blatant contrast between Systems 1 and 2.

The definition of expertise to humans is subjective, because to someone who doesn't play an instrument, expertise could be a few nice sounding chords. However, when musicians constantly judge and push themselves (and others) in order to reach the 'next level' and become 'better', only they know how high their goal exactly is. Dreyfus defines everyday expertise in this case, not as technique or skill, but as how well each individual meets their own needs. He mentions 'expert flow in action', or in other words, full engagement in one's actions without overthinking it. This suggests that mindfulness is present in their actions, but detached contemplation is not involved. A common example would be making coffee everyday for years; at some point the experience– in Dreyfus's argument– leads to expert motions without the requirement of explicit thought [9]. Some people may not add an ideal amount of cream or sugar, but it suits their needs perfectly. Now, if 'expertise' truly is what Dreyfus argues, AI in music is nowhere near reaching the expertise level of humans. Since the model of the AI is its brain, depending on the kind of model the AI is being fed, it could involve thinking in the moment, or already have all the instructions ahead of time. Either way, this level of 'thought' in the action does not correspond to Dreyfus's argument. Some similarities could be seen, because if the AI has a genetic system (which was explored earlier on), only the good outcomes would sustain, and so the AI's system may have less qualms to reproduce those outcomes (which again could be seen as "less thought"). Still in the end, AI has no subconscious, so it's impossible for it to have 'mindful actions' without actually thinking. Philosopher John McDowell defends his claim that rational mindedness is 'pervasive in our lives' by arguing that Dreyfus's idea is a myth at its very core [10]. McDowell specifically focuses on Dreyfus's case of a chess master absorbed in lightning chess . Joseph K. Schear explains McDowell's argument well in the introduction of his novel *Mind, Reason, and Being-In-The-World*:

> "McDowell urges that the chess master's absorption does not prevent him from knowing what he is up to, and that, moreover, if the chess master really is a master, he will be able to give rational explanations of his moves as intelligible responses to the forces on the board. So understood, the chess master's expert play is a case of "cultivated rationality" in operation. This is precisely the "actualization" of conceptual capacities in experience and action that McDowell is keen to highlight".

In terms of humans, philosophers believe McDowell went too far by saying that there is no slow, reflective thinking (as described in *system 2*), only completely rational thought without actively using deep thinking. Although we previously disputed Dreyfus's argument, there is a truth to the connection he made between overthinking and disruption of an action. Stage fright is common for performers– however, experienced musicians can overcome this to some extent by becoming more absorbed in the music than focusing on their heightened awareness of their surroundings (which will no doubt be a crippling distraction). American Philosopher Taylor Carman takes a similar view by regarding McDowell's theory as "scholastic fallacy" [10]. He believes that McDowell is attempting to fit structured and rational thinking into experiences that are supposed to be immediate and not deeply thought out. In other words, he is trying to use complex, organized thinking where it doesn't belong, which again is similar to our previous thoughts. Having rational contemplation as McDowell argues, may prove disastrous to a human musician then. However, looking at this in terms of artificial intelligence, if McDowell's theory is correct, AI is (in contrast to Dreyfus's theory) already an expert. AI does think rationally after all. Still, the whole point of this section of the paper is to discuss AI's ability to express emotions and make listeners *feel* the same way they do when human musicians perform. Based on the hypothesis that emotions, and therefore understanding of both the context and the piece beyond the surface level of rationality is truly vital, we can discard McDowell's theory momentarily.

Barbara Montero, a professional ballet dancer turned philosopher takes another view on Dreyfus's theory in her "A dancer reflects," [10]. First, she dismisses Dreyfus's principle of automaticity, by arguing that there is not enough psychological evidence to support the theory. She mentions that action-in-the-flow is not necessarily mindless, and that if it were it would conflict with scenarios where experts may need to reflect or monitor their actions. This is similar to classical musicians who often view mindless performance in-the-moment as a mistake. Musicians like cellist Inbal Segev believe that eliminating thought during performance eliminates the artist's ongoing conception and image of the music [6]. Of course, classical music in comparison to jazz music will be discussed further later on. Furthermore she diverts from Dreyfus's theory when she categorizes 'esoteric expertise' separate from 'everyday experience' [9]. Esoteric expertise refers to specialized knowledge or experience in a certain thing: such as performance in a particular instrument. Joshua A. Bergamin acknowledges this in *Being-in-the-flow: expert coping as beyond both thought and automaticity:*

> "But for Montero, esoteric expertise is marked by a continuous desire to improve and she insists that this highly-motivated attitude– called kaizen in Japanese– requires explicit cognitive attention if it is to succeed" [9].

Therefore once again the idea that thought and 'explicit cognitive attention' is important surfaces. However, there is also another idea here: for AI to be an esoteric expert, it must have the desire to improve. In revisiting both the autopoietic perspective and autonomous creativity (2.1a), for AI to be self-sustaining, it must also have these human-like attributes of wanting to evolve. This is fundamentally not possible for AI currently, as it is the person who has created the model who is thinking, and pushing the AI to improve. The AI is a clone of the engineer, not its own individual— at least not yet. Still, if we take Montero's theory and apply it here to AI musicians, there is a significant way to gain esoteric expertise. It also relates to the previous idea of a 'constant, unseen effort', because one would only put in the needed effort should they have the desire to (or in the terms of artificial intelligence are being programmed to). Comparing humans and AI in this manner doesn't help however, because we cannot expect a machine to replicate the emotions of humans. However, in terms of the acceptance of AI musicians in

society, we can see that there will be a struggle to consider a machine as an expert when it cannot truly be classified as its own individual. Although there are several more philosophers such as Gallagher, Toner, and Moran, who contribute immensely to this discussion, the debate on expertise and mindfulness is helpful in understanding several perspectives on how human/AI performers (for jazz or otherwise), think and act in the moment.

2.1c) Classical Music vs Jazz Music

Looking at the difference between classical and jazz music allows for a continuity in the philosophies explored in the last segment, and of course, brings new ideas about the relative success of both genres of AI in music. First, there is a clear distinction between jazz and classical music. Jazz thrives on imperfection, with musicians embracing and owning their "mistakes" to create artistic expression. In contrast, classical musicians are known for attention to technique in order to attain a performance as close to perfection as possible. Due to the expectation that classical musicians replicate a score to technical perfection, they are described by some as ‘robots’ [11]. The fact that classical musicians are even ranked by objective ability and seated as such in an orchestra, only adds to this conclusion. However, this is not to say that classical music has less feeling involved. Though the restrictions on how the music of the time period and written style by the specific composer, only make it more difficult to express emotions in phrases, a good classical musician is able to make every perfected note contain beautiful emotion. However, the sound each genre produces is different– jazz is freeing in a way that classical music can’t always be. It lies less so in the score, but more in the soul of the performance itself. Along with this, improvisation is rarely ever in the equation when it comes to classical performance. Most classical musicians in the 20th century cannot improvise at all. These boundaries between jazz and classical have resulted in segregated studies for both genres of music, and similarly require different types of AI systems to replicate/create [11].

We see this difference with AIVA, an AI program which created a 2016 album “Genesis”, and drew from classical master composers such as Beethoven and Mozart to create its own music. The piece “Symphonic Fantasy in A minor, Op.21” specifically drew attention as an impressive creation, because of its use of dynamics and intensity [8]. The AIVA team gathered opinions through *Turing Tests*, in which humans try to distinguish whether the AI generated music is distinguishable from human compositions. As Laidlow (2.0) called it, a “musical guessing game” of sorts [2]. They found that the majority were not able to tell there was a difference, and that the AI took the styles of the composers it had samples from and created something new and profound. However, upon listening to the recordings, they also found that it still lacked the ‘human touch’, and tended to be repetitive. Although motifs are often used by composers, the piece often seemed like it was meant to be leading somewhere, but still ended up back at ground zero. Classically trained musician Hoh Xi Ting, compared the Symphonic Fantasy and another composition “Scottish Fantasy” by Max Bruch (1800). She explains that they both begin as “a slow, mellow, low brass melody, later joined by a violin solo”, but the big difference is that “Humans can feel, and their emotions are reflected in the composition as themes, harmonies, and the occasional, but well-placed discordant chord [12]. There is direction, tension building up to a climax, and a satisfying finale {in the human composed Scottish Fantasy}”. The imperfection shown in these Turing Tests and the comparison above illustrate— as James Vincenta, a senior reporter who has covered AI, robotics, and more for eight years at The Verge writes— that AI may be better suited for less structured music. In this

case, jazz would prove to be an extremely suitable genre for AI composition, as its lack of rigid structure aligns with the short-sightedness of deep learning systems (2.1a) [13].

In order to test this conclusion and attempt to combine the best of both worlds, classical and jazz, together, Professor Geraint Wiggins of Queen Mary's University introduces "the first concert consisting almost entirely of music composed by artificial intelligence" (Wiggins) [13]. The first AI-written song was a jazz combo led by Mark d'Inverno– and it proved to be a remarkable success that "sounded just like the real thing". However, the team still hadn't yet experimented with a less structured classical piece. In order to make it more likable and easier to comprehend to the audience, the music was presented in the style of a familiar composer: more specifically, Mozart (with a jazz spin). The two pieces tested by Stefan Lattner were *Mozart Unchained* and *Mozart Constrained* [13]. In Mozart Unchained, the professor gave the deep learning system little guidance, and allowed the AI more freedom to explore and create. The result was music that sounded less recognizable as classical music, with a melody that was "all over the place". Mozart Constrained on the other hand, limited the system's ability to generate melodies outside of the most statistically likely patterns. This attempt was certainly more coherent, but seemed to be nearly imitating whole 20 second phrases from Mozart's own work. With the direct comparison of live performances of jazz and two structures of classical music, we can draw the conclusion that AI makes better independent judgments in jazz music, but these judgments are out of place in less structured classical music. We can further connect the results of Mozart Constrained back to Montero's question of whether AI is able to function on its own and understand the image of a piece. In this case, the answer is no once again, because it was Lattner's thoughts and limitations that allowed the constrained version to perform well, not the AI's. However, we can see that AI is suitable in its own way to both genres, and as we will learn in the next segment, it has the potential to overcome its obstacles and reach a point where both its rational and irrational 'thinking' could be useful.

2.2) Micro and Macro Evolution of AI

The AIVA team contradicted their statement in the previous section that 'Genesis' lacked a human touch, by stating that times are changing, and that people have been getting used to AI generated music in the past 20 years [8]. AI 's evolution in the past few years is certainly characterized by its increased use to people in the macro-world, but can also be analyzed in the micro-world of each new individual musical piece.

A recent Pitchfork article by Marc Hogan is titled, "Musicians Are Already Using AI More Often Than We Think". The article details Grammy-winning engineer and producer Shawn Everett, experimenting with Open AI's song creation tool. Everett input a chord progression from an unreleased Killers song, and requested from the AI a piece in Pink Floyd's style. He found that the song was unique and a little strange, but overall he was surprised because he thought a human couldn't have done something as creative with the instructions as the AI did [14]. Another credited artist Danny Wolf, an Atlanta-based producer, actively integrates AI into his work. Wolf (as of May 2023) has been collaborating with ChatGPT to generate the concept for his entire solo album. He even mentions that he frequently uses AI to create symphonies, which he then breaks down using another AI program to remove vocals from beats, and then incorporates them in his work [14].

Peter Martin, an Oscar nominated film producer "... recalls how it would have taken 11 hours of a voice for an AI to be able to mimic it just three years ago. "That is now down to less

than two minutes," he says" (Hogan). Commenting on the recent Drake AI remake, Martin explains: "Generating a fake Drake song might involve four or five different AI tools now… But he's beginning to see one-stop shops like Uberduck" (Hogan).

The micro-implications however more so affect development within a single piece. We will be revisiting the ideas of Collaging and Hidden Layers, as well as the performance of additional AI-generative platforms (2.0) to see this. A 2019 composition for mezzo-soprano and ensemble called "Alter" uses several algorithms which serve different purposes: MuseNet as a primary symbolic-generative algorithm, WaveNet as an audio-generative algorithm, and a combination of a text-RNN and 'GPT-2' as a text generative algorithm [2]. The composition is divided into three narrative-driven sections, and is defined by early, middle, and end points in the AI's development. In the early AI development of "Alter", the AI algorithm was simpler and less refined, and the music/lyrics were naive. Here 'MuseNet generations explore repetition as a hidden layer; starting from small-scale repeats, then evolving as the narrative develops, into larger, structural repetition. At the same time, 'WaveNet' is used to create evolving electronic soundscape elements of vocalizations, speech, and singing, that mirror the narrative development of "Alter". The AI becomes increasingly self aware in the middle section of the piece, and gains more maturity according to Laidlow [2]. At the end of "Alter", each instrument or voice performs its own line, creating a collage effect. By this point, AI-generated material (in this case MuseNet), completely takes over the decision making process, and creates more intricate musical textures.

If AI has the ability to learn to go through the above process on its own, it may be able to reach the level of human emotional understanding we expect it to. Both in the macro and micro world of AI-generated music, AI is taking a firmer stance and each new system is becoming a stronger version of its previous self. The truth is that AI is evolving at a faster rate because humans are looking for faster ways to produce music. The ethics of AI use in music must be touched upon in the future in order to fully understand the risks and benefits of AI-generated music.

3. Conclusion

In conclusion, this research has illuminated the complex relationship between AI-generated music, and human created music. As we have reviewed, AI-generated music, although sometimes revolutionary for the time period, is insubstantial in comparison to the music human musicians produce.

AI's largest problem is regarding the low level of emotional capacity within its music. Its inability to create music that makes a significant emotional impact on listeners is no doubt a great hindrance to its advancement. Not only this, but AI is also unable to create a lasting connection with the audience, because it cannot interact with fellow humans the way living-performers can. As discussed before, AI is completely different from humans, and as such this comparison may not be appropriate. However, because music is almost wholly ruled by emotions, AI must be able to produce emotions in music to some extent.

We can see this need further in jazz music, where AI has a lot of trouble in creating the energy and atmosphere required of the genre. AI is unable to improvise without some level of imitation, and is repetitive to a fault due to limitations in its systems long term memory. In terms of the question of improvisation, we looked over arguments concerning how AI could be suitable– or unsuitable in terms of spontaneity. Here some positive light was shed on

AI-generated music being ‘unpredictable’ in a surprisingly creative manner. However the counterargument to this is AI simply being ‘out of touch’ with music-theory in live circumstances. Then again it can be argued that making mistakes is important for AI to be considered an evolving ‘organism’.

It is obvious that several questions are left unanswered: Is improvisation meant to be completely in-the-moment, or slightly practiced and structured? Is intrinsic motivation truly possible for AI? Is expertise described as detached or contemplative? Is the uncanny-valley theory pervasive or could AI’s advancement lead to growing acceptance in the new generation? AI’s definition in relation to music, and the way scientists and composers view its evolution will affect future research on this subject matter, and perhaps bring to attention new claims that could harm or help the case of AI- generated music.

As we contemplate the significance of AI in music, it becomes evident that understanding its role is not just a matter of technological advancement but also a profound exploration of human versus machine in terms of emotion and creativity. AI’s advancement in music could lead to a future of increased collaboration with human musicians, or complete independence once it reaches appropriate cognitive ability (in terms of humans subjective critique). In essence, the acceptance of AI in music means redefining our spheres of creativity and embracing new, unconventional, artistic expression.

4. Acknowledgements

PHD Rob. S, mentor and guide

5. References

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

[2] Laidlow, R. (2022). PDF. 11 Jan 2023; The Royal Northern College of Music in collaboration with Manchester Metropolitan University.

[3] Network, T. L. (2023b, May 11). What students are saying about a.i.-generated music. The New York Times. https://www.nytimes.com/2023/05/11/learning/what-students-are-saying-about-ai-generated-music.html

[4] Ramalho, G., & Ganascia, J. G. (2019, March 9). Retrieved October 15, 2023, from https://www.researchgate.net/profile/Jean-Gabriel-Ganascia/publication/2671507_Simulating_Creativity_in_Jazz_Performance/links/02e7e51867c15dc574000000/Simulating-Creativity-in-Jazz-Performance.pdf

[5] Weiner, K. (2019, April 29). Machines Can Create Art, but Can They Jam? - Scientific American Blog Network. Scientific American Blogs. Retrieved October 15, 2023, from https://blogs.scientificamerican.com/observations/machines-can-create-art-but-can-they-jam/

[6] Torrance, S., & Schumann, F. (2019, March 9) Retrieved October 15, 2023, from https://doi.org/10.1007/s00146-018-0838-4

[7] Baxt, J. (2022, February 23). Computers in a Jazz Ensemble? Inventing Improvisational AI. UC IT Blog. Retrieved October 15, 2023, from https://cio.ucop.edu/computers-in-a-jazz-ensemble-inventing-improvisational-ai/

[8] Hillier, L., & Isaacson, L. (n.d.). Could music affect jazz? New York Jazz Workshop. Retrieved October 15, 2023, from https://newyorkjazzworkshop.com/could-ai-generated-music-affect-jazz/

[9] Bergamin, J.A. Being-in-the-flow: expert coping as beyond both thought and automaticity. Phenom Cogn Sci 16, 403–424 (2017). https://doi.org/10.1007/s11097-016-9463-1

[10] Schear, Joseph K. Mind, Reason, and Being-in-the-World: The McDowell-Dreyfus Debate. Routledge, 2013.

[11] Pellegrin, R. (2021, October 5). Why improvisation is the future in an AI-dominated world. The Conversation. Retrieved October 15, 2023, from https://theconversation.com/why-improvisation-is-the-future-in-an-ai-dominated-world-166724

[12] Ting, H. X. (2022, December 16). Scottish Fantasy' and AIVA: Machine Learning in Classical Music. Medium. Retrieved October 15, 2023, from https://medium.com/@kyrinstitute/scottish-fantasy-and-aiva-machine-learning-in-classical-music-8a3b78936418

[13] Vincent, J. (2016, October 12). A night at the AI jazz club. The Verge. Retrieved October 15, 2023, from https://www.theverge.com/2016/10/12/13247686/ai-music-composition-jazz-club-london-deep-learning

[14] Hogan, M. (2023, May 11). Musicians Are Already Using AI More Often Than We Think. Pitchfork. Retrieved October 15, 2023, from https://pitchfork.com/thepitch/musicians-are-already-using-ai-more-often-than-we-think/