

**Title:** “Can AI Understand Music?”

**Author:** Daniella Prado

**Mentor:** Professor Daniel Lopresti

**Abstract:**

This paper explores the complex nature of understanding music, examining two distinct definitions applied to both humans and Artificial Intelligence (AI). The first definition centers on technical proficiency, while the second focuses on emotional and spiritual connections. Human brain responses during music creation and perception are examined through fMRI studies, revealing cognitive patterns and emotional triggers. AI, exemplified by models like ChatGPT and Musilingo, showcases in music analysis, composition, and emotion detection. However, it struggles to express genuine empathy or emotions. The study concludes that AI, with its remarkable technical capabilities, has the possibility of understanding music emotionally in the near future.

### **Introduction:**

Understanding music has a different meaning for everyone. One definition for “understanding music” is to be able to read and learn it, as well as understanding compositional features which enhances appreciation (What Does it Mean to Understand Music?, 2022). By this definition, the epitome of a human that truly understands music would be a trained professional. Would this mean that an untrained music enthusiast does not understand music at all? A competing definition for understanding music states that this sense of understanding can only occur when one grasps the music’s intention and the listener has a spiritual or emotional connection to a piece (Hallam, S). Now, understanding music now opens up to the general public whose emotions can be evoked by a musical piece. These definitions explain how humans can understand music, but could the same be applied to Artificial Intelligence. Could AI understand music under the guidelines of the first definition, the second, both, or none?

### **Main text:**

#### **I. Humans “Understanding” Under Definition I**

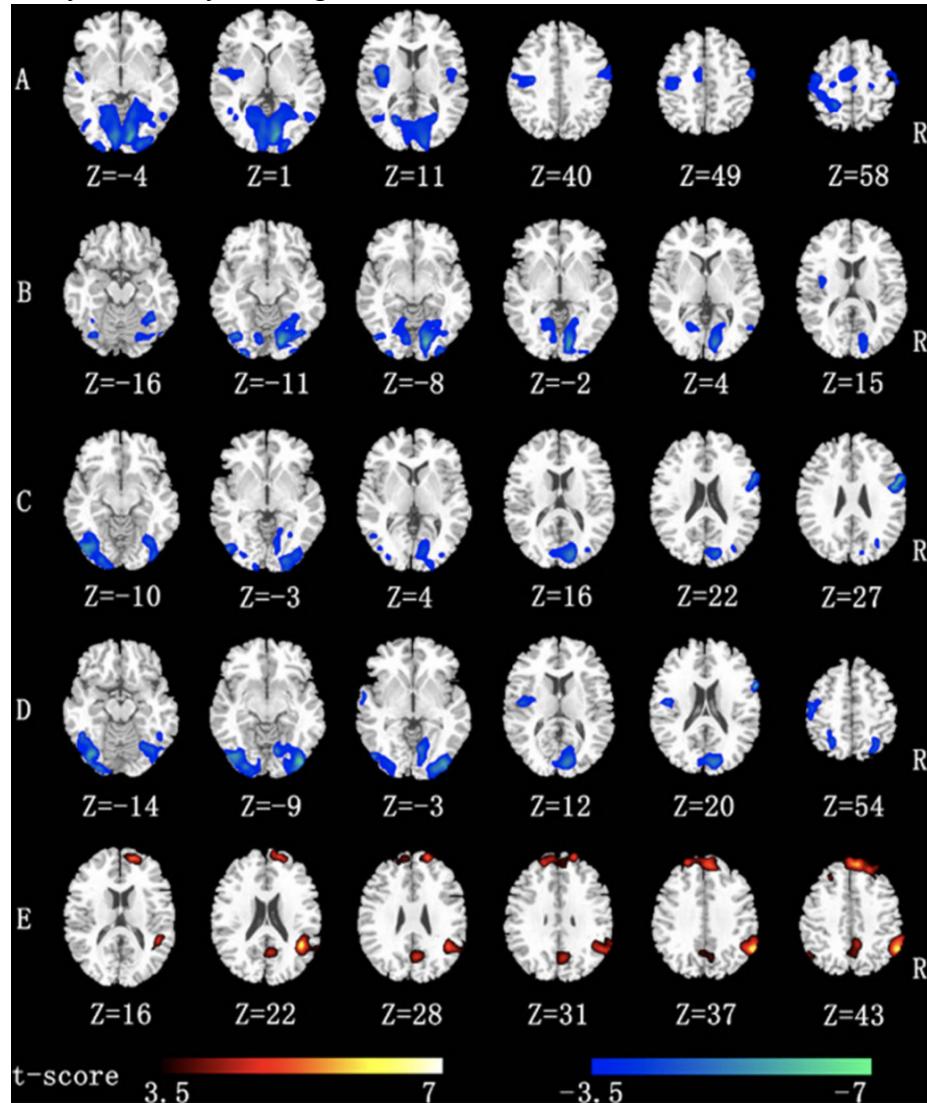
A person that can understand music by having the ability to read, learn, and understand compositional features could be a songwriter, music Professor, or musician. The way they have learned to appreciate music has been comparable to how one knows how to speak a language fluently. This deep appreciation is the result of various neurological changes and responses to be able to get to this level.

A study conducted by Scientific Reports leveraged fMRI to study how a music composer’s brain worked. Seventeen professional composers, each with over five years of experience in the industry, participated in the experiment. They were subjected to two sets of scans- one where they were in a “resting” state and a “composing” state. The study aimed to find what parts of the brain lit up during the composing state to get a sense of what characteristics allow humans to understand music at a technical level.

The results showed that during the composing state, the functional connectivity of primary networks such as the visual and motor areas decreased compared to the resting state. However, some networks, such as the Anterior Cingulate Cortex (ACC) (focuses on emotions and motivation) and Default Mode Network (DMN) (responsible for how the brain integrates information), showed stronger functional connectivity. This means that when professional composers were actively creating music, the areas of their brains associated with basic functions

like seeing and moving communicated less with each other. Areas related to attention and daydreaming were more active and coordinated (Lu, J. et al, 2015).

This finding contradicted previous studies that found visual and motor areas to be highly active when artists performed on instruments. The fact it is the inverse for composition highlights the complexity with music and the human brain. It is believed that composers have developed other unique ways to create music, so they rely less on their motor skills since they already have an idea of how a piece might sound without having to play every note. Similarly, the activation of the DMN was also surprising from the study. The DMN is not activated when a human thinks of a current task, but rather when there is something considered long-term or for the future. The reason for this could be that composition does not view writing as simply coming up with a single note - it implies someone must find a series of notes that go well together to create a harmony or melody. The figure below shows the results of the fMRI scans.



**Figure 3.** Results of functional connectivity analysis ( $p < 0.05$ , FDR-corrected, cluster threshold  $k > 600 \text{ mm}^3$ ). Row 'A' represents the significantly decreased functional connectivity of the left occipital seed. Row 'B' reveals that of the right occipital seed. Row 'C' represents the significantly decreased functional connectivity of the left postcentral seed. Row 'D' reveals the significantly decreased functional connectivity of the right postcentral seed. Row 'E' reveals the significantly increased functional connectivity of the left cingulum seed.

## II. AI “Understanding” Under Definition I

The way Artificial Intelligence is commonly thought to be able to understand music is through the more technical route the first “understanding music” definition describes. After ChatGPT was prompted to explain the musical elements of “Go Your Own Way” by FleetWood Mac, these were the results:

 **You**  
what are the musical elements that are used in "Go Your Own Way" by FleetWood Mac

 **ChatGPT**  
"Go Your Own Way" by Fleetwood Mac is a classic rock song known for its catchy melody and memorable lyrics. Here are some of the musical elements found in the song:

1. **Chords and Harmony:**
  - The song is based on a series of chords that create the harmonic structure. The harmonies used in the vocals and instruments contribute to the overall sound.
2. **Melody:**
  - The vocal melody, particularly in the chorus, is a defining feature of the song. Lindsey Buckingham's distinctive vocal style adds character to the melody.
3. **Instrumentation:**
  - The song features the standard rock band instrumentation, including electric guitars, bass, drums, and vocals. The use of acoustic and electric guitars contributes to the song's texture.
4. **Rhythm:**
  - The rhythm, driven by the drums and percussion, gives the song its energetic feel. The use of syncopation and dynamic changes adds to the rhythmic interest.
5. **Lyrics:**
  - The lyrics tell a story of personal relationships and emotions, reflecting the song's title "Go Your Own Way." The emotional content of the lyrics is an essential element of the song's impact.
6. **Vocal Style:**
  - The vocal delivery, with Lindsey Buckingham's distinctive voice and the harmonies provided by other band members (including Stevie Nicks), contributes to the song's signature sound.
7. **Guitar Solo:**
  - One of the standout elements is Lindsey Buckingham's guitar solo, which occurs in the latter part of the song. The solo is known for its expressive and emotive qualities.
8. **Arrangement:**
  - The arrangement of the song, including the build-up in intensity as it progresses, is an important element. Changes in dynamics and instrumentation contribute to the overall structure.

"Go Your Own Way" is a well-crafted song that combines these musical elements to create a timeless and memorable piece of rock music.

The results show that ChatGPT does have a very good understanding of elements that makes this specific song such a hit. Even the vocabulary it uses to describe these specific elements is very similar to what one would expect from a professional critic.

As self defined by ChatGPT, its purpose is to “generate human-like text based on the input it receives. It’s designed to understand and respond to natural language prompts in a coherent and contextually relevant manner”. There are OpenAI technologies designed to perform musical tasks, such as feature extraction, and composing.

### **I. Feature Extraction**

One groundbreaking use of AI in the music industry was the creation of the “Final Beatles record”. To create “Now and Then”, released in 2023, Paul McCartney used stem separation to extract John Lennon’s vocals from an old recorded version of the song (Marr, Bernard, 2023). Stem separation is a tool used in AI where a song is broken down into many components, or stems, and based on that the desired features are extracted. It is this same technology that allows AI to have such versatility in the process of creating songs.

Shortly after this article was released, another scandal surrounding AI and music emerged, except this one was met with various repercussions. Through AI, a fake track was generated using the voices of two of the most famous singers today: The Weeknd and Drake (The Learning Network. 2023, May 11). Upon the song’s viral spread, legal action was used to take down the song. This incident exemplified AI’s capability to create songs by utilizing stemming technology without the original author’s permission, showcasing its technical understanding of musical components.

### **II. Composing**

In addition to these stemming capabilities, AI also has the ability to compose music. In a recent article on Billboard, it was examined how AI could potentially revolutionize the music industry. A big emphasis on the article was placed in music production and composition. For example, it was stated that BandLab’s Songstarter is an OpenAI source that can create lyrics and a simple beat when prompted with emojis. Similarly, it was shown how other algorithms can create a personal theme song for users if prompted with a situation specific to the user. It went even as far to say that pitch technology can be used by AI to match a song that would fit an artist’s range or style (Leight, E.,2023).

A study from Mexico conducted a survey of AI’s compositional practices, investigating various composition methods shown in the figure below. After all the models were thoroughly analyzed, it was concluded that there was still no real measurement of success for music composition just yet. It was concluded that humans have subjective criteria in this sense, and are therefore unreliable. If an algorithm was trained to measure the success of a composition, it would be hard-coded, and therefore, be limited to only the information it has at hand. As such, AI for compositional purposes exists and is working, but its improvement and acceptance by the public cannot be quantitatively measured yet (O. Lopez-Rincon, 2018).

AI based music composition	
Area	Subareas
Soft computing based music composition methods	<b>Heuristic Composition Methods</b> (Evolutionary based Methods and Dynamic Programming)
	<b>Deep Learning Composition Methods</b> (Deep Belief Networks, Convolutional Networks, Recurrent Networks)
	<b>Stochastic Composition Methods</b> (Markov Models, Generative Adversarial Networks)
Symbolic AI based music composition methods	<b>Agent Composition Methods</b>
	<b>Declarative Programming Composition Methods</b>
	<b>Grammar Composition Methods</b>

### III. Humans “Understanding” Under Definition II

#### I. Brain Responses

The second definition of understanding music appeals to a more sensitive aspect for humans. Having an emotional or spiritual connection to a piece is something that most humans experience. There are certain songs that connect with an audience on a deeper level because of the situation, the artist, or just personal experiences.

Similar to the fMRI studies conducted a composer’s brain activity whilst composing, similar studies have been conducted specific to understanding an audience’s emotions to a song. In this study, subjects had to rate their mood after listening to a classical music piece. They had to categorize it as happy or sad, and briefly explain their reasoning behind it. While this is happening, subjects were under an fMRI machine that would track brain activity. The list below is the song selections for the study, accompanied by their categorization as “happy”, “sad”, or neutral. The mean rating is how many people categorized the song in the correct mood.

TABLE I. Musical pieces selected for the fMRI paradigm

Musical pieces	Composer	Mean ratings
<i>Happy pieces</i>		
Carmen: Chanson du toréador	Bizet	0.87 (0.001)
Allegro—A little night music	Mozart	0.90 (0.001)
Rondo allegro—A little night music	Mozart	0.65 (0.021)
Blau Danube	Strauß	0.84 (0.001)
Radetzky march	Strauß	0.77 (0.003)
<i>Sad pieces</i>		
Adagio in sol minor	Albinoni	0.80 (0.001)
Kol Nidrei	Bruch	0.60 (0.04)
Solveig’s song—Peer Gynt	Grieg	0.79 (0.001)
Concerto de Aranjuez	Rodrigo	0.62 (0.03)
Suite for violin & orchestra A minor	Sinding	0.80 (0.001)
<i>Neutral pieces</i>		
L’oiseau prophète	Schumann	0.56 (0.02)
Claire de lune	Beethoven	0.57 (0.05)
Claire de lune	Debussy	0.79 (0.001)
Symphony no. 2 C minor	Mahler	0.60 (0.002)
La traviata—Prelude to the 1st scene	Verdi	0.56 (0.30)
Pictures at an exhibition	Mussorgsky	0.63 (0.01)
Water music—passepied	Händel	0.57 (0.02)
Violin romance no. 2 F major	Beethoven	0.54 (0.02)
Water music—minuet	Händel	0.57 (0.02)
The planets—Venus	Holst	0.74 (0.001)

The results showed a few key findings, one of them being that happy songs had a higher “retention” or classification rate than the neutral or sad songs. This means that overall, happiness

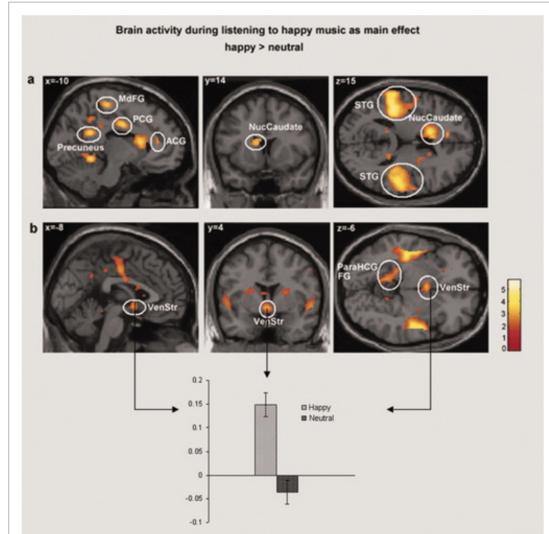
was the emotion that was easiest to detect. Sadness and neutrality seemed to be interchangeable by some of the subjects. It was also mentioned in the study that gender played no distinction during these classifications. Overall, men and women had the same abilities in distinguishing between songs and emotions.

Another key finding from the study was what parts of the brain showed the most responsiveness depending on the song being played. Figure A shows a participant's brain activity during a happy song, whereas Figure B shows the same participant's brain activity during a sad song. Different parts of the brain light up in each scan.

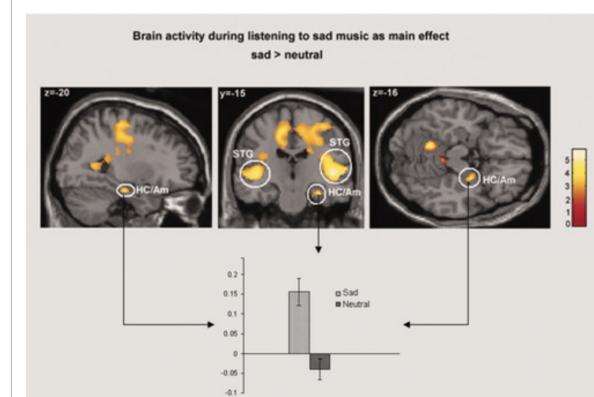
Happy music resulted in more activation in the ventral striatum and the dorsal striatum. The ventral striatum is activated during reward responses. This means that the brain identifies happy music as a type of reward, and this releases dopamine. It is not the first time it has been uncovered that happy music is associated with being a part of a reward system. In nature it is shown by mothers singing to their children and then responding happily, or children learning music in school. The dorsal striatum is associated with motor functions. When this is activated by happy music, it is associated with any sort of movement such as dancing, or even simple tapping to the beat. It is also associated with physiological arousal, as well as attention, recognition of musical melody, and mental imagery.

Sad music however, had higher activation in the amygdala and the hippocampus. These are colloquially known as the ones responsible for "fight or flight" responses. These are the parts of the brain responsible for fear and trauma responses. The study argued the high activation in these areas could be attributed to humans trying to evaluate this music deeper since sadness is a more complex emotion to understand for humans. This theory would also explain the higher attention rate found in people who listened to sad music. However, it has been previously argued that sad music could be attributed to euphoric reactions, so it may be a possibility that brain activation during sad musical pieces differs from brain activation during genuine sadness (Mitterschiffthaler, 2007).

**Figure A:** Activity during a happy song



**Figure B:** Activity during a sad song



## **II. Artist Connection**

Another important aspect to a human's connection to a specific artist. In an article by the New York Times, students were asked their opinions on AI generated music. Though most admitted they would listen to it out of curiosity, most said they felt a human connection was missing from the song (The Learning Network, 2023). British researchers tried to solve this phenomenon by studying why teenagers liked "problem" music as it emerged. Problem music was considered rap or heavy rock. This was during the early 2000's when parents were concerned this music was having a poor influence on kids, and the goal was to find out if it was negatively impacting them. For this study, undergraduate students were asked what type of music they enjoyed the most, and then participated in a psychological evaluation.

During the study, a theory the researchers had was that since some teens viewed these artists as their heroes or role models, they would want to model the values or behaviors they projected in their songs. The study revealed that while people who do enjoy problem music more are more prone to deviant behaviors and have a higher level of psychotism than those who preferred other genres (North. A, 2005). What was interesting about this study was that although these teens do feel a strong personal connection to the artists they like, it is not the sole factor in them enjoying this type of music. However, there was a big emphasis at the beginning of the study that the connection between an artist and the audience is important.

Amidst the COVID pandemic in 2020, one of the biggest struggles for artists was connecting with their audiences. They were no longer able to perform for live audiences, so remote shows started to gain popularity. However, audiences were hesitant to spend money to essentially watch a video of an artist performing. The University of Japan researched this phenomenon that same year, and investigated how Virtual Reality, VR, could be enhanced to improve virtual concert experiences. In the VR used, an avatar of the user would be placed in a venue where the concert would be, and other users could join in real-time.

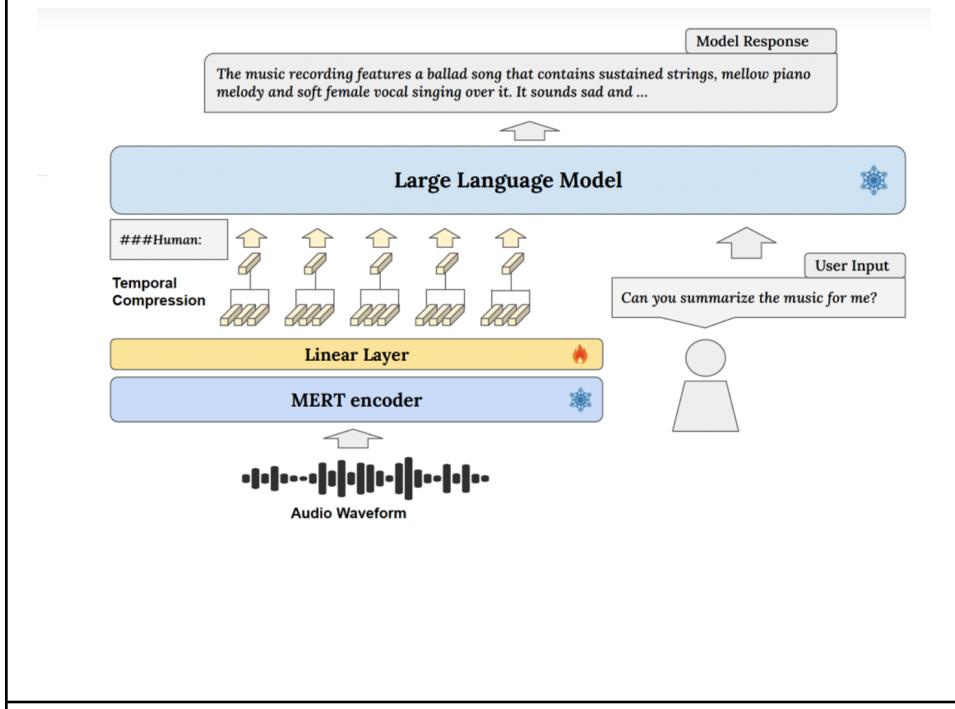
The four solutions explored in this study were: "copying the user's own movements, copying other users' movements , repeat beat-synchronous movements, synthesizing machine-learning based movements". Ultimately, it was found that by having avatars in the venue, these became too distracting for users. They would move to try to mimic others, or would get too distracted by their own avatars. The study understood that the shared concert experience was important for audiences to connect with an artist and with each other, but VR might not be the best approach to achieve this (Yakura, H. 2020).

## **IV. AI "Understanding" Under Definition II**

One of the big debates about AI nowadays is its lack to generate emotions. If ChatGPT was prompted with a question such as "How are you feeling today?" the response will most likely be "I do not have feelings" and redirect the question to the user. This is a common response for AI when faced with a question they are unsure how to answer.

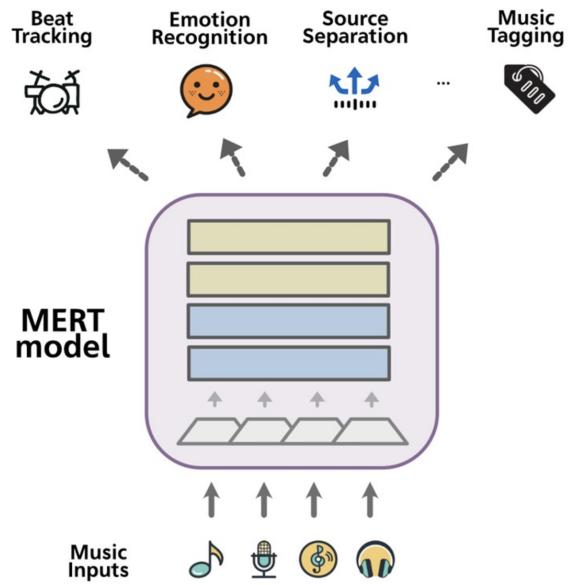
Though AI algorithms do not innately feel emotions, they are able to identify cues for certain emotions. New technology such as Musilingo has shown that even if AI cannot come up with an emotion by itself, it can identify them. Musilingo is a pre-trained language model used to describe the feelings behind a song, as well as the visuals and symbolism of a music video. Musilingo, shown by Figure A, makes use of MERT (Figure B) and Vicuna for closed captioning and answering questions the user might have (Ma, Y. 2023).

**Figure A: Musilingo Architecture**



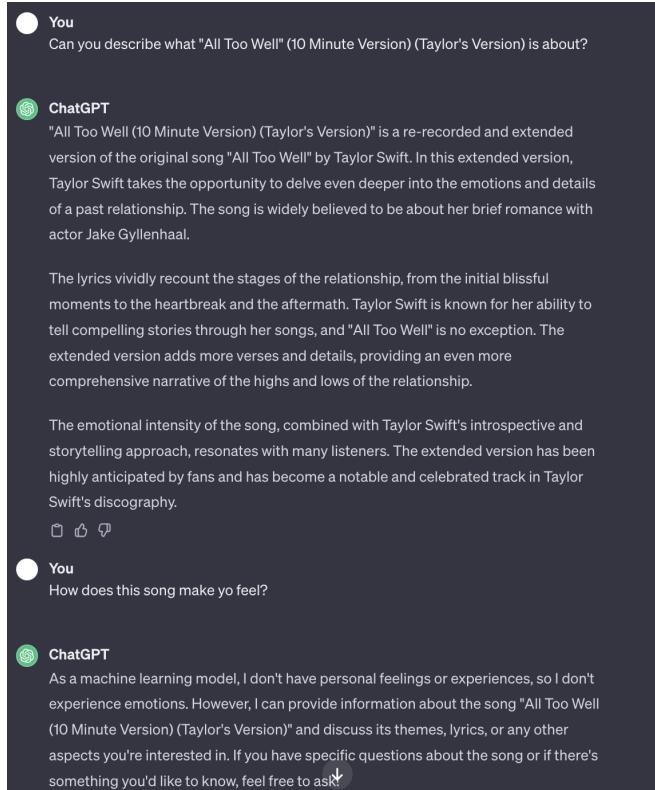
**Figure B: MERT model**

**MERT: one lightweight model for general music understanding** ✨



Similar to this, there are also several other AI models that serve this same purpose through a different model. Though these technologies have revolutionized the music industry, they are still flawed. Similar to ChatGPT, if prompted with a question they would not know the answer to, such as “Why do you think this artist was heartbroken when writing this song”, the algorithms revert the question back to the user, or simply say they are unsure.

In a small experiment using ChatGPT to comprehend a sad song, the algorithm demonstrated proficiency in understanding the song's theme but failed to express the song's impact. Although ChatGPT recognized the sad undertone of the song, it lacked the ability to empathize with the emotions presented. The figure below shows the results.



## Discussion:

Understanding music is a multifaceted concept for both humans and AI. This paper investigated two different definitions of understanding music, and applied each definition to both humans and AI. The complexities of what was found are the following.

### I. Understanding Music: Definition One Findings

Exploring both human and AI's understanding of music through this definition focused on a technical aspect. With this definition, one who has a deep understanding of music might be able to create it. It was revealed through fMRI scans of various composers that the brain activity shown was very different from someone's who does not compose. Unique cognitive patterns have been developed by these individuals to be able to create something new. Something noteworthy about the study's results was the increased activity in attention or daydreaming and decreased activity in motor skills. This emphasized how composing relies on a creative mindset. Artists such as Taylor Swift have tried to explain their songwriting process in many ways, but it all seems to depend on multiple factors. The last three tracks of Swift's 1989 album are voice memos of her explaining her songwriting process for three different songs on the album. She

explained how she's written to a melody, track, and to shouting out ideas. Swift did not use a formula to create any of these songs. At this point in her career she had a lot of experience in songwriting, but the creative process she underwent to develop these tracks was just as important as the technical skills she would use to compose and produce them.

AI algorithms also contain neural networks (NN) of their own that can also create music. The creation of "The final Beatles record" showed that through stemming technologies, AI can create music by retrieving old information. Similarly, the creation of The Weekend and Drake song is also an example of how stemming can be used to create something new. Still, there is a stark difference between a composition Swift could create compared to the one AI could create. Swift can think of a melody or a lyric by herself and work at trying to make it into a tune. Though AI does have the capability of creating a song, it needs to stem the components from somewhere else. AI needs to be prompted to know what to write, as it does not have an equal emphasis on creativity as technicality as others do.

## **II. Understanding Music: Definition Two Findings**

Interpreting music through emotive responses involves understanding the relationship music, emotions, and brain responses have in humans. Similar to the previous definition, this aspect of understanding music was also explored through fMRI's. In these, brain activity was tracked when subjects would identify a song to be happy, sad or neutral. The principal discovery of this was that different parts of the brain are activated depending on the song's mood. This also goes to show how a human's brain is prepared to accept, understand, and feel certain emotions. A human's brain already has the tools it might need to tackle an emotion quickly, regardless of *why* they could be feeling it. Songs sometimes convey a strong message that can evoke emotions, but there might be other factors that come into play. Music might stir emotions because of a personal connection or admiration for an artist, or a strong memory that is linked specifically to a song. As humans, we are able to sympathize with artists or situations at times, and this emotion can be transmitted. Therefore, it is possible for humans to have strong emotional or spiritual experiences when listening to certain pieces of music.

For AI, the key lies in emotion detection. It is already known AI does not have the ability to have emotions of its own (yet). While AI models like Musilingo demonstrated the capability to identify emotional cues in songs and music videos, they fell short in expressing or empathizing with these emotions. Still, they are able to identify why something would be considered upsetting or not even if the AI does not directly understand this emotion. Similar to Musilingo, ChatGPT has shown in various occasions it does not perceive questions regarding feelings in a normal way. Still, because the AI is designed to pass the Turing Test, it simply redirects these confusing questions back to the user. Despite this limitation, AI's ability to identify and respond to emotional aspects, opens the doors for a lot more research and development in this aspect of the technology.

### **Conclusion:**

In conclusion, the concept of understanding music is intricate, involving technical knowledge and emotional understanding. AI demonstrates remarkable technical knowledge in music creation, but struggles with truly understanding complex human emotions. As such, under the first definition provided, AI does understand music. This has been proved through its deep technical understanding of music components that would require a human to have professional training. Though there is still room for improvement, AI does understand music under the second definition provided to an extent. The technology does not personally undergo an emotional change during a song, but it can identify some of these emotional features, and that is a start.

## References:

- Hallam, S., & Papageorgi, I. Conceptions of Musical Understanding. 1-38, [https://discovery.ucl.ac.uk/id/eprint/1537697/1/Hallam\\_UCL%20Musical%20Understanding.pdf](https://discovery.ucl.ac.uk/id/eprint/1537697/1/Hallam_UCL%20Musical%20Understanding.pdf)
- Leight, E., & Robinson, K. (2023, August 4). 5 Ways AI Has Already Changed the Music Industry. *Billboard*. <https://www.billboard.com/lists/ways-ai-has-changed-music-industry-artificial-intelligence/>
- Lu, J. et al. The Brain Functional State of Music Creation: an fMRI Study of Composers. *Sci. Rep.* 5, 12277; doi: 10.1038/srep12277 (2015).
- Ma, Y. (2023, November 10). Pretrained models for music captioning and query response. Aihub. <https://aihub.org/2023/11/10/pre-trained-language-models-for-music-captioning-and-query-response/>
- Marr, Bernard. (2023, June 14). AI Creates Final Beatles Song: What Does It Mean for the Future of Music? *Forbes*. <https://www.forbes.com/sites/bernardmarr/2023/06/14/ai-creates-final-beatles-song-what-does-it-mean-for-the-future-of-music/?sh=74e9a247546a>.
- Mitterschiffthaler, M.T., Fu, C.H.Y., Dalton, J.A., Andrew, C.M. and Williams, S.C.R. (2007), A functional MRI study of happy and sad affective states induced by classical music. *Hum. Brain Mapp.*, 28: 1150-1162. <https://doi.org/10.1002/hbm.20337>
- North, A. C., Desborough, L., & Skarstein, L. (2005). Musical preference, deviance, and attitudes towards music celebrities. *ELSEVIER*, 1903–1914. <https://doi.org/xxxx>
- O. Lopez-Rincon, O. Starostenko and G. A. -S. Martín, "Algoritmic music composition based on artificial intelligence: A survey," *2018 International Conference on Electronics, Communications and Computers (CONIELECOMP)*, Cholula, Mexico, 2018, pp. 187-193, doi: 10.1109/CONIELECOMP.2018.8327197.
- The Learning Network. 2023, 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>
- What Does it Mean to Understand Music?, Sintelli App, 2022. <https://sintelly.com/articles/what-does-it-mean-to-understand-music/>
- Yakura, H., & Goto, M. (2020). Enhancing Participation Experience in VR Live Concerts by Improving Motions of Virtual Audience Avatars. In *Proceedings of the 2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. University of Tsukuba, National Institute of Advanced Industrial Science and Technology (AIST), Japan.