THE MONSTER: ASSESSING THE NEGATIVE OPINIONS ON AI GENERATED MUSIC

A Research Paper submitted to the Department of Engineering and Society

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

Shannon Lankford

Spring 2024

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Advisor

Kent Wayland, Department of Engineering and Society

Introduction

One of the recent concerns brought about by modern technology in the creative world is the rise of artificial intelligence, and its ability to develop artwork. In the world of music, AIs such as SOUNDRAW, Suno, Loudly, and Musicfy are able to take prompts and generate songs that fit the user-entered description. AI generated music has been met with a lot of backlash from members of the music industry, some even going so far as to back an open letter sent to Medium.com from the Artist Rights Alliance, calling for AI companies to "pledge that they will not develop or deploy AI music-generation technology, content or tools that will undermine or replace the human artistry of songwriters and artists or deny [artists and songwriters] fair compensations for [their] work." (Artist Rights Alliance, 2024). While many of the arguments against AI are centered around the fiscal loss that the artists take and the potential copyright violations, there is also the idea of undermining human artistry, as referenced in the open letter. This idea of "undermining human artistry," and why that poses an issue for many can be evaluated and explained under Smits's Monster Theory.

Background and Context

Artificial Intelligence has been making headlines all over the world for its various advancements, as well as fears about AI taking over industries. AI has four primary types: Reactive, Limited Memory, Theory of Mind, and Self-Aware (Johnson, 2020). Reactive AI, when given an input, will deliver the same output every time without learning, like the Spotify recommendations. Limited Memory AI, which is the most advanced kind that currently exists, is able to adjust actions based on experience, unlike reactive AI. Theory of Mind AI and Self-Aware AI do not yet exist but would be more capable of emotional intelligence and a more

developed decision-making process, likened to the ones that humans have. Since Theory of Mind AI and Self-Aware AI do not yet exist, this paper will focus on Limited Memory AI.

One of the most popularized and advanced cases of AI music was in May 2023 when an AI generated song called "Heart on My Sleeve" that used AI to replicate the voices of the music artists Drake and The Weeknd, creating original lyrics and melody (The Learning Network, 2023). This sparked conversations and provoked analysis of how the public, and not just the artists who have strongly spoken out against it, view AI generated music. In a survey conducted by The Learning Network, they found that most people interviewed believe that AI generated music does not have the potential to replace what many referred to as "real music" (The Learning Network, 2023). While this example deals with the potential legal issues with AI music and prompted public conversation about AI generated music that allowed for public opinions to be compiled, it does not influence the topic of this paper. This paper deals with the creation of music solely using AI. At its current state, it has not become commonplace to see AI generated music being used in various settings. It has been mainly used as a tool for composers to take unique musical loops created and use them as inspiration or as a starting point for a song.

In an attempt to analyze the negative response to AI-generated music, Martijnte Smit's Monster Theory can be applied to the theory. Smit's theory outlines an explanation of how public perception can lead to differing views of new technologies, which applies to AI-generated music seamlessly. The Monster Theory argues that this technology comes to its most controversial when it combines two tenants of society that are seen as mutually exclusive, such as living and dead or nature and culture. This fusion can create both adverse and positive feelings about technology in a human. Some view technology as innovative, and they become optimistic about the potential positive effects this technology can have on society by fusing these norms of

society. Under Smit's theory, these people have a "monster embracing" perspective, and support or even advocate for these new technologies. Others fear this technological fusion because they view it as unnatural and contradictory to the world they know. Under the Monster Theory, this group is referred to as having a "monster exorcism" perspective and become apprehensive or even become adversarial to the new technological developments (Smits, 2006).

Since it is widely understood that most innovative technologies will find most people under the "monster exorcism" perspective, the Monster Theory also details how these technologies can become more widely accepted in society. These two potential approaches are called "monster adaptation" and "monster assimilation." Monster adaptation involves slightly changing the technology to place it more firmly into one cultural category rather than making it a fusion. This makes it fall more into the current world order and curbs the fear of the technology exhibiting a massive change in societal norms. Monster assimilation takes a more radical approach by requiring cultural categories to be changed to fit this new technology, changing the perspectives of society by fully changing the fabric of society (Smits, 2006).

Smits' Monster Theory offers an insightful framework for analyzing the causes of polarized responses to AI-generated music of embryos because this technology is perceived as fusing the cultural categories of organism and machine. "Organism" is used in this context to refer to the capabilities of humans, with original thoughts, feelings, and autonomy, and "machine" is used in this context to refer to the idea of human-made operators that execute commands sent by an organism (Smits, 2006). For the analysis of the negative opinion of AI-Generated music, the Monster Theory will be used to highlight reasons for objection and evaluate its potential to be integrated into society, and to see if that potential hinges on the state of the art.

When evaluating the current "state-of-the-art," as it references A-generated music, one of the important things that is evaluated is the AI's ability to convey emotion through music. Many of the most listened to songs in history deal with deep seeded human emotion that AI would need to be able to replicate in order to properly create a song. However, researchers have been able to mimic these facets that convey emotion in songs by using machine learning tools that evaluate the harmony, melody, and rhythm of the song (Imasato et al, 2023). While these are not how we evaluate emotion outside of the realm of song, these are tools that have been used by artists to project a certain emotion onto a person, which is important not only in the music industry for recording artists, but it is also used by people attempting to use music to curate an energy, such as movie producers, restaurateurs, and fitness instructors. However, emotions conveyed in a song is hard to assess because individual people have their own emotions and interpretations of said emotions within songs (Imasato et al, 2023).

Another aspect of AI music generation that is used to evaluate its "state-of-the-art" is the originality of the music, deciding whether this system that uses existing music to generate new music can ever be "original." The idea of "music ownership" can be difficult to define, as there has been many cases, especially in recent years, where musicians will sue other musicians because they believe that a work sounds too similar to a work they have done that predates the other's. In AI, Generative Adversarial Networks (GANs) have been used to combat this problem. GANs have two distinct parts: a generator that creates the data and a discriminator that evaluates the authenticity of the data (Yu et al, 2023). Another facet of music being "original" other than copyright is the idea that it is different from anything else they had heard and can foster creativity rather than churning out all similar sounding music. Both the originality of the music

and the ability to convey emotion through the music are aspects typically associated with human artistry, which shows a potential fusion between organism and machine.

Methods

To analyze the data that currently exists rather than creating a new data set, two studies were selected that were blind studies comparing AI generated music to composer-made music (or manmade music). One of these was a study that used two different types of AI music generation models: unconditional, where the music was generated without a prompt to convey, and conditional, where there was a prompt to cohere to (Chu et al, 2022). For the purposes of this paper, the conditional music generation models were used to more accurately measure the effectiveness of AI generated music in conveying an emotion to a listener. The other study that was used compared classical string quartets (CSQ) and classical piano improvisation (CPI) to the output generated by an AI music generator using the data from the CSQ or CPI (Yin et al, 2023). For the purposes of this paper, CPI was used, as well as Translational GANs because those two variables best matched the variables already provided with the first study found.

To analyze public opinion on AI-generated music in an attempt to gauge the negative opinion in the context of "undermining human artistry," the two different studies that were evaluated included one that used evaluation data from fifty participants with a high-level of knowledge about music (Yin et al, 2023) and one that used evaluation data from fifty participants (although only forty were deemed "valid") declared at a "novice" level of music knowledge (Chu et al, 2022). Both studies were used in this instance to create a comparison between participants that can take both industry expert's negative perception and see if the same reasonings and results appear across the board, or if the experts have a different interpretation of the music than

"novices" do when listening to the AI-generated music. Both studies were conducted by having the participants listen to multiple songs, some generated by the same AI music generator (Translational GANs), and rank them based on certain metrics, both using the seven-point Likert scale. The studies had different metrics for evaluations, but five metrics for each study were found to be comparable to a metric in the other study based upon the definitions given in each study.

Metric, (Yin et al, 2023)	**Metric, (Chu et al, 2022)**	**Definition**
Stylistic Success	Coherence	The extent to which it adhered to the reference given.
Aesthetic pleasure	Naturalness	The measure of how "pleasing" or "expressive" the piece is, evaluating how it conveys emotion.
Repetition or self-reference	Structureness	The reuse of notes, melodies, harmonies, and rhythm throughout the piece
Melody	Melodiousness	The organization of notes and pitch. A successful piece is organized and recognizable.
Rhythm	Rhythmicity	The extent to which the music has a unified rhythm.

Table 1: Metric Comparison between (Yin et al, 2023) and (Chu et al, 2022) and the definition of the metrics

Results

Metric	Yin et al, 2023 (Professional)			Chu et al, 2022 (Novice)		
	Composed	*AI*	*Difference*	*Composed*	*AI*	*Difference*
Coherence	4.640 ± 2.027	2.980 ± 1.655	-1.660	5.400 ± 1.248	4.192 ± 1.802	-1.208
Naturalness	5.340 ± 1.531	2.720 ± 1.613	-2.620	5.100 ± 1.453	3.700 ± 1.645	-1.400
Structureness	5.660 ± 1.336	3.060 ± 1.618	-2.600	5.375 ± 1.284	3.892 ± 1.512	-1.483
Melodiousness	5.580 ± 1.415	2.720 ± 1.600	-2.860	5.442 ± 1.368	3.867 ± 1.695	-1.575
Rhythmicity	5.360 ± 1.480	2.800 ± 1.720	-2.560	5.467 ± 1.276	3.792 ± 1.698	-1.675

Table 2: Likert Scale scores for the five metrics defined in Table 1, compiled from all users and song variants

The table above (Table 2) depicts the average Likert score, and the standard deviation of the scores, from the study with the professionals (Yin et al, 2023) and the study with the "novices" (Chu et al, 2022). From analyzing the data, while the Likert scale scores between the two studies for the composed works were comparable with one another, the AI scores compiled from each study were not, with a greater deviation from the Likert scale score from the composed score being observed when comparing the AI Likert scale score in the professional study. When considering the differences of the individual metrics between the Likert scale score of the composed song vs. the AI song within their own study, there appears to be no significant difference between the individual scores, especially considering the large standard deviation that accompanies the average scores. However, the only outlier to this appeared to be the difference

in the coherence metric in Yin's score, which is almost a whole unit difference from the rest of the metrics' differences.

Since both of these were blind studies, meaning that the participants didn't know which was AI generated music and which was user generated, the results seem to provide evidence to suggest that the distaste for AI generated music does not solely stem from the knowledge that the music is not made by a human but rather there is something that feels intrinsically "off" about the music. This becomes even more evident when comparing the Likert scale score differences between the professionals and the "novices." As the professionals have been exposed more to music and have a greater schema by which to compare and evaluate the music in, anything that contradicts that schema will be more likely to be seen as "wrong" rather than "different" due to the vast knowledge of these professionals.

Using the framework of Smit's Monster Theory, it appears that the overarching view from both studies conducted is one of "monster exorcism." The differences in the AI generated music versus the music created by a composer that followed the typical conventions of music that the participants were used to caused them to rank the AI generated music lower, across the board, for all aspects of the music. The AI music generator takes the music that already exists and generates something unique from it, which does not guarantee that the songs will follow standard conventions and detect emotion or purpose in things like rhythm and melody. This creates this "monster" through a combination of the human emotions and creativity conveyed in music and the machine that churns it out. The creativity and emotion can be interpreted through the metrics that were set from both studies, with rhythm, melody, and structure being formulas used to convey emotion and add individual creativity to a song, and the ideas of "naturalness"

and “coherence” being how the user is able to define how they feel about the song and the emotions that they are able to gather through it.

In tandem with evaluation under Smit’s Monster Theory, a question is posed: will AI-generated music be best implemented into society using a “monster assimilation” approach or a “monster adaptation” approach? In this case, a “monster assimilation” approach would mean societal acceptance that the overarching umbrella of art, and not just music, can be generated by machine without the direct hand of organism interaction. This would mean a redefinition of what art is, since its current definition is rooted in the idea of it being human creative skill rather than just any piece of media, like a song or artwork. Since art has been an integral part of society since its very beginnings, there is a low chance that art will be redefined to create a subspace between organism and machine that will allow for “monster assimilation.” In the words of author Leo Tolstoy, “Art is a human activity consisting in this, that one man consciously, by means of certain external signs, hands on to others feelings he has lived through, and that other people are infected by these feelings and also experience them” (Tolstoy, 1897). This quote depicts a commonly-held definition of art as it remains rooted in the idea of humanity. However, this does not mean the complete abandonment of AI generated music. This merely means that a “monster adaptation” approach may be more feasible in this scenario. A potential approach to “monster adaptation” would be to program in more guidelines or structures within the AI music generator that will allow for generated music that fits more into the current societal bounds of “organism” and “machine” while providing less crossover. With a more guided approach to generation, this “monster” will become less of a fusion “organism” and “machine” and maintain a more currently held relationship wherein a heavy influence from the organism allows the machine to exhibit the traits and emotions organisms can. Another potential “monster adaptation” approach would be to

continue to use AI generated music as a tool to generate ideas or basic song structures. This current utilization of it has become widely accepted amongst music professionals, although not used by all. This “monster adaptation” keeps the final creative control held by an organism rather than held by just the machine, which solidifies the divide between “organism” and “machine.” A “monster adaptation” approach for AI generated music is likely to be accepted due to the current bridges of art that exist today between “organism” and “machine” including photoshop and web drawing applications.

Conclusion

This paper dove into the complex landscape of AI-generated music, and its reception by both industry experts and novices. Applying Martijnte Smit’s Monster Theory to the analysis, it becomes apparent that AI-generated music occupies a space that challenges the traditional cultural categories of “organism” and “machine,” blending elements of human creativity with machine-generated output. This fusion evokes reactions of apprehension, as seen and detailed from the Artist Rights Alliance’s open letter, as well as the public dislike of the AI generated Drake and The Weeknd song, reflecting a "monster exorcism" perspective among participants. Blind studies showed that, even when it is not known that the music is AI generated, there is still a lowered “acceptance” of the songs. The findings, as well as the evaluation of how art fits into our society, suggest that for AI-generated music to find broader acceptance, it must find a way to adapt to or change the bounds of “organism” and “machine” that are currently accepted today. The question arises whether to pursue a "monster assimilation" approach, redefining societal perceptions of art to include machine-generated works, or a "monster adaptation" approach, wherein AI systems are refined to align more closely with existing cultural norms. While a complete redefinition of art seems unlikely, incremental adjustments to AI music generators

could bridge the gap between human artistry and machine generation. Furthermore, the potential for AI-generated music lies not only in its role as a standalone creative output but also as a tool for inspiration and idea generation within the music industry. By maintaining a balance between human creativity and machine assistance, AI-generated music can complement traditional artistic processes rather than supplanting them entirely. In essence, this paper underscores the intricate interplay between technology, creativity, and societal norms in the realm of music composition. As AI continues to advance, navigating these complexities will be crucial in shaping the future landscape of music creation and consumption.

Resources

Artist Rights Alliance. (2024, April 23). *200+ artists urge tech platforms: Stop devaluing music*. Medium. https://artistrightsnow.medium.com/200-artists-urge-tech-platforms-stop-devaluing-music-559fb109bbac

Chu, H., Kim, J., Kim, S., Lim, H., Lee, H., Jin, S., Jin, S., Lee, J., Kim, T., & Ko, S. (2022, October 1). *An empirical study on how people perceive AI-generated music: Proceedings of the 31st ACM International Conference on Information & Knowledge Management*. ACM Conferences. https://dl.acm.org/doi/10.1145/3511808.3557235

Imasato, N., Miyazawa, K., Duncan, C., & Nagai, T. (2023). Using a language model to generate music in its symbolic domain while controlling its perceived emotion. *IEEE Access*, 1–1. https://doi.org/10.1109/access.2023.3280603

Johnson, J. (2020, June 8). *4 types of artificial intelligence*. BMC Blogs. https://www.bmc.com/blogs/artificial-intelligence-types/

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

Smits, M., Achterhuis, H., Kousbroek, R., Midgley, M., Soper, K., Swierstra, T., Willigenburg, V., Winner, L., Douglas, M., Meikle, J., Yarsley, V. E., & Carson, R. (2006, November 13). *Taming monsters: The cultural domestication of new technology*. Technology in Society. https://www.sciencedirect.com/science/article/abs/pii/S0160791X0600039X

Tolstoy, L. graf, & Maude, A. (1897). *What is art?* Project Gutenberg.

Yin, Z., Reuben, F., Stepney, S., & Collins, T. (2023, March 21). *Deep learning's shallow gains: A comparative evaluation of algorithms for Automatic Music Generation - machine learning*. SpringerLink. https://link.springer.com/article/10.1007/s10994-023-06309-w

Yu, X., Ma, N., Zheng, L., Wang, L., & Wang, K. (2023). Developments and applications of Artificial Intelligence in music education. *Technologies*, *11*(2), 42. https://doi.org/10.3390/technologies11020042