

Striking a Chord: The Viability of AI-generated Pop Music Among Generation Z

Srida Ayeireddy

Independence High School

Abstract

The emergence of generative AI in the music industry has presented new possibilities but also raises questions regarding its reception and competitiveness with human-produced music. Thus, this study explores Generation Z's perception of AI-generated pop music created by Udio, an AI model utilizing Generative Adversarial Networks (GANs), in comparison to human-produced pop music on the basis of five key attributes: creativity, emotion, authenticity, replay value, and overall rating. Results show an overall preference for the human-produced sample, specifically for creativity, emotion, replay value, and overall rating. Qualitative data corroborated this pattern, indicating recurring concerns with dissonance, genericness, and balanced composition within the AI-generated sample, while the human-produced sample was recognized for its overall design and variability. While some participants recognized the innovative abilities of AI in music production, a majority expressed ethical concern for traditional musicians and the lack of emotional translation and creativity. The results of this study demonstrate that current AI-generated pop music generated through AI utilizing GANs does not yet rival human-produced pop music among a Generation Z audience.

Keywords: Computational creativity, generative-AI, AI-music generators, uncanny valley, human-produced music, competition, Udio, reception to AI-generated music, Generation Z.

Introduction

Recently, new advancements in machine and deep learning have allowed for the creation of AI-generated content: media “derived from patterns in existing works, [that] blurs the line between derivative and original

creation” in the forms of videos, pictures, artwork, and even music, potentially posing new opportunities for innovation within the industry (Bukhari et al., 2023). While these new tools can increase efficiency and provide new outlets for creative expression, the use of such advanced tools raises ethical issues regarding potential job displacement, specifically within media-related industries (Pugachev et al., 2023). Additional concerns include complications with copyright, authorship uncertainty, a reduced demand for human-produced music, and the potential shift towards a more homogenous pop music market that lacks creativity. As such, it is critical to evaluate the performance and reception of AI-generated music in comparison to human-produced music to ultimately determine what threat, if any, generative-AI tools pose to creative industries. Thus, this paper aims to compare the popularity of human-produced content to that of AI-generated content, specifically within the pop music industry.

With the recent introduction of AI-music generators and synthesizers, AI technologies can analyze large volumes of music data, extract certain patterns and features, and generate new music compositions based on a few input prompts (Briot & Pachet, 2018). Most current AI-music generators utilize Generative Adversarial Networks (GANs), a specific type of neural networks that consist of “a framework that uses an adversarial process to estimate generative deep learning models” (Chen et al., 2024), allowing the model to generate predictions, based on training data, to create a final musical composition. Thus, this study specializes in AI-music generators utilizing GANs, particularly Udio. Specifically, this paper’s research question is: “with the increasing use of neural networks in generative AI, how does the perception of pop music generated by AI-music generators utilizing GANs compare with that of traditional human-made pop music according to Generation Z consumers?”

Literature Review

AI-generated Music Process

Currently, there are various types of AI-music generators available, each with their own set of limitations and specializations. Jourdan and Caramiaux, in their study, analyzed large volumes of AI-music generators, delving deeply into each step of the process that an AI-generator typically carries out. However, they found that most AI-music generators follow a similar process in which the computer takes a set of data defining the task as input and proposes an output capable of performing this task (Jourdan & Caramiaux, 2023). More specifically, the system generates a final output through recognizing patterns and making predictions about a style learned from an arbitrary corpus of music, otherwise known as training data, according to Briot and Pachet (Briot & Pachet, 2018). However, this process can severely limit the amount of personalization and human input permitted in the final output. For instance, 54% of the publications in

Jourdan and Caramiaux's systematic review do not allow users to act on the training data or on the model parameters. While traditional human-produced compositions can be refined, many AI-music generators lack interactivity and require whole regeneration as opposed to the typical refinement of a composition (Briot & Pachet, 2018). As such, a significant portion of AI-music generators propose a "black box" system for users in which there is little room for personalization. This disparity can be attributed to the machine learning models' reliance on large datasets, inhibiting the deviation from certain styles and producing varied results (Jourdan & Caramiaux, 2023).

Generally, very few AI models actually "explain the reasoning behind their process, making it difficult to debug such models" (Bryan-Kinns et al., 2023). Thus, Bryan-Kinns and his colleagues suggest the implementation of eXplainable AI (XAI) features, where creative AI presents information that is meaningful to people – such as by "presenting visual cues between mappings, or visualizing levels of mutual trust with emoticons" to better streamline collaboration (Bryan-Kinns et al., 2023). However, there are some exceptions to this overarching outlook, namely within specific, newer models such as Mubert and DeepBach. These models have already integrated features, similar to those of XAI, that increase user interactivity and input, replicating the iterative process of producing traditional music. For instance, "[e]ach of Mubert's streams represents an ever-evolving "live" track that continually transforms" through the use of a "dislike" and "like" feature that allows live input from users, replacing the need for regeneration or to rewind (Pugachev et al., 2023). Likewise, DeepBach, another generative model, allows for "user-defined constraints such as rhythm, notes, parts, chords and cadence" (Hadjeres et al., 2017). This model allows for users to isolate specific attributes of a composition, enabling them to be more intentional with their choices and provide room for further experimentation. However, it is important to note that the choices within the DeepBach model are "very general and do not involve expert knowledge about harmony or scales but are only mere observations of the corpus," reiterating the prevalent role of training data in the variability of the final output as found by Briot and Pachet's findings (Hadjeres et al., 2017).

Capabilities of Various AI-Music

Furthermore, within AI-music generators each model has slightly different abilities and limitations. For instance, Chen, Huang, and Gou explain how each model has distinct limitations as "symbolic generation models often lack expressiveness in timbre, and audio generation models struggle with long-range structural modeling," (Chen et al., 2024) hindering the quality of the final output. A potential reason for these limitations is that "[d]ue to copyright protection, many high quality music datasets cannot be publicly released" (Chen et al., 2024). Consequently, researchers often have access only to limited datasets, resulting in a lack of certain music styles in

research; this limitation prevents the research from being generalized to broader music domains and often results in “less diverse outputs or favoring certain styles” (Chen et al., 2024). Since the models have less training data to study, this limitation also hinders the algorithms’ ability to recognize patterns, potentially limiting complex AI-music generation only to genres with adequate training data.

It is crucial to note that there are some systems better suited at replicating human-produced music. For instance, AMG-systems (AI-based affective music generation systems) fine-tuned their emotional translation ability, invoking feelings of valence and arousal thought to previously be confined to “traditional” music according to Dash and Agres (Dash & Agres, 2024), potentially creating competition by improving on a primarily “human” aspect of music. However, this paper aims to compare final outputs developed by models utilizing GANs, a widely used and accessible type of neural network utilized for short-term structure (Figuera & Vaz, 2022), making them adept for the short, 30-second-long musical sample generated for this study. Thus, other AI-music generator models will not be delved into further, nonetheless it is important to note that models utilizing other neural networks such as CNNs and RNNs are present and an area for further research.

Computational Creativity vs Traditional Creativity

The number of publicly accessible AI-music generators has grown tremendously as companies continue to experiment with the variability, creativity, and quality of the produced output of AI-music generators. While the breakthrough of AI-music generators has lowered the cost and increased accessibility of music creation, it also poses potential impacts to pre-established musicians and artists. Creativity specifically has been an issue of contention in AI-generated music as computational creativity (CC), defined by Carnovalini and Rodà as actions computational systems exhibit that neutral observers would define as creative (Carnovalini & Roda, 2020) is difficult to directly compare to traditional creativity. For instance, not all music generation systems prioritize creativity as most “[systems are] simply designed to tackle tasks that are commonly considered to involve creativity (usually artistic tasks like painting or composing music), but the process involved in the generation of such output does not necessarily involve creativity,” making it difficult to assess the overall effectiveness of CC (Bukhari et al., 2022). Furthermore, there are also a wide diversity of goals regarding CC, for example, psychologists prioritize using CC to understand human creative processes while artists prioritize their use in expressing their own creativity (Carnovalini & Roda, 2020), making it difficult to directly compare the two in terms of competition on a standardized scale.

Reception to AI-generated Music

Instead of viewing AI-music generators as competition, many encourage their use as tools for musicians to enhance their own expression and creativity. For instance, in their academic article, Ijiga and their colleagues, argue that generative AI models expand the “horizons of creative expression by enabling artists to manipulate and modulate voices in unprecedented ways”, such as through altering vocal timbre, pitch, or emotion; these technologies serve as versatile tools for artists and increase inclusivity and diversity (Ijiga et al., 2024). By eliminating common barriers that inhibit music production, such as access to expensive equipment or a producer, AI-music generators fill the role of a collaborator, allowing for small musicians and even consumers to still express themselves creatively. Reinforcing this perspective, in their peer-reviewed article, Anantrasirichai and Bull predict that “AI will be adopted much more widely as a tool or collaborative assistant for creativity, supporting acquisition, production, post-production, delivery and interactivity,” rather than replacing human creativity entirely (Anantrasirichai & Bull, 2021). Similarly, Avdeeff, in her peer-reviewed article, builds off Anantrasirichai and Bull’s argument, stating AIPM (artificial intelligence pop music) software is akin to a producer, offering support as opposed to competition (Avdeeff, 2019). However, the training databases used by systems such as Flow Machines require large amounts of unbiased, diverse data, which can be difficult to obtain. Additionally, in terms of legal implications, ethical regulations and guidelines will need to be implemented “to ensure the integrity of developers and users” surrounding these technologies (Anantrasirichai & Bull, 2021). Namely, a legal system of consent and responsibility of the creator is required, along with clarification on intellectual property rights and ownership of the final work (Ijiga et al., 2024).

Furthermore, reception to AI-generated music in real-world scenarios has been mixed. In her peer-reviewed article, Avdeeff, discusses the first AI-collaborated album, *Hello World*, by SKYGGE, stating that Flow Machines, the software used, created a much more musically engaging work, “meant to blur the boundaries of conceptual and commercialized pop music” (Avdeeff, 2019). However, when tested with audiences, many had adverse reactions attributed to a noticeable dissonance within the final output. Avdeeff describes this phenomenon as an uncanny valley, where the “small discrepancies between reality and expectation are highlighted to the point of provoking unease” (Avdeeff, 2019).

The Gap

As indicated by the review of pre-existing literature, research regarding user experience with AI-music generators is prevalent. However, research regarding reception to the final outputs produced by these generators was limited, and even more so within the pop music genre. Notably, even the few existing studies, such as Avdeeff’s, have limitations; for instance, due

to the relative newness of AIPM (artificial intelligence pop music), at the time, her paper relied primarily on artificial intelligence music as a whole, rather than pop music specifically. Thus, this paper seeks to address this disparity by focusing on AI-generated pop music explicitly and generating a new sample utilizing an AI-music generator instead of relying on a predetermined database. Additionally, this study seeks to isolate the younger demographics of pop music, namely Generation Z, as research regarding this generation's reception specifically was lacking.

Assumptions

This paper assumes that conclusions can only be generalized to AI-music generation models utilizing GANs, as opposed to all AI-music generators. Additionally, conclusions drawn from this research assume that all participants are being truthful in their responses. To reiterate, this paper's research question is: “with the increasing use of neural networks in generative AI, how does the perception of pop music generated by AI-music generators utilizing GANs compare with that of traditional human-made pop music according to Generation Z consumers?”

Methodology

In order to compare the popularity of AI-generated music with the popularity of traditional human-produced music, this study utilized a non-experimental, causal-comparative, mixed research method to collect Generation Z's opinions on samples of AI and non-AI-generated pop music. While it may have been possible to conduct an experimental design for greater control over variables, feasibility and ethical concerns hindered the practicality of such a study. An individual's music preferences and perception are highly subjective as they are dependent on a wide range of confounding variables such as previous genre exposure, environment, and experiences. Thus, regulating these variables would be impractical and unethical in a fully experimental setting. Therefore, a non-experimental comparative research method was chosen as it allowed for data observation while reducing subjectivity and was designed to align with the causal-comparative nature of this paper's research question. While this approach limits the generalizability of results due to the small sample size and number of AI-music generators being used, it stimulates a real-world environment in which participants would naturally come across such music.

Research Instrument

The first part of this study's method utilized semi-structured interviews to gather insight on the viewpoints of Generation Z individuals on AI-generated music in comparison to human-produced music. A semi-structured interview, according to Dovetail, is defined as a

“qualitative research method used to gain an in-depth understanding of the respondent's feelings and beliefs on specific topics” through a conversation-like interview as opposed to a set number of questions in a specific order (“Semi-Structured Interview,” 2023). Additionally, utilizing semi-structured interviews allowed for the preparation of a few guiding questions ahead of time, and the freedom to create new follow up questions in real time based on popular subtopics that arose. Ultimately, the researcher felt that simply having participants state whether they “liked” or “disliked” each sample of pop music through a survey or other structured method would not fully address the research question, as it did not lend itself to identify what specific factors in each sample justified such responses.

Thus, instead of merely disliking or liking each sample of music, participants were able to fully explain their thought process freely. They also had room to make connections to various topics that were brought up while incorporating their own personal experiences that came up extemporaneously. Furthermore, the flexibility of semi-structured interviews allowed the researcher to clarify any potential doubts regarding a participant's response, reducing errors, the risk for misunderstandings/miscommunication, and increasing the quality of the qualitative data responses that surveys lacked. Semi-structured interviews were preferred as opposed to secondary data collection sources, such as a meta or content analysis, because they allowed for the collection of primary data on the viewpoints of Generation Z individuals.

Participants

Initially, small pop musicians were selected as the primary group of participants to further explore their opinions on AI-generated tools and competition regarding job displacement. However, due to feasibility concerns, this group of participants was later omitted due to inadequate sample size and feasibility issues. Thus, this study pivoted towards a more consumer-based approach focused on exploring the viewpoint of Generation Z on the increased saturation of AI-generated pop music and its viability compared to human-produced music. All 14 participants were classified as Generation Z between the ages of 13-28, as people within this age range were the primary target audience for the pop music genre specifically. This group was selected as Generation Z is the primary target audience for pop music, and research regarding this specific demographic was limited, as addressed in the gap section. Professors in music theory or artificial intelligence were omitted from this study as interviewing them would not have addressed the younger demographic gap in my topic regarding the viability of AI-generated music in practice versus simply in theory. Additionally, Generation Z's reception to AI-generated pop music was crucial as their preferences play a significant role in influencing the

current competition and trends within this genre of the music industry, determining the overall viability of the technology.

Additionally, all participation was voluntary and all participants interviewed were currently enrolled in the AP Capstone Program at Independence High School, as selecting from this program fulfilled the Generation Z requirement and offered diverse participants of any gender, race, and socioeconomic background. More specifically, the participants were composed of 12 females and 3 males, with all participants falling between the ages of 15 and 16, living in Frisco, TX, and considered to be from a high socioeconomic background. Information regarding interviews was sent out to all Frisco students in the AP Capstone Program. Due to ethical reasons, all participants in this study volunteered to participate, limiting the generalization of the results as the sample was not fully randomized. Future research could be conducted addressing this limitation.

Sampling Procedure

1. How old are you?
 2. Are you an AP Capstone student with your consent form on file?
 3. Provide a rough estimate of on average, how many hours a day do you spend listening to music?
 4. Do you have any experience with creating music?
 - a. If so, please describe your experience (Ex. making covers to popular songs)
 5. Have you had any past experience with using AI-music generators?
 - a. If so, please describe your experience, describing any challenges you faced throughout the collaboration process.
 6. Have you ever listened to an AI-generated piece of music?
 - a. If so, please state the name of the music piece and describe what you liked and disliked about it
 7. How do you determine if a piece of music is creative?

This next set of questions is asked twice. Once per each music sample. (1 being the lowest 10 being the highest)

 1. On a scale of 1-10 please rate the overall creativity/innovation of this sample of music.
 2. On a scale of 1-10 please rate the emotion and expression of this sample.
 3. On a scale of 1-10, please rate the authenticity of this sample.
 4. On a scale of 1-10 please give this sample an overall rating based on how much you liked it.
 5. Please describe aspects of this sample that you did and did not like. (Ex. rhythm, melody, lyrics, sub-genre/style, etc.)
 6. On a scale of 1-10, how likely are you to listen to this sample of music again? Briefly justify your answer explaining why or why not.
- Overall/Wrapping up Questions:
1. Which sample of music did you prefer (1 or 2) and why?
 2. Based on your responses to the questions above, did one sample stick out to you over the other?
 - a. If so, please describe why or why not.
 3. With the increased saturation of AI-generated music, how has your experience and definition of creativity changed?

FIGURE 1. Guiding Interview Questions.

The first step to collecting data was to create a list of potential guiding questions for the interviews, which can be seen in Figure 1 and targeted the overall goal of the study, while being cognizant of confidentiality and ethical implications. These questions focused specifically on 5 main attributes: emotion, authenticity, creativity, overall rating, and relay value which was defined as a participant's likeliness to listen to a particular sample repeatedly. These attributes were selected based on their prominence in the limited prior literature evaluating outputs of AI-music generators. To find two similar samples of music, one AI-generated and the other human-produced, the first music sample was generated through Udio and with the input prompt: "a indie pop song about love, meaningful, catchy lyrics." From there, a random number generator was used to determine which produced output to use. This was identified as Sample 1. Then, a similar process was repeated for the human-produced music sample. The same prompt was imputed onto artlist.io, a royalty free music website, to reduce bias. Again, a random number generator was used to sort through the results and picked, "California Dreams" by Joseph Brothers, which was referred to as Sample 2.

To account for practicality and participants' safety, emails were chosen as the sole means of communication between the interviewer and participants as they were convenient, professional, and allowed for the opportunity to schedule a coordinated time to conduct a virtual interview that aligned with the participants' schedules. Prior to the interviews themselves, the procedure and list of potential questions in this study were cleared by a Campus Internal Review Board and deemed ethical, allowing continuation onto the following steps. The sampling group consisted of 14 students enrolled in the AP Capstone Program, which were identified after an advisor shared information regarding the interviews with all AP Capstone classes to avoid convenience sampling. The information provided to potential participants contained information regarding the purpose, risks, time-commitment, and qualifications for participating in the interview, which included having filled out the FIRD informed consent form at the beginning of the academic school year provided by their AP Capstone teachers. Through filling out this form, the students acknowledged that their participation in the interviews was voluntary, anonymous, and that they could withdraw from the study at any time.

From there, students reached out via email to schedule an interview. Only once students expressed their interest and confirmed their consent forms were on file, were online interview times virtually scheduled through Zoom. It took approximately one week for participants to express interest and fortunately the sample size received from this initial round of emails was sufficient to progress. Each interview ranged from 10-20 minutes and participants were aware that the interviews were recorded to later be transcribed using Restream AI ("Transcribe Audio," 2024). During the interview, participants were first asked seven general questions

regarding their music consumption and any past experiences with AI-music, then listened to the first 30 seconds of the AI-generated sample, rated various aspects of the composition on a scale from one to ten and briefly asked to justify why, and the same steps were repeated for the human-produced sample. Follow-up questions were also asked based on popular topics that arose and when clarification was needed. Participants were not aware that either sample was AI-generated. This single blind technique was utilized to recreate a real-world environment in which consumers would come across songs naturally. Finally, participants were asked a few overall questions comparing the two samples. All participant information was then stored in a password-protected computer with assigned code names for each participant to maintain confidentiality.

Data Analysis

As the study collected both quantitative and qualitative data, two different analyses were performed to cater towards both data types. For the quantitative data, a t-Test: Paired Two Sample for Means was performed for each of the five attributes (creativity, emotion, authenticity, overall rating and replay value).

Additionally, a thematic analysis was utilized to identify recurring trends and themes. A thematic analysis, according to ATLAS.ti, is “particularly valuable when integrating qualitative and quantitative approaches in mixed methods research” as it offered “a structured yet flexible way of extracting and analyzing themes from a diverse set of data” that a quantitative analysis alone may have lacked (“Thematic Analysis,” 2024). Thus, this method allowed conclusions to be effectively drawn from open-ended responses and supported the inductive reasoning of this paper’s research question to help compare the opinions of Generation Z on the 2 different music samples. Each interview transcript was printed out, then general codes were made through highlighting significant words in each transcript. From there, these codes were grouped into larger themes that were found to be recurring across a majority of the transcripts.

Results and Findings

Attributes	Difference in Means ($\mu_{human} - \mu_{ai}$)
Replay Value	2.785714285
Overall Rating	2.035714285
Emotion	1.428571429
Creativity	0.892857143
Authenticity	0.571428571

FIGURE 2. Mean Difference in Ratings per Each Attribute. Note. Mean Difference = Sample 2 Mean Rating - Sample 1 Mean Rating.

Attributes	P(T<=t) Two-Tail Values
Creativity	0.0376461897
Replay Value	0.0001729500
Emotion	0.0011131537
Authenticity	0.2927570031
Overall Rating	0.0009904013

FIGURE 3. t-Test: Paired Two Sample for Means for All Attributes. Note. All p-values values shown in the figure resulted from a Two-tailed t-Test for probability. While white indicates that the p-value > 0.05, grey indicates that the p-value < 0.05 for each attribute respectively.

Initially, the difference in mean scores was calculated for each attribute through calculating the mean, average rating, for Sample 1 (AI-generated) and Sample 2 (human-produced) individually. Then, the mean of Sample 1 was subtracted from the mean of Sample 2 to create the values displayed in Figure 2, which were organized from the highest attribute variance to the lowest. While the values in the figure do demonstrate higher mean ratings for Sample 2 in every attribute, this difference in sample means alone does not consider the variability in the sampling distribution of the differences in sample means and could be attributed to random chance. Meaning that if another sample were to be taken following the same procedure, it might result in an opposite conclusion. Thus, an inference procedure, such as a Paired Samples t Test, was needed to determine if the observed difference in sample means was statistically significant or if it simply occurred by random chance.

Thus, to adequately understand if the the difference between the popularity of AI-generated music versus the popularity of human-produced music was statistically significant, a Paired Samples t Test was conducted between the ratings for each attribute rated by participants with results displayed in Figure 3.

A Paired Samples t Test, was defined by Kent State University, as a statistical test that compares the means of two measurements taken from related units to determine if there is a significant difference between the groups (“LibGuides: SPSS,” n.d.). Thus, this statistical test was the ideal inference procedure as it suited the comparative nature of the paper’s research question. Additionally, a two-tailed test was chosen as opposed to a one-tailed test to reduce the possibility of falsely declaring a significant effect and test for any significant difference whether that be in the positive or negative direction as the initial null hypothesis was that there was no significant difference between Sample 1 and 2 in any attribute. The Paired Samples t Test was run through inputting raw participant data into Microsoft Excel. Through the test, it was concluded that there was a

significant difference between the mean scores of each respective attribute for AI-generated and human-produced pop music in each of the five categories except authenticity as depicted by the color coding in Figure 3, where the darker gray was correlated with a statistically significant difference in means for that attribute whereas attributes in white were found to not be statistically significant.

More specifically, the two tail p-values for creativity, emotion, authenticity, overall ratings, and replay value were found to be 0.037646, 0.001113, 0.2927570, 0.00099, and 0.000172, respectively. All p-values, with the exception of authenticity, were found to be significantly smaller than the alpha value, which was set to the standard 0.05. The low alpha value ensures that there is a low risk of error with only a 5% chance of a Type I error or a “false positive” occurring. Thus, there is convincing evidence to reject the null hypothesis for these attributes and claim that the alternative is, in fact, true. To clarify, this means that the population mean, is not equal to the hypothesized difference of zero, and thus, there is a significant difference between the ratings of creativity, emotion, overall ratings, and replay value in Sample 1 versus Sample 2. To further solidify these results, since the mean rating for each attribute was higher for all attributes in Sample 2, it can be concluded that there is convincing evidence that creativity, emotion, replay value, and overall rating values were significantly better in Sample 2, the human-produced sample. However, authenticity is concluded to not be statistically significant in the data because its high p-value of 0.2927570 was higher than the alpha of 0.05; thus, there is not convincing evidence to suggest a significant difference between the means and thus it fails to reject the null hypothesis. This means that the difference in means between Sample 1 and 2 was insufficient to deem statistically significant for authenticity, implying that the variation in means for this attribute is likely due to random chance. A potential reason for this result could be due to confusion surrounding the definition of authenticity, which is discussed further in the limitations section.

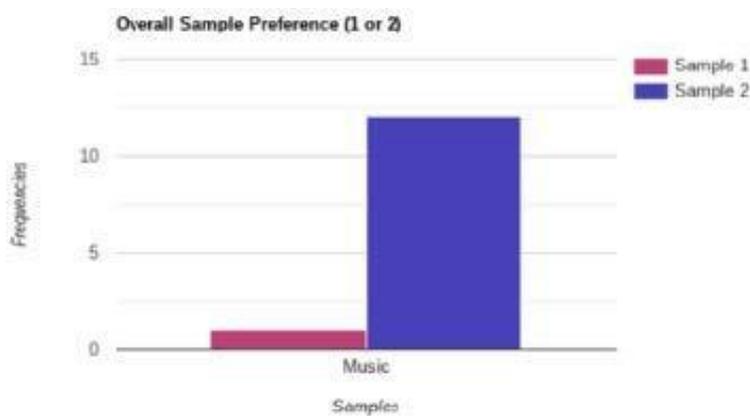


FIGURE 4. Overall Sample Preference. Note. Sample 1 represents the AI-generated Pop Music Sample while Sample 2 represents the Human-produced Pop Music Sample.

Additionally, when asked to consider each sample holistically, nearly all participants, 13 out of 14, preferred Sample 2, the human-produced sample, as shown in Figure 4. Potential reasons as to why will be further detailed through the thematic analysis performed.

Thematic Analysis

TABLE I
INTERVIEW THEMES

Table 1: Definition of Themes		
Sample Number	Recurring Themes Within Specific Sample	Description
Sample 1	Genericness	Throughout interviews, participants made it clear that they had differing opinions on the uniqueness of each sample. Codes for these sections included original sounding, stock photo, auto generated, autotune, copied, generic, monotone, robotic, and repetitiveness often used in response to Sample 1. For example, Participant A described Sample 1 as "not exactly copied, but [it] definitely wasn't like innovative, kinda, just felt like it pulled from very generic audios."
	Childhood References	When describing the replay value of Sample 1 primarily, a common pattern seen among participants was childhood references. Codes included references to elementary school songs, nostalgia, upbeat, preschool, funny, silly, younger audience, and Disney. For example, Participant C stated that Sample 1 was "kind of childish in a way, kind of something like a little kid would probably listen to" and was "silly in a way."
	Dissonance	During interviews, the majority of participants identified an off-putting dissonance within Sample 1. Codes included weird, sounding off, voice cracks, robotic, disconnect, and autotune, chaotic, overwhelming, unbalanced, and harsh. For instance, when discussing Sample 1, Participant E mentioned that "the voice sounded a little bit, like it sounded weird" . . . "like autotuned or something."
	Style alignment/Genre	During interviews, the majority of participants mentioned genre/style issues in some capacity. Codes included expressing dislike due to incompatibility with participants' music genres, country music, and references to redundant similarities with older music. For instance, Participant A mentioned that Sample 1 "was not really like [their] style or genre", stating that "it was really upbeat" and that they did not "like how it [was] so happy". Additionally, other participants such as Participant I and G assumed that Sample 1 "[was] country music" as they felt "it [sounded] the most similar to country music."
	Poor quality	Many participants mentioned the poor lyrical and vocal quality in Sample 1 and periodically in Sample 2. This is further addressed in the limitations section. Codes for this section include, enunciation difficulty and poor translation. For instance, when referencing Sample 1, Participant G mentioned that they could not "hear the lyrics that well" and that "it took a little bit extra effort to like, hear the lyrics . . . it might have been because it was like muffled within the music."
Sample 2	Variability	When describing positive aspects of Sample 2, many participants contrasted it to the genericness observed in Sample 1. Responses to Sample 2 included codes such as variety, anticipation, and beat drops. For instance, when examining Sample 2, Participant K mentioned that they enjoyed that everything "flowed really well together" and that "every single time he stops singing, the [whole chorus] would stop . . . and they would pick it back up again."

	Lyrics	Many participants correlated strong lyrics with a higher emotion score. Codes for this section included personalized lyrics, human experiences, relatable, and story-telling. For instance, when discussing Sample 2, Participant G mentioned they “preferred number 2 because the lyrics were telling a story.”
	Design	Throughout interviews, participants mentioned the overall balance of the piece as a factor when considering an overall rating. Codes included in-depth, thoughtful, intentional, thoughtful, blended, thoughtful composition, and audio balance. For instance, Participant N stated that they definitely prefer Sample 2 as “it definitely resembl[ed] more of, like, the modern pop music scene” and was “definitely something that, like, [they] could see [themsell] listening to normally just because . . . the melody was nice . . .” and conveyed an appeal.
Real-World Experiences	Previous Exposure to AI-generated Music	Five participants shared their previous experiences with listening to AI-generated music. Codes included covers, references to pre-existing music, social media, and short-form content in particular. However, no participants shared prior experience with using AI-music generators. For instance, Participant C mentioned they listened to an AI-generated song replicating a certain Kpop artist while scrolling through Instagram. Participant C mentioned that the song “did sound very much like him, but at the same time it sounded kind of off.” “It was as if you could tell immediately that he was not singing it.”
	Competition	Many follow-up questions regarding direct competition between AI-generated music and human-produced music arose during interviews. Codes included uncertainty, job displacement, neutrality, cheating, integrity, and AI limitations. For instance, Participant M stated that “AI does not really pay attention to . . . make sure, like, you can hear both sides, like, the background music and the vocals, which [the participant] felt like an artist would focus more on.”

TABLE 1. Recurring Interview Themes.

Discussion of Major Recurring Themes

As seen in Table 1, there were multiple recurring themes for participants. These subthemes were grouped under 3 main groups: Sample 1, Sample 2, and Real World Experiences. Overall, Sample 1 was regarded more negatively in comparison to Sample 2.

For Sample 1, prevalent themes were genericness, childhood references, dissonance, poor style alignment, and poor quality. Poor style alignment could be attributed to the few genre specializations mentioned due to limited training data in certain genres of music, potentially including pop music as described by Dash and Agres (Dash & Agres, 2024). Additionally, dissonance as a recurring theme is also bolstered by Avdeeff’s study in which she attributed a similar phenomenon, uncanny valley, as reasoning for participant hesitancy to AI-generated pop music.

For Sample 2, opinions were generally more positive with themes such as high variability, relatable lyrics, and complex design. The inability of Sample 1 to have a complex design is supported by the limited user input and choices within the DeepBach model that “do not involve expert knowledge about harmony or scales but are only mere observations of the corpus,” (Hadjeres et al., 2017) potentially providing similar reasoning for the simple compositions in Sample 1.

Lastly, 5 out of the 14 participants mentioned having previous exposure to AI-generated music in the past, mainly through social media such as Instagram, suggesting the increased saturation of AI-generated music not only within but also outside of the music industry. Furthermore, when asked about competition, many participants expressed negative

views regarding the ethics and integrity of AI-generated pop music while others maintained a neutral position, mentioning they were not opposed to AI-generated pop music as long as the effects on pre-established artists, such as job displacement were limited.

Limitations

Before drawing conclusions, it is necessary to address potential limitations of this study, as they could hinder results. Initially, Udio, the AI-music generator that was used to produce Sample 1 was the best quality generator that the researcher had access to; therefore, the quality aspect of Sample 1 could be attributed to the limitations of this particular model. Additionally, a potential reason for authenticity being the only attribute deemed not significantly different could be due to the common confusion surrounding the definition of authenticity as many participants showed signs of confusion and only a select few asked for clarification before answering, potentially skewing their ratings for that particular section.

Another limitation was that both music samples were played over Zoom, as all interviews were conducted virtually. Therefore, the quality aspect of both samples may and thus the ratings of participants may have been skewed by the Zoom quality of each sample as mentioned by a few participants.

Additionally, the single blind procedure might not have been effective due to the semi-structured format of interviews, as the researcher and participant may have inadvertently introduced biases throughout the duration of the interview. This may have potentially skewed participant responses. Thus, for future research, the methodology of this study could be slightly refined to further avoid potential biases, potentially through adopting a double blind procedure.

Finally, the sample size of the study was relatively small, with 14 total participants; therefore, to generalize the results to the population of Generation Z consumers of pop music, a larger sample size of Generation Z would be optimal with more diverse demographics.

Conclusions

As identified in the quantitative analysis section of this study, the alternative hypothesis was true as there was a significant difference in the popularity of AI-generated and human-produced pop music with the latter being preferred holistically. As for individual attributes, the Paired Samples t Test further exacerbated the significance of popularity of human-produced music over that of AI-generated music in 4 of the 5 attributes tested within this study: creativity, emotion, overall rating, and replay value. This disparity can be attributed to dissonance, genericness, poor style alignment, and poor quality within the AI-generated pop music sample found through the thematic analysis. This supports the findings of

Avdeeff, who also found dissonance as a common reaction to AI-generated pop music which she described as uncanny valley and attributed much of the audience's hesitancy to. These conclusions greatly differ from the initial null hypothesis, which suggested that there was not a significant difference between the AI-generated and the human-produced samples in any category. However, authenticity was the only attribute deemed not statistically significant, potentially due to definition confusion among participants.

Thus, further research should be conducted on this attribute in particular. Overall, these conclusions reveal 4 possible target areas for improvement in current AI-music generators utilizing GANs for further appeal to younger consumers of pop music, specifically Generation Z. Additionally, while AI-music generators have evolved significantly over the past few years, models involving GANs still have limitations that are evident in their outputs.

Furthermore, the significantly lower popularity of AI-generated music does suggest that, as of now, musical outputs generated by GANs are not yet on the level to provide significant competition to human-produced pop music as conveyed by the strong opinions regarding the continued dissonance and easily differentiable aspects of current AI-generated music such as vocals and lyrics. Moreover, further research should be conducted on the reception of AI-generated musical outputs within other types of music genres, such as folk and jazz music, as this study focused explicitly on the pop genre. Additionally, comparing results across different music genres could help generalize conclusions regarding reception to AI-generated music across the music industry as a whole. Such findings could also guide the implementation of regulations to promote the balance of innovation while protecting works of pre-established musicians for if and when AI-music generators close the gap in the coming years.

References

- Anantrasirichai, N., & Bull, D. (2021). Artificial intelligence in the creative industries: a review. *Artificial Intelligence Review*, 55(1), 589–656. <https://doi.org/10.1007/s10462-021-10039-7>.
- Avdeeff, M. (2019). Artificial Intelligence & Popular Music: SKYGGGE, Flow Machines, and the Audio Uncanny Valley. *Arts*, 8(4), 130. <https://doi.org/10.3390/arts8040130>.
- Briot, J., & Pachet, F. (2018). Deep learning for music generation: challenges and directions. *Neural Computing and Applications*, 32(4), 981–993. <https://doi.org/10.1007/s00521-018-3813-6>.
- Bryan-Kinns, N., Banar, B., Ford, C., Reed, C. N., Zhang, Y., Colton, S., & Armitage, J. (2023). Exploring XAI for the Arts: Explaining Latent Space in Generative Music. *Arxiv*, 1-14. <https://arxiv.org/abs/2308.05496v1>.
- Bukhari, S. W. R., Hassan, S. U., & Aleem, Y. (2023). Impact of Artificial

- intelligence on Copyright Law: Challenges and Prospects. *Journal of Law & Social Studies*, 5(4), 647–656.
<https://doi.org/10.52279/jlss.05.04.647656>.
- Carnovalini, F., & Rodà, A. (2020). Computational Creativity and Music Generation Systems: An Introduction to the state of the art. *Frontiers in Artificial Intelligence*, 3, 1-20.
<https://doi.org/10.3389/frai.2020.00014>.
- Chen, Y., Huang, L., & Gou, T. (2024). Applications and Advances of Artificial Intelligence in Music Generation: A Review. *Cornell University*, 1-19. <https://arxiv.org/abs/2409.03715>.
- Dash, A., & Agres, K. (2024). AI-Based Affective Music Generation Systems: A Review of Methods and Challenges. *ACM Computing Surveys*, 56(11), 1–34. <https://doi.org/10.1145/3672554>.
- Figueira, A., & Vaz, B. (2022). Survey on synthetic data generation, evaluation methods and GANs. *Mathematics*, 10(15), 2733.
<https://doi.org/10.3390/math10152733>.
- Hadjeres, G., Pachet, F., & Nielsen, F. (2017). DeepBach: a Steerable Model for Bach Chorales Generation. *PMLR*, 1-10.
<https://proceedings.mlr.press/v70/hadjeres17a.html?ref=https://githubhelp.com>.
- Ijiga, N. O. M., Idoko, N. I. P., Enyejo, N. L. A., Akoh, N. O., Ugbane, N. S. I., & Ibokette, A. (2024). Harmonizing the voices of AI: Exploring generative music models, voice cloning, and voice transfer for creative expression. *World Journal of Advanced Engineering Technology and Sciences*, 11(1), 372–394.
<https://doi.org/10.30574/wjaets.2024.11.1.0072>.
- Jourdan, T., & Caramiaux, B. (2023). Machine Learning for Musical Expression: A Systematic Literature Review. *New Interfaces for Musical Expression (NIME)*, 1-13.
<https://hal.science/hal-04075492/>
- LibGuides: SPSS Tutorials: Paired Samples t Test*. (n.d.). Kent State University.
<https://libguides.library.kent.edu/SPSS/pairedsamplesttest>.
- Pugachev, A. A., Kharchenko, A. V., & Sleptsov, N. A. (2023). Transforming the future: a review of artificial intelligence models. *RUDN Journal of Studies in Literature and Journalism*, 28(2), 355–367. <https://doi.org/10.22363/2312-9220-2023-28-2-355-367>.
- Semi-Structured Interview: Explanation, Examples, & How-To*. (2023). Dovetail. <https://dovetail.com/research/semi-structured-interview/>.
- Thematic Analysis in Mixed Methods Approach*. (2025). ATLAS.ti. <https://atlasti.com/guides/thematic-analysis/thematic-analysis-mixed-methods#:~:text=Thematic%20analysis%20is%20a%20widely%20diverse%20set%20of%20data>.
- Transcribe Audio to Text - AI Transcription Software | Restream*. (2024). Restream AI. <https://restream.io/tools/transcribe-audio-to-text>.