Can Chat-GPT Rap: Analyzing the ability of Chat-GPT to replicate and Analyze the Lyrical Style of Young Thug.

Eli France, Jon Chun, and Katherine Elkins

IPHS 300 AI for Humanity (Spring 2023) Prof Elkins and Chun, Kenyon College

Abstract

Initially, it was believed that while automation would quickly replace manual labor, it would be more difficult for artificial intelligence to replicate human creativity and produce art. However, in the last five years, this assumption has been proven quite wrong. AI art, music, and literature are all burgeoning genres in their own right. While there are questions regarding the creative merit of their work, the ascendancy of these programs seems to indicate that AI can produce art that replicates that created by humans. Nevertheless, artists have been typically reluctant to embrace these innovations in artificial intelligence out of fear that they may be replaced by these programs.

While AI art is rapidly gaining popularity, one medium which has yet to be infiltrated by AI is songwriting. As of the time of this work, there are no AI programs designed to engage in songwriting. However, in my research, I show that through the proper usage of one-shot learning and prompt engineering, LLMs possess the ability to replicate the songwriting style of rapper Young Thug accurately. These results suggest that AI may not replace artists but rather make their jobs easier.

Introduction

While musical creation was once considered one of the pinnacles of human creativity, AI has made significant advancements in the field. While the first known AI music experiment was conducted by Alan Turing in 1951, it would not be until when Google Magenta and Watson Beat could generate musical compositions. While these initial compositions were elementary and often one minute, many new tools for AI music generation, such as AIVA, now allow users to generate entire songs. A recent study found that the AI-music generation could create compositions indistinguishable from those of humans. The paper further predicted that by 2030, AI music would dominate the industry. While it remains to be seen if their prediction is validated, an AI-generated song entitled "Heart on my sleeve," created by anonymous tik-toker @ghostt977, which utilized the voices of popular Canadian musicians, The Weeknd and Drake, was streamed millions of times in just a week before being removed from major platforms due to copyright issues on the request of Universal Music Group. However, while this song is no longer accessible, it sparked larger debates about the validity and legality of AI music, especially as it pertains to the creation of songs in the style of a specific performer.

Opponents of AI-generated music claim it can potentially destroy the music industry. AI-Music, especially when used to replicate pre-existing sounds, can potentially devalue human artists. Considering the capability of artificial intelligence (AI) to generate music at a significantly faster rate and lower cost compared to humans, it is reasonable to suggest that the music industry, upon adopting AI-generated songs, could result in a decreased demand for human musicians. The prospect of such an outcome raises concerns about the potential obsolescence of human creativity and the implications of relying on AI-generated content. Moreover, given that AI-song generations rely upon training models on pre-existing music, which is thus replicated, AI music's use of intellectual property renders its legality questionable. Additionally, given that AI music only serves to synthesize pre-existing patterns, some critics doubt its capabilities regarding innovation and creativity. Finally, there are questions about whether AI can truly create music that encapsulates the nuances and subtleties of human emotion and expression, as expressed in a 2022 research paper that claimed AI struggles most to make music humans consider to be emotionally resonant. Thus, they suggest that humans must be involved or provide feedback.

Defenders of AI music see it not as a replacement for traditional artistry but rather as a new tool artists can use to enhance their work. Proponents of this technology argue that AI music can be created more efficiently and cost-effectively than traditional methods of music creation. In this way, AI music can provide an opportunity to express musical creativity for people who may, for various reasons, such as lack of formal training or limited monetary access, have been limited from doing so. Moreover, AI music can also allow for new possibilities, such as collaborations that may not have otherwise been possible or the creation and release of new posthumous music. Finally, as Carlos Hernandez-Olivan argued in his 2022 paper, A Survey on Artificial Intelligence for Music Generation: Agents, Domains, and Perspectives, individual creativity is a finite resource, and thus, AI music generation may allow for the creation of innovative new sounds and compositions. Finally, AI lyrics can assist in ending snippet culture. Snippet culture, a term defined by popular music publication Pitchfork, refers to when artists preview unfinished versions of their songs yet never release them as they struggle to finish the song due to writer's block. Thus, Chat-GPT and other LLMS can be used to help artists complete songs.

Currently, AI-music creation largely revolves around the generation of original compositions. Services such as AIVA, Amper, and OpenAI Muse-Net allow users to compose royalty-free AI-generated instrumental tracks. These programs allow the user to create technically-proficient compositions in any given style, and some even allow the user to provide "inspiration" in the form of a pre-existing song. However, songs often contain not only music but also lyrics. While significant AI innovations have occurred regarding the former, far less attention has been paid to the latter. As of the time I write this, there are no songwriting AI services, and the existing music generation programs do not include lyrics in their compositions. Many of the most popular pieces of AI music, such as the aforementioned "Heart on My Sleeve" track, utilized AI vocal and music composition but human-written lyrics. Nevertheless, Large Language Models (LLM) provide an exciting opportunity for the generation of lyrics. LLMs, including OpenAI's Chat-GPT, generate responses to prompts or questions by utilizing natural language processing and deep learning algorithms. These models undergo a training process utilizing a vast corpus of textual data, enabling them to learn common patterns and associations within language. The training data is tokenized, which means the text is divided into units of meaning, such as individual words or phrases, which the AI can comprehend. When a user inputs a prompt, the LMM employs its knowledge of tokenized language to predict the most probable word or phrase to follow. In this way, it generates textual passages through a series of predictions. This process allows the AI to simulate natural communication and produce outputs often indistinguishable from human authors.

Consequently, LLMs have shown the ability to create lyrics that conform to standard songwriting practices, including continuous rhyme schemes and verse-chorus format. Moreover, given the vast textual data on which these programs were trained, they can often create lyrics in the style of certain artists. This feature is the one that interests me most. The art of songwriting has always held a deep fascination for me. Although the composition of music(which here refers to the actual instruments being played or simulated) is undoubtedly a large aspect of songs, it is often the lyrics of my favorite artists that truly captivate me. While genre conventions may restrict instrumental pieces, the lyrical aspect of a song provides abundant room for creativity and, in this way, allows artists to stand out truly. Songwriters can express their individuality through a unique vocabulary, thematic focus, flow, and rhyme scheme and cultivate an innovative style that sets them apart from their peers. In this way, the lyrics of a song serve as a medium for conveying an artist's personality and artistic vision. Therefore, I want to discern the extent to which LLMS can successfully replicate the songwriting styles of a specific. I specifically chose rappers, as they are far more likely than other artists to write their lyrics rather than relying on a team of songwriters. Unlike traditional singing, rap music is characterized by its rhythmic vocal delivery, which provides a distinct and diverse approach to songwriting. The use of rhythmic patterns in rap allows for greater creativity and experimentation in the songwriting process. Finally, rappers often use their lyrics to express their verbal prowess rather than tell a defined story. Consequently, as songs do not have to differ as much due to a lack of a theme, more recognizable patterns can be discerned in their songwriting.

I believe that the results of the project have valuable implications concerning the future of AI-Music. The extent to which LLMs can accurately generate lyrics similar to those of existing rappers will be quite revealing regarding the ability of AI music to replace human-made music. Moreover, comparing the generated lyrics will allow insight into the legality of AI-lyric generation. Specifically, it can be seen if the songs created plagiarize lyrics or produce content that, while abiding by the standards of a certain artist, has enough distinct material to be classified as fully unique. I am aware of certain limitations to this project. First, LLMs must abide by certain content policies that prevent them from generating content that is considered too controversial. Consequently, these lyrics generated may lack the use of profanity and references to illicit activity common in rap lyrics. Due to content policies that limit potentially offensive or controversial language, AI-generated lyrics may prioritize conforming to mainstream standards of acceptability and therefore lack the distinctive creativity and originality that characterizes the work of the human artists that served as their inspiration. Moreover, given that LLMs work by predicting on a word-by-word or phrase-by-phrase basis, they lack long-term foresight and thus may struggle to create intricate rhyme schemes or coherent narratives. Nevertheless, I believe many of these limitations can be avoided through advanced prompt engineering, and thus it is worthwhile to investigate the songwriting capabilities of LLMS. Therefore, my project will answer the question: To what Extent Can LLMs replicate the songwriting styles of popular Rappers?

QR Codes For Resource

Lyrics

Methodology

The first step in my research would be to select a rapper whose style was to be replicated. After researching various artists, I ultimately decided that Young Thug, real name Jeffrey Williams (henceforth referred to as "Thug" or "Williams"), would be the focus of my study. Williams was selected due to his unique and highly recognizable songwriting style, which Pitchfork has described his writing as "extraordinarily distinctive." Moreover, Williams' has been categorized as the most influential rapper of the 21st century by the BBC, who noted that many young rappers attribute their style to Thugs' distinct use of slang, wordplay, and adlibs. My criteria for selecting the song whose sentiments would replicate were that it had to include hallmarks of William's style, a clear narrative and had to have been released after the knowledge cutoff date for Chat-GPT. Accordingly, from his most recent 2021 album "Punk," the song "Hate The Game" was chosen due to its hallmark use of slang, clear narrative, and triumphant message, which aligns with the themes which I determined to be most present in Young Thug's music. It is important to note that the song was released after the knowledge cutoff date for Chat-GPT, meaning that the AI would not have any knowledge of the lyrics.

Eight different prompts were used to simulate different possible ways in which AI could be implemented into songwriting practice. The first prompt utilized zero-shot learning, which allowed the AI to create lyrics with as little human involvement as possible. This prompt simply asked CHAT-GPT to create a song entitled "Hate The Game" in the style of Young Thug. The lyrics created by this prompt are classified as "Standard," and from here on out, this will be referred to as the "standard" prompt. Subsequently, I attempted to use One-shot learning, providing an example of real Young Thug Lyrics for which the LLM was to reference for vocabulary and phrasing. Next, I explored the capabilities of detailed prompt engineering, providing details of the mood, themes, and topics with which to reference in addition to the standard prompt. After this, I engaged in Human-AI collaboration, providing the LLM with lyrics to the chorus of "Hate The Game" and asking it to write two additional verses.

The attributes of these prompts were not mutually exclusive. Consequently, I combined them to form 4 different combination prompts. The first was a prompt with both an explanation and an example, the second was a prompt with an explanation and chorus, the third was a prompt with a chorus and example, and the fourth included all three.

I judged the generated songs not on their ability to replicate one aspect of Young Thug's songwriting but rather on encapsulating the various features which make his style unique. The process of manually assessing the similarities of the outputs to the real song began with me analyzing the actual lyrics to "Hate The Game." I identified several of the most apparent sentiments, themes, and topics(shown in Chart 1) and next assessed and recorded the frequency at which these themes or sentiments were apparent in the lines of the songs. This process was repeated for each of the eight songs. To calculate the average disparity percentage (the metric I created to assess similarity) for each category, I took the absolute value of the difference between the frequencies with which each theme or sentiment appeared in the real and generated songs. For each prompt attribute type, I added the absolute value scores together and divided them by the number of categories such that I could determine the average disparity concerning the frequency with which topics and themes appear between the real and generated songs(henceforth known as the Average Frequency Disparity Percentage or AFDP. I then repeated this process with a smaller set of categories so I could focus specifically on sentiment and topical disparity percentage. (these will be referred to as SAFDP and TAFDP, respectively). Finally, I identified key attributes of Thug's music, such as multi-syllabic rhyme scheme, adlibs, and his distinct phrases, and subtracted 25 from the scores of each of the generated lyrics for how many times they appeared while adding .25 for every lyric which I deemed unnatural. Finally, I recorded and charted each of the three scores for each similarity type.

To get a second opinion on the similarities and to gain further insight into the process by which LLMs understand lyrics, I also had CHAT-GPT assess the generated lyrics. To achieve this goal, I asked CHAT-GPT for the most common sentiments, themes, and topics of Young Thug Songs(displayed in Chart 2). Subsequently, I asked GPT to rate the lyrics of "Hate The Game" by Young Thug on the basis of how present each sentiment, theme, and topic was. I repeated this process with the lyrics generated by all eight prompts and once again took the average difference between their scores and the score of the original song. I next analyzed and compared the results to see which prompts produced the most similar text and how my evaluations differed from those of the AI.

Chart 1:

Graph 1:
Manually Chosen Themes or Topics
Triumphant
Sad/Neutral
Braggadocious Statements
Relationship Toxicity
Violence
References to Brand Names
Weapons
Transportation
Cops
Drugs
Clothing, Jewelry and Accessories
References to Women or Sex
Lines about Thug

Chart 2:

AI suggested theme or topic	Explanation
Money and success	Young Thug frequently raps about his wealth, luxurious lifestyle, and success in the music industry. He often mentions expensive cars, jewelry, and designer clothes, and emphasizes his status as a wealthy celebrity.
Drugs and partying	Drugs and partying: Another recurring theme in Young Thug's music is drug use. He often talks about using drugs like marijuana, lean, and Xanax, and portrays himself as a hedonistic lifestyle that involves partying and having fun.
Relationships and love	Young Thug also frequently raps about relationships and love, often focusing on romantic entanglements and the challenges that they bring. He expresses both heartbreak and infatuation.
Struggle and perseverance	Despite his success, Young Thug has also faced significant struggles in his life, including poverty, violence, and legal issues. He often raps about his resilience and determination to overcome these challenges and the importance of working hard to achieve one's goals.
Violence and street life	Some of Young Thug's music also deals with themes of violence and street life. He sometimes talks about his experiences growing up in a rough neighborhood and the dangers of gang culture and violence.
Confidence and swagger	Young Thug often exudes confidence and swagger in his music, projecting an image of himself as a cool, confident, and unapologetic artist.
Joy and happiness	Many of Young Thug's songs have an upbeat, joyful vibe that encourages listeners to dance and have fun.
Melancholy and sadness	Some of Young Thug's music can be introspective and melancholic, exploring themes of heartbreak, loneliness, and the

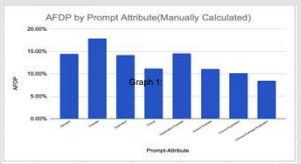
Results and Analysis Part 1

The AI showed a remarkable propensity to create songs quite similar to those of Young Thug. Each piece GPT created incorporated a standard rhyme scheme throughout the verses as well as elements of wordplay and adlibs. From my own personal analysis, a vast majority of lyrics seemed to align with those of Thugs, as only the standard prompt had more than three lines which I marked as being unnatural. Most of these unnatural lines used dated vernacular, suggesting that Chat-GPT may be limited in its ability to understand the nuances of language usage fully.

Nevertheless, the AFDP (the metric defined in the Methodology section) remained below 20% among all eight different prompts utilized. As illustrated in Graph 1, the standard prompt, which aimed to provide the LLM with minimal information, achieved an ADFP below 15. Surprisingly, the LLM seemed to struggle with one-shot learning. Prompts that included examples resulted in an AFDP of 17.9, the highest among all prompts attribute types. It seems this is the result of overfitting, as I noticed when analyzing the output that it repeated a few of the lyrics in the prompt nearly verbatim. Explaining to the LLM what to include proved much more successful in replicating the frequency of different topics and sentiments appearing in Young Thug's song. As seen in Chart 1, the prompts with explanations attached achieved an AFDP of 14.3, showing improvement over the standard prompts. However, the lowest ADFP of any of these one-attribute prompts, indicating the highest similarity, belonged to the prompt, which included the song's chorus and simply asked the LLM only to write verses. As seen in Graph 1, the AFDP of these prompts was 11.2%. The fact that the chorus performs the best suggests that the usage of human-AI collaborations might be an effective means of songwriting, allowing artists to complete unfinished ideas.

The results from multi-attribute prompts further support the idea that LLMs work most effectively when asked to complete songs rather than create them from scratch. Prompts that included both an explanation and examples produced an average disparity percentage of 14.6, a score that was significantly higher than prompts with only the chorus. Moreover, as shown in Graph 1, while prompts with both explanations and chorus, as well as examples and chorus, scored lower than those with just the chorus or just examples, the prompt with only the chorus outperformed the prompt with both an example and explanation by a notable margin 1.1. Notably, the prompt that included both explanations and examples, along with the chorus, achieved the lowest score of all prompts, with 8.5. This score means the average difference in the frequency with which topics were discussed in the generated and real song was under 10% or approximately two lines of difference. These results are quite encouraging as they suggest that AI can write a song in a nearly identical manner to an existing artist, with only a slight variation in topics and mood.

Graph 1:



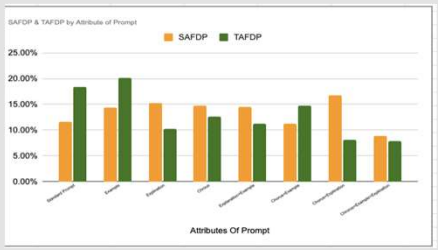
Results and Analysis Part 2

Graph 2 illustrates that while the standard prompt could generate outputs that captured the song's sentiments, it struggled to replicate the topics without specific details. Thus, it achieved a SAFDP score of 11.6 but a TAFDP score of 18.4. The same was true for the prompt utilizing one-shot learning through a Young Thug verse example. This suggests that while one-shot learning can generate output that "feels" like a Young Thug song (with a SAFDP of 14.4%), it faces difficulties in addressing the song's themes based on limited examples, as seen by the TAFDP of 20.2%. However, the inclusion of an explanation appeared to help the LLM more accurately replicate the frequency of topics discussed in Young Thug's lyrics. Graph 2 demonstrates that this prompt achieved one of the lowest TAFDP (10.2%), likely due to the suggestion of certain themes. While it faced challenges in replicating the mood of the song, with a SAFDP of 15.2%, these findings highlight the value of advanced prompt engineering in songwriting. However, providing the chorus resulted in relatively low scores for both SAFDP and TAFDP, at 14.8% and 12.55%, respectively. While the chorus had achieved the lowest AFDP, it neither achieved the highest SAFDP nor TAFDP. This indicates that while prompts like this that specifically engage in human-AI collaboration can replicate lyrics to the highest degree of accuracy of any single-attribute prompt, they are as strong at identifying sentiment as one-shot learning prompts or topics as prompts that use advanced engineering.

The multi-attribute prompts seemed to balance SAFDP and TAFDP with more ease than the single-attribute prompts. The strength of multi-attribute prompts largely lies in their ability to complement each attribute's strengths. As shown in Graph 2, the prompt with both explanation and example achieved one of the lowest TAFDPs (11.22%) and a relatively low SAFDP (14.55%). Consequently, this prompt combined the advantages of prompt engineering and one-shot learning. The same pattern emerged in prompts with a chorus and explanation, which successfully replicated topic frequency (8.1%) but faced challenges with replicating sentiment (16.8%). These findings suggest that advanced prompt engineering is valuable for replicating thematic usage styles in songwriting. Consequently, this indicates that without the use of one-shot learning, CHAT-GPT has difficulty discerning emotions. This suggests that LLMS may better be able to understand the emotive qualities of songs through examples than specific directions.

Alternatively, the scores for the prompt with the chorus and examples indicate that one-shot learning is most successful in capturing the mood of artists' songs. However, the prompt, which included an example, an explanation, and a chorus, had both the lowest average topic and sentiment disparity percentages(7.8 and 8.8, respectively. This suggests that AI songwriting can most accurately replicate the lyrical styles of artists through a combination of one-shot learning, prompt engineering, and Human-AI collaboration.

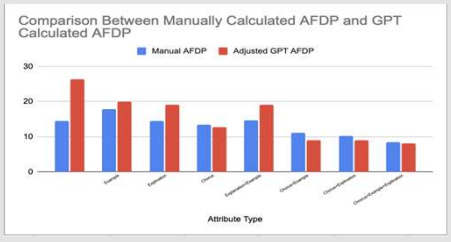
Graph 2:



As can be seen from Graph 3, for the most part, CHAT-GPT assessed the different prompts similarly to me. One notable variation was our assessment of the Standard Prompt, which Chat-GPT found to be far more dissimilar from the actual song than I. I believe this results from this prompt creating more vague lyrics due to lack of information, which Chat GPT could have more difficulty analyzing for themes or sentiments. CHAT-GPT also gave the same rankings to the lyrics produced by prompts with explanations and examples. While I had determined the version with examples to be more notably similar to the original song, CHAT-GPT apparently considered them to be equally similar. The version I believed to be more similar was penalized by Chat-Gpt for not being as violent or sensual as the original song. However, in my own analysis, I found many lines that could be classified as violent or sensual, even if they did not explicitly reference violence or sex. This further points to the fact that Chat-Gpt has difficulty analyzing subtlety in lyrics. Nevertheless, like me, Chat-GPT gave the lowest disparity score of any single-attribute prompt to the prompt with the chorus, indicating the highest similarity.

With the exception of the example and explanation version, which I also considered to be the worst of the advanced prompts, Chat GPT gave all multi-attribute prompts similar scores within .1 of each other. As I have noted, it seems that prompts with explanations allow for better replication of topics. Prompts with examples produce lyrics that better encapsulate the emotions of the song. Given that Chat-GPT gave both the explanation and chorus, as well as the chorus and example prompts, the same score, it seems that Chat-GPT values thematic similarity equally to topical similarity. The fact that the explanation and example-only prompts also got the same score provides further evidence for this theory. Personally, as I identified more topics than moods, topical disparity affected my rating system more. This difference might explain the disparity between my calculations and those of Chat-GPT. However, both Chat-GPT and I graded the prompt, combining all attributes the highest. Consequently, this reveals that CHAT-GPT assesses the attributes of lyrics similarly to humans. Nevertheless, just as Chat-GPT could not fully replicate the styles of actual artists, there were also irregularities in its ability to critique music. Not only did Chat-GPT miss subtleties, but it would occasionally make up lyrics in its explanations for its scores. Consequently, this casts doubt on the reliability of CHAT-GPT to assess lyrics, suggesting that it made need the guidance of a human.

Graph 3:



Conclusion

These results suggest that LLMs can understand and replicate the lyrical styles of rappers such as Young Thug. Even the standard prompts created a song that expressed sentiments and emphasized topics similar to those in "Hate the Game" by Young Thug. Moreover, the experiment shows how different types of prompts help replicate different attributes of an artist's lyrical style. For example, advanced prompt engineering results in a high topical similarity, while one-shot learning helps the LLM to replicate the sentiments expressed by an artist. Consequently, it can be seen how combing these different attributes into one prompt produces the best results. Additionally, the fact that the chorus-only prompt outperformed all the other single-attribute prompts suggests that the best results will emerge from human-AI collaboration. On a broader scale, this suggests that LLMs will merely be another tool in their kit. Rather than replace songwriters, I foresee this helping artists with writer's block to complete song ideas and consequently address the issues of snippet culture.

From comparing my assessments of the songs with those of CHAT-GPT, it appears that LLMs still do not understand lyrics like a human. While I cannot say that my scores are objectively more accurate, I will note that the divergence between my scores and the scores of the AI seems to have several implications. First, my scores seemed to vary more, suggesting that CHAT-GPT is less sensitive to small changes in the lyrics or presents a less numerically-based analysis. Moreover, it seems that CHAT-GPT has trouble detecting more subtle themes. This was because it often tended to rate the songs very high or low on their inclusion of a certain category. Therefore, it struggled to identify song themes or decode the topics discussed in some less prominent lines. Nevertheless, the fact that the LLM agreed with me regarding which song was the most similar to "Hate The Game" shows that, on some level, it can understand lyrics similarly to a human. Thus, while the ability of CHAT-GPT to understand lyrical themes may lag behind its ability to generate music, it is still capable of decoding these lyrics to some extent.

Limitations, Future Research and Considerations

The findings of this study are subject to limitations due to time constraints and available technology during the project's completion as well as the sample size. Had I had more time to conduct this experiment, I would have been able to experiment more with the creation of different types of prompts. This would have allowed for further insights into the best practices for AI songwriting. Moreover, experimenting with more types of prompts would have allowed me to better assess how CHAT-GPT analyzes and creates lyrics. Additionally, had I had more time to complete this experiment, I could have created multiple sets of lyrics to better understand how each prompt's different attributes generated different lyrics. Furthermore, while running this experiment, I only had access to CHAT-GPT 3.5 rather than the most updated version, CHAT-GPT 4. Therefore, my results and their implications on the ability of LLMS to recreate lyrical styles may be slightly outdated. Moreover, I only had CHAT-GPT recreate the lyrical style of one specific artist. Therefore, these results may not be reflected when analyzing the ability of LLMs to recreate other artists' styles or lyrics in genres besides Rap. A final limitation is that my manual assessment of themes and emotions is purely based on my interpretation of the songs. While I tried to research any existing explanation for the lyrics of the song "Hate the Game," when these were unavailable, I relied upon my subjective analysis of what topics I believed were being discussed. This limitation also affected my evaluation of the AI-generated songs.

Ultimately, while my results show promising signs regarding the ability of AI to produce song lyrics in the style of Young Thug, I believe that further research should be conducted concerning the capability of CHAT-GPT to generate lyrics. Specifically, I believe further research should assess the ability of CHAT-GPT to replicate the styles of various artists. Likely I believe that it will have an easier time creating lyrics similar to those of some artists than others. By analyzing which styles this LLM can best copy, I believe valuable insight can be gained regarding how AI understands different songwriting elements. I also believe future studies should look at how CHAT-GPT comprehends and replicates the songwriting styles of artists in different genres. By seeing which genres it can replicate with the highest accuracy, I believe that much can be learned in terms of how LLMs process and understand songs of different genres and differences in the songwriting styles of the genres themselves. This exploration can provide valuable insights into the linguistic, thematic, and cultural characteristics that define different musical genres and contribute to creating distinct lyrical content within each genre.

While I used an example of one-shot learning, I believe that future studies should attempt to incorporate many-shot learning. In doing so, these studies will provide a better understanding of how Chat-Gpt learns about lyrics from examples. Moreover, I believe future studies should delve deeper into prompt engineering. Due to time and technology constraints, I only provided one prompt, which included in-depth song creation details. By experimenting with different advanced prompts, I believe more could be learned about the effective use of prompt engineering to create lyrics. Specifically, I think it would be interesting to see how different levels of specificity resulted in the generation of distinct lyrical patterns. By designing prompts that call for different degrees of attention to detail, it can be examined whether highly specific prompts lead to a more accurate replication of an artist's lyrical patterns. Alternatively, research on this field can also identify if more open-ended prompts allow for greater creativity and innovation. This exploration has immense value because it would help reveal the optimal balance between providing explicit guidance and allowing room for the LLM to explore its own interpretations.

Finally, I believe that future studies should address innovations in technology. Most importantly, I believe that there should be research on the ability of CHAT-GPT 4 to produce lyrics. Not only will the results of this data better represent the ability of current generation AI to write lyrics, but also, by comparing them to the results of my study, much can be gleaned about the differences between GPT 3.5 and GPT 4. Moreover, as AI technology advances, I believe research should be conducted on the ability of an LLM to create a unique and advanced lyrical style. These results will have massive implications regarding the ability of LLMs to produce works of genuine creativity and innovation.

Acknowledgements

I would like to extend my heartfelt gratitude to Professor Elkins and Professor Chun for their unwavering support and guidance throughout my academic journey. Their dedication to empowering students and fostering a passion for learning has equipped me with invaluable skills that enable me to thrive in today's rapidly evolving world. I am sincerely grateful for their mentorship and the impact they have had on my personal and professional development.

References

Cole, Rose. "The Problem with AI Music: Song and Cyborg Creativity in the Digital Age: Popular Music." Cambridge Core, August 27, 2020. https://www.cambridge.org/core/journals/ popular-music/article/problem-with-ai-music-song-and-cyborg-creativity-in-the-digital-age/F42497C1D564D33C15ABF5495E4BEEAC.

Dash, Adyasha, and Kat R. Agres. "AI-Based Affective Music Generation Systems: A Review of Methods, and Challenges." arXiv.org, January 10, 2023. https://arxiv.org/abs/2301.06890.

Gonsalves, Robert A. "Vicemorphia: Using AI for Style Transfer of Song Lyrics." Medium, October 23, 2021https://towardsdatascience.com/vicemorphia-using-ai-for-style-transfer-of-song-rics-fe22c052128d.

Hernandez-Olivan, Carlos, Javier Hernandez-Olivan, and Jose R. Beltran. "A Survey on Artificial Intelligence for Music Generation: Agents, Domains and Perspectives." arXiv.org, November 3, 2022 ttps://arxiv.org/abs/2210.13944.

Joe. "The Rise of Artificial Intelligence in Music." t3blog, January 20, 2023. https://www.thomann.de/blog/en/the-rise-of-ai-in-music-2/.

Locke, Charley. "Young Thug Isn't Rapping Gibberish, He's Evolving Language." Wired, October 15, 2015. https://www.wired.com/2015/10/young-thug-evolution-of-language/.

Thompson, Paul A. "Playlist Cart: Whole Lotta Red." Pitchfork, January 5, 2021. https://pitchfork.com/reviews/albums/playlist-cart-whole-lotta-red/.

Wales, Jeff. "Why Young Thug Is the 21st Century's Most Influential Rapper." BBC Culture, January 23, 2023. https://www.bbc.com/culture/article/20191021-why-young-thug-is-the-21st-centurys-most-influential-rapper.