From Words to Waves: How Gender Bias Perpetuates AI-Based Music Generation

Bruce Wu

Shanghai High School International Division，Shanghai，Republic of China

Brucewu1108@outlook.com

Abstract. This study investigates the influence of gender bias in song lyrics on AI-generated music, focusing on how biases in textual content can impact various acoustic features in the resulting compositions. Using a dataset from the Billboard Hot 100, gender bias was quantified through a Transformer-based model, allowing for the categorization of songs into the most biased and least biased groups. These songs served as inputs for the Suno AI platform, which generated new music based on the provided lyrics and genres. Acoustic features such as Aggressiveness, Danceability, Approachability, Engagement, Valence, and Arousal were then analyzed to identify differences between the two groups. The results revealed significant disparities in several features, particularly in Valence and Arousal, indicating that gender bias in lyrics can influence AI-generated music's emotional and rhythmic qualities. These findings demonstrate the potential for generative AI to perpetuate societal biases and highlight the importance of developing bias mitigation strategies. The study concludes by discussing theoretical and practical implications, proposing several methods to reduce bias in AI-generated content, and suggesting avenues for future Research to enhance fairness and inclusivity in AI-driven creative industries.

Keywords: AI-Generated Music, Media Bias, Acoustic Analysis.

1. Introduction

The concept of creating intelligent AI to assist with human research and activities has long been considered a key element of the future. This vision has rapidly developed and taken shape in various ways, leading to the emergence of generative AI as we know it today. Unlike traditional AI systems, which are primarily designed to recognize patterns and make decisions based on pre-existing data, generative AI models can create entirely new data instances that closely resemble the datasets they were trained on [1][2].

Generative AI can be applied in many fields, such as healthcare, art, and etc. The introduction of models such as DALL-E 2 by OpenAI in June 2022 and Stable Diffusion from Stability AI in December 2022 has made advanced generative models more accessible, enabling automatic generation of text, images, videos, and music. These models have revolutionized industries by enhancing efficiency, spurring creativity, and broadening access to information.

The impact of generative AI is substantial, with industry revenues surpassing $100 billion by 2023-2024 and projections showing continued growth [5]. A notable 70% of Gen Z has already interacted with generative AI tools, and nearly 90% of jobs in the United States could be influenced by AI [6]. By 2025, it is estimated that 95% of customer interactions will involve AI[7]. This level of adoption signals AI's dominance as a primary source of information and inspiration in society.

However, these advancements come with huge concerns, particularly regarding the spread of biases. Generative AI models are trained on large datasets from human culture, so they may inadvertently replicate and even amplify biases, discrimination, and prejudices present in the data [8][9]. Bias has always been a persistent issue in human society, often alienating certain groups based on differences from societal norms—norms that are frequently rooted in misunderstanding and ignorance.

Bias in generative AI has far-reaching consequences. For instance, when asked to generate images of specific professions, AI may exhibit bias toward certain age groups, genders, or races, such as predominantly depicting doctors as middle-aged white men [11][12]. Recent controversies, like those involving Google Gemini—which was criticized for failing to depict white individuals in its images and for producing content that lacked factual accuracy—highlight ongoing biases in AI [13]. Similarly, a Nature article discussed the prevalence of racist and sexist content in AI-generated material [14]. Biases in AI-generated content can subtly shape users' perceptions, especially among younger individuals who may not yet have fully developed critical thinking skills. Moreover, bias in AI erodes public trust in technology and institutions, exacerbates economic inequalities, and can lead to psychological and social harm. The ethical and legal concerns around accountability and fairness in AI further fuel debates about its role in society [15].

While much research has focused on bias in text and image generation, other emerging fields, especially AI generated music, remain relatively unexplored. As depicted in Figure 1, academic attention and publications on AI-generated music lag significantly behind those on image generation [16]. The release of music-generating models, like Suno v3.5 in March 2024, signals the increasing development of AI-driven music generation.

Similar to other forms of generative AI, it is very likely that biases will perpetuate in the music AI tools due to the training data. If the bias in the AI-generated music is underestimated or ignored, these contents, which are potential media for information dissemination, will not only limits the diversity of creativity, but also perpetuates social norms and marginalizes underrepresented voices in the music industry. However, the potential biases, including gender biases embedded in such systems have received little attention.

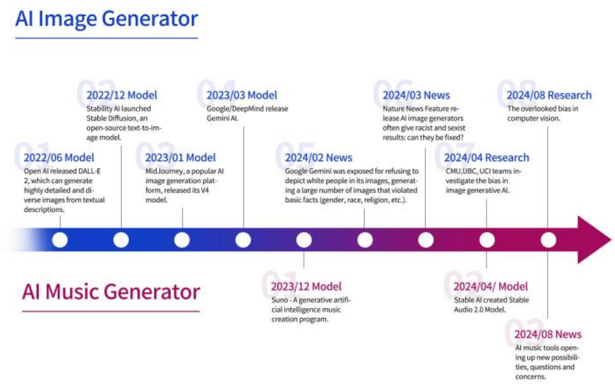

Figure 1: Development of Image Generative AI vs. Music Generative AI

Therefore, this study aims to address this gap by examining gender bias—defined as a preference or prejudice toward one gender over the other, often based on cultural norms and stereotypes, leading to unequal treatment in opportunities, rights, and representation [40]—in AI-generated music by analyzing both the lyrics and acoustic features of AI-generated music. This research contributes to the broader understanding of bias in generative AI, particularly how gender bias can shape music production. This innovative approach is significant as it explores an under-researched area, shedding new light on how AI technologies may inadvertently reinforce societal biases through music.

The research begins by collecting a dataset of songs composed by human artists, specifically selecting entries from the Billboard Hot 100. Using a pre-trained transformer model, I assess and quantify the gender bias in the lyrics of these compositions. I then selected the 30 songs with the highest bias and 30 with the lowest bias as inputs for the Suno AI model. The AI-generated music was then analyzed to explore how gender bias influenced the generation process. This study outlines a framework for generating music using biased lyrics, examines the impact of gender bias on AI music generation, and evaluates the outputs through acoustic analysis.

The rest of this paper is organized as follows: Section 2 reviews existing research on AI-generated art and bias in AI-generated content. Section 3 introduces the research framework and the data collection process. Section 4 investigates the acoustic features of the AI-generated songs and their correlation with the biases in the lyrics. Section 5 discusses the findings and presents the conclusion.

2. Related Work

Generative AI has become one of the most versatile and widely used tools in contemporary society, significantly improving processing capabilities and reshaping how tasks are performed. These AI models enhance efficiency, reduce the need for human labor, and increase productivity across various sectors. However, since generative AI models are trained on large datasets derived from human culture, growing concerns have emerged about the inherited biases that these systems may have. As the use of AI expands, the presence of bias has become more apparent, prompting governments, tech companies, and research institutions to prioritize addressing this issue.

Recognizing both the potential and risks of generative AI, governments around the world have begun to introduce policies aimed at managing these technologies. For example, China implemented the "Interim Measures for Generative Artificial Intelligence Service Management," which ensures AI development adheres to existing 2laws while encouraging innovation and international cooperation. In February 2024, National Public Radio (NPR) reported that Google's CEO Sundar Pichai demanded fixes to Google's generative AI model, Gemini, following criticism of the model's anti-white bias. Similarly, MIT Technology Review analyzed AI image-generation systems such as DALL-E 2 and Stable Diffusion, stressing the need to mitigate biases in AI-generated outputs. These incidents reflect the growing recognition that generative AI tools must be developed and used responsibly and equitably.

In recent years, a considerable amount of research has been conducted to analyze bias in AI, with a significant focus on image generation. The following studies, conducted between 2023 and 2024, highlight various societal biases in generative AI applications. For example, Anna M. Gorska et al. (2023) [17] examined gender bias in AI-generated images and found that men were represented in 76% of the images, while women appeared in only 8%. This study, although focused on earlier models, revealed the profound gender biases embedded within generative AI. Similarly, Federico Bianchi et al. (2023) [18] used self-edited text prompts to examine gender and racial bias in models like Stable Diffusion and DALL-E, finding systemic biases in how different genders and races are portrayed.

Further research by Jordan Vice et al. (2023) [19] explored social bias in Stable Diffusion models, showing how these biases manifest in AI-generated content. Using custom text prompt datasets, they uncovered biases targeting specific social groups. Similarly, Mi Zhou et al. (2024) [20] studied systematic gender and racial biases in AI image generators like Midjourney and Stable Diffusion. Through the use of the Face++ API, they analyzed facial features in AI-generated images, finding significant biases against women and African Americans.

Beyond image generation, text generation models like ChatGPT have also been subject to examination for bias. Xiao Fang et al. (2024) [21] assessed gender and racial biases in Large Language Models (LLMs), such as ChatGPT and LLaMA, by comparing AI-generated news content with original headlines from the New York Times and Reuters. They discovered substantial biases in the AI-generated text. In addition, Jochen Hartmann et al. (2023) [22] analyzed political and social biases in ChatGPT, concluding that the model exhibited a pro-environmental, left-libertarian ideology, showing clear preferences for certain political and social viewpoints.

These studies represent just a fraction of the growing research on bias in generative AI. As research into AI bias deepens, it becomes increasingly evident that generative AI has the potential to perpetuate societal biases in various ways. In this study, I aim to explore gender biases in generative AI from a new perspective—music. AI and music have already converged, as detailed in the survey by Yueyue Zhu et al. (2023) [16], which traces the development of AI music generators from the early 21st century. Techniques such as Markov chains, rule-based models, and evolutionary algorithms paved the way for more sophisticated music generation models like MUSPY [31], GenJam [32], and Jukebox. More recently, large transformer-based models such as MusicGen AI and Suno have expanded the capabilities of AI-driven music production. However, like other forms of generative AI, these systems may harbor latent biases. Therefore, it is crucial to uncover and evaluate these biases to ensure AI is used positively and integrated into society in ways that benefit humanity.

3. Methodological Framework

This study follows a structured methodological framework to explore how gender bias in song lyrics influences AI-generated music. The research process is divided into multiple steps, beginning with the collection of a comprehensive dataset from the Billboard Hot 100. Afterward, gender bias within the lyrics is evaluated using a Transformer-based model. The songs, categorized by their bias levels, are then input into Suno AI to generate new music. The AI-generated songs are subsequently analyzed for acoustic features, focusing on key aspects such as valence and danceability. Finally, statistical analysis is performed to examine the relationship between the bias present in the original lyrics and the characteristics of the AI-generated music. The flowchart below (Figure 2) illustrates the sequence of this process.

Figure 2: Flowchart of the Research Process

3.1. Data Collection

For this research, a dataset was built using compositions by human artists, with the Billboard Hot 100 serving as the primary data source. The Billboard Hot 100 is widely regarded as a reliable platform that ranks popular songs based on sales—both physical and digital—dating back to 1958. As such, it is considered a benchmark for success in the music industry. It also serves as one of the most comprehensive collections of popular music, making it the perfect source for this study's dataset. I compiled all songs recorded on the Billboard Hot 100 chart from 51965 to 2021, resulting in a dataset that includes song titles, lyrics, composers' names, and years. After removing redundant data, such as repeated songs that appeared on the chart in different years, the final dataset consisted of 5,702 unique songs that are varied in styles and genres. The word cloud in Figure 3 provides an overview of the most common words found in the lyrics from this dataset.

Figure 3: Word Cloud of Lyrics from Billboard Hot 100 (1965-2021)

However, while these high-frequency words reflect common themes or repetitive language in the lyrics, they do not offer direct insights into the underlying biases. These words cannot reveal the extent or presence of gender bias, nor can they show how such bias might influence the characteristics of AI-generated music. This limitation is one of the key reasons this study is necessary. Moving beyond superficial analysis, we aim to uncover more profound and nuanced relationships between bias in human-generated lyrics and the features of AI-generated music.

3.2. Bias Estimation

The next important step was to evaluate the gender bias present in each song in the Billboard Hot 100 dataset. This evaluation lays the groundwork for understanding the potential impact of gender bias on music generated by Suno AI. It provides an initial benchmark for how biased the original lyrics are, which is essential for subsequent analysis.

To carry out this task efficiently, I used a Transformer-based bias detection model from the Dbias open-source package, created by Shaina Raza et al., [33]. and hosted by Hugging Face, a popular platform for machine learning tools. More specifically, DistilBERT, a more compact and faster version of the BERT (Bidirectional Encoder Representations from Transformers) model, is especially suited for dynamic word embedding, which makes it ideal for detecting gender bias. Since detecting bias requires understanding semantic and contextual meanings in text, DistilBERT is well equipped for this purpose. The model's architecture also makes it efficient for processing large datasets like the one used in this study.

The DistilBERT bias detection model was fine-tuned and pre-trained on the MBIC (Media Bias Annotation Dataset Including Annotator Characteristics) dataset [34]. This dataset contains 1,700 sentences from around 1,000 articles, all of which may contain various forms of media bias. This model was chosen because it is highly effective for evaluating gender bias in textual data, such as song lyrics, which share similarities with media texts.

To perform the bias detection, I set up a Python virtual environment using Google Colab. From there, I built a pipeline using the bias detection model from the Dbias package. The pipeline included a tokenizer to break the lyrics into tokens, which could then be input into the model for evaluation. I set the maximum token sequence length to 500 to ensure thorough processing. Once the model completed its run, a new column with gender bias scores was added to the dataset.

3.3. AI Music Generation

Once the gender bias evaluation was completed, the next step was to generate and collect the AI music. I focused on organizing the initial dataset and gathering the necessary lyrics and genres for the music generation phase. In the end, I formed a dataset of AI-generated music.

For this step, I selected 60 songs from the original dataset—30 with the highest gender bias and 30 with the lowest. This selection strategy was essential to create a clear contrast between AI-generated music produced from 6high-bias and low-bias lyrics, facilitating a more detailed analysis of how gender bias affects the music generation process.

I gathered the lyrics and genre information needed for generating the music. Rather than extracting the lyrics directly from the original dataset, I used Genius, a well-known online music encyclopedia. Genius was particularly useful for this purpose because it provides well-structured lyrics, with sections like the chorus and verse clearly delineated. It also includes genre tags, which are important for ensuring that the generative AI models have consistent and structured input. This setup helps minimize variables and reduces uncertainty in the AI-generated music outputs.

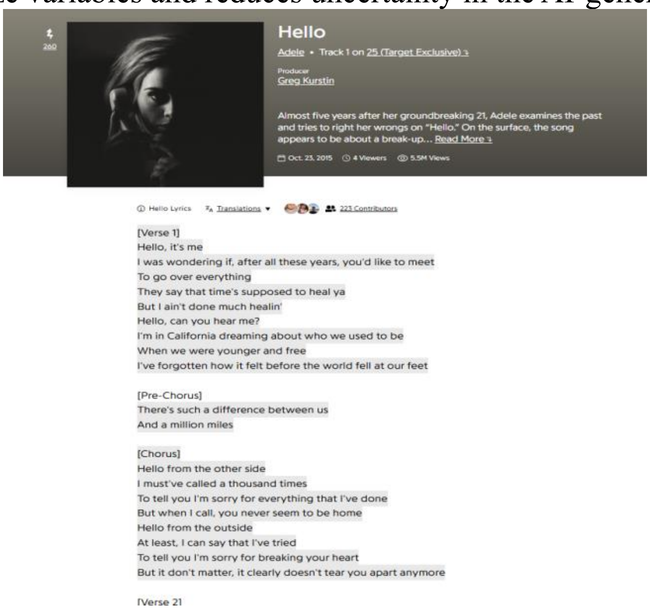

Figure 4: "Hello"(By Adele) Lyrics from Genius' webpage

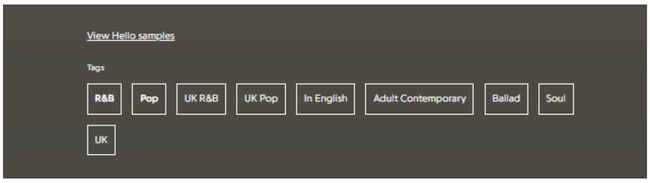

Figure 5: Tags with the theme of the song "Hello"(By Adele) on the Genius webpage.

I used the Suno AI platform, accessible through its official website, to generate the AI music. Figure 6 shows the main page of the platform. Suno AI allows users to create songs by inputting both lyrics and genre information, which are then processed through its AI model. For this study, I enabled the customization mode to ensure that both lyrics and genre information closely adhered to the original structure of each song. Figure 7 shows how the lyrics and genre information were input for music generation. However, during this process, one song from the low-bias category violated Suno AI's generation criteria, reducing the number of songs for analysis from 60 to 59.

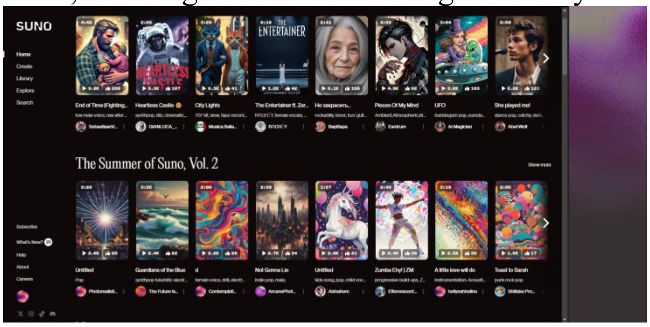

Figure 6: Home page of Suno offfcial website

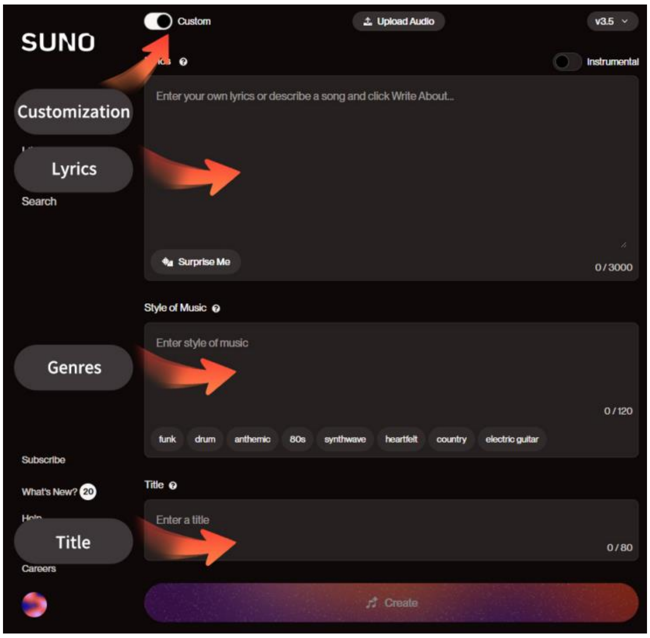

Figure 7: Suno AI Music Creation Interface

3.4. Acoustic Features Extraction and Analysis

After inputting the lyrics, title, and genre, Suno AI processed and generated the songs. The songs were then presented on the platform along with an AI-generated cover image. The final dataset included 59 songs, as one song from the low-bias group had been excluded due to a violation of Suno AI's generation criteria.

To investigate how gender bias in the lyrics affected the characteristics of the AI-generated music, I conducted a computational analysis focused on various acoustic features. While the analysis of lyrics targets the textual content, the acoustic feature analysis examines the musical components of the song. This approach allowed for a detailed comparison of both obvious and subtle features in the AI-generated music, providing insight into how bias in the input lyrics influences the musical output. Specifically, to clearly reflect the potential impact of gender bias on the generated songs, I selected high level acoustic features that best represent the emotional and thematic expression of the songs

For acoustic feature extraction, I used the Essentia library, an open-source C++ library designed for audio analysis and music information retrieval [35]. Essentia offers a variety of pre-trained state-of-the-art music tagging and classification models tailored to extracting different acoustic features, and it provides metadata, demo data, model weights, and Python code for each feature on its official website. This makes it an ideal tool for conducting comprehensive acoustic analysis in a virtual Python environment, which was set up in Google Colab.

The analysis focused on the following acoustic features:

Arousal/Valence: This feature assesses the emotional content of the music. I used the deam-msd-musicnn model, which was trained on the DEAM dataset [36]. This dataset consists of 1,802 music excerpts and full songs annotated with continuous valence and arousal values. The msd-musicnn model was used as the embedding extractor, feeding into the deam-msd-musicnn model [37].

Engagement: The engagement level of the music, which measures how captivating the song is, was evaluated using the engagement regression-discogs-effort model trained on Discogs-4M dataset [38]. This model outputs continuous engagement values, extracting embeddings using the discogs-effort-bs64 model using contrastive learning method to learn the music similarity.

Approachability: Approachability, reflecting how inviting or accessible the music is perceived to be to the public (e.g. belonging to mainstream or niche and experimental music genres, was analyzed as continuous values using the approachability regression-discogs-effort model [38], which also uses the discogs-effort-bs64 model for embedding extraction.

Aggressiveness: To evaluate the degree of aggressiveness or intensity in the music, the mood-aggressive-discogs-effnet model was applied to classify the music into aggressive vs. non-aggressive [39], again utilizing the discogs-effnet-bs64 model for embeddings.

Danceability: The danceability of each song, which indicates how suitable the music is for dancing, was assessed using the danceability-discogs-effort model [38]. As with the other features, the discogs-effort-bs64 model provided the necessary embeddings for this evaluation.

By applying these models, I systematically evaluated the acoustic features of the AI-generated music to identify any correlations between the gender bias in the input lyrics and the resulting musical characteristics.

This analysis is an important contribution to understanding how biases in human-generated content can be transferred—and possibly amplified—in AI-generated outputs. It underscores the need for careful consideration of bias when developing and using AI systems.

4. Results

4.1. Bias Estimation Result

After assessing the gender bias in each song within the initial dataset using the DistilBERT bias detection model, the gender bias values ranged from 0.500 to 0.996. Figure 8 presents a histogram that provides an overview of the gender bias values across the entire dataset, revealing the distribution of data across different value ranges. Notably, approximately 50% of the data exhibits a gender bias value exceeding 0.95, indicating a high level of gender bias.

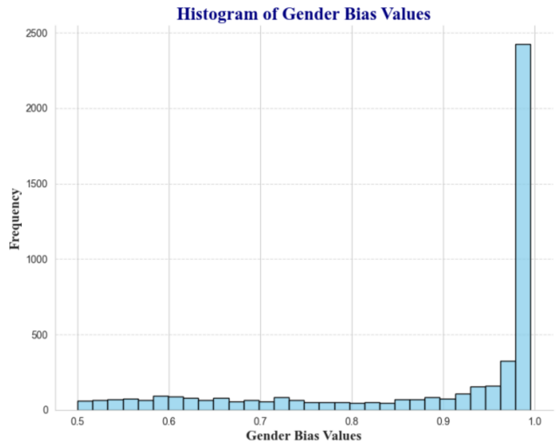

Figure 8: Histogram of Gender Bias Values

Based on these evaluated gender bias values, I sorted the dataset and selected the top 30 songs with the highest gender bias and the bottom 30 songs with the lowest gender bias. As outlined in the framework section, these selections are crucial for ensuring a clearer insight into the effect of gender bias on AI-generated music compositions. Figure 9 illustrates the 60 data points extracted from the initial dataset, with bars 1 to 30 representing the songs with the lowest bias (the least biased group) and bars 31 to 60 representing the songs with the highest bias (the most biased group).

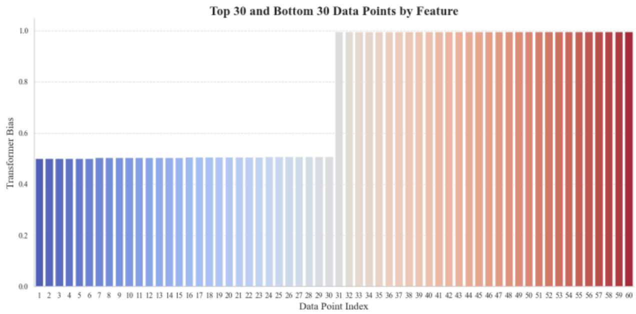

Figure 9: TOP 30 Bias and Bottom 30 Bias Songs

4.2. Music Generation Result

Except for one song that violated the generation criteria of Suno AI, 59 AI-generated music tracks were successfully produced using Suno AI, based on the original lyrics and genres of the selected songs. These tracks include 29 from the least biased group and 30 from the most biased group. Originally output as .mp3 files, all tracks were converted into .wav format and stored in separate categories (least biased and most biased) for subsequent acoustic feature evaluations. Figures 10 and 11 display example waveforms from the generated .wav files, with Figure 10 showcasing tracks from the most biased group (e.g., Wide Awake by Katy Perry, We Built This City by Starship) and Figure 11 illustrating tracks from the least biased group (e.g., The Way You Move by OutKast, Miss You by The Rolling Stones).

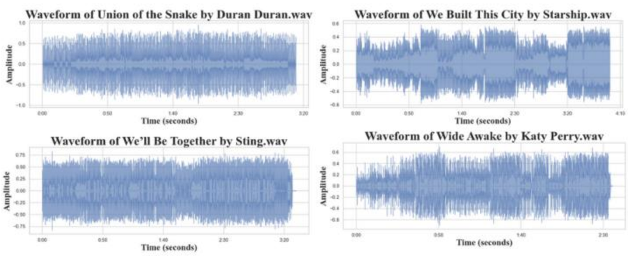

Figure 10: Examples Waveform from Top Bias

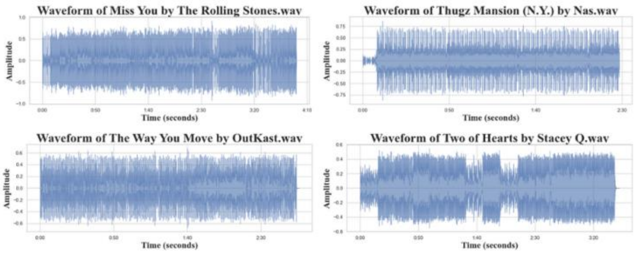

Figure 11: Examples Waveform from Bottom Bias

4.3. Acoustic Feature Results

The results of acoustic feature evaluations are key to assessing how gender bias in the original lyrics may perpetuate in Suno AI's generated music. Differences between the least and most biased groups across various acoustic features provide direct evidence supporting the study's main objectives. The analysis focused on six acoustic features: Aggressiveness, Danceability, Approachability, Engagement, Valence, and Arousal. I calculated the mean values of each acoustic feature for both groups and conducted t-tests (statistical hypothesis tests) to examine 11the differences between the means of the two groups. The null hypothesis (H0) posited that songs from both groups would exhibit similar levels of acoustic features. A swarm plot was generated for each acoustic feature to visualize the differences between the two groups and provide a preliminary understanding of the numerical values for each dataset.

Table 1: Acoustic Feature Comparison Between the Most Biased Group and the Least Biased Group

Acoustic features	High group	Low group	t-stats	p-values
Mood Aggressive	0.135	0.068	1.84	p<0.1*
Danceability	0.773	0.599	1.99	p<0.05**
Engagement	0.800	0.743	-1.83	p<0.1*
Approachability	0.834	0.810	0.789	p>0.1
Valence	5.650	5.142	3.41	p<0.01***
Arousal	5.806	4.954	4.78	p<0.01***

p<0.01***, p<0.05**, p<0.1*

From the six acoustic features analysis, the p-values for Mood Aggressiveness and Engagement approached the 0.05 significance level, with values of 0.07 and 0.07, respectively. These results suggest a noteworthy difference between the two groups for these features, challenging the null hypothesis (H0) and indicating that the observed differences are not due to random chance. A p-value of 0.07 implies that there is only a 7% probability of concluding the null hypothesis is true, highlighting a statistically significant difference between the most biased and least biased groups in terms of aggressiveness and engagement.

The Danceability feature exhibited a p-value of 0.05, indicating a significant difference between the two groups, which is more pronounced than the differences observed for Mood Aggressiveness and Engagement. Furthermore, Valence and Arousal reflected the most statistically significant differences, with p-values of 0.001 and close to 1e-5, respectively. These findings indicate that these two features vary significantly between the most biased and least biased groups, suggesting that gender bias in the input lyrics plays a distinct role in shaping these acoustic characteristics in AI-generated music.

However, the t-test for the Approachability feature did not reveal significant differences between the two groups, with a p-value of 0.43. This high p-value suggests a strong likelihood that the null hypothesis holds true for this feature, indicating no substantial difference in Approachability between the least biased and most biased groups.

Figure 12 consists of six swarm plots, each representing a comparison of a specific acoustic feature between two song categories: the low-bias group (songs with the lowest gender bias) and the high-bias group (songs with the highest gender bias). These visualizations allow us to assess how different acoustic features are distributed across the two groups and to detect any trends or patterns in the data:

The values for aggressiveness tend to be higher in the high-bias group, with more songs clustering around the 0.1-0.2 range. In contrast, the low-bias group shows a more compact distribution with generally lower aggressiveness values. This suggests that songs generated from high-bias lyrics tend to have a more aggressive sound.

Regarding engagement, the high-bias group has a slightly higher distribution of engagement values compared to the low-bias group, with most values falling between 0.7 and 0.9. The low-bias group shows more clustered values around 0.7, indicating that songs from the high-bias group might be perceived as more engaging.

In terms of danceability, the figure shows that the danceability scores of the high bias group are higher, concentrated in the range of 7.5-8.5. The danceability values of the low bias group are lower, generally between 6.5 and 7.5. This shows that songs generated by lyrics with a higher degree of gender bias may tend to have more rhythmic and physical attractive elements.

The values of approachability are closely distributed across both the low-bias and high-bias groups, indicating that gender bias in lyrics does not significantly influence the approachability of the generated music.

The arousal plot highlights a noticeable difference between the two groups. The high-bias group has higher arousal values, with several songs clustering between 50 and 60. The low-bias group shows generally lower arousal values, indicating that songs generated from high-bias lyrics tend to evoke stronger emotional intensity or excitement.

In the subplot of valence, songs from the high-bias group tend to have higher valence values (closer to 5.5 and 6), suggesting a more positive emotional tone. In contrast, songs from the low-bias group exhibit lower valence values, clustering closer to 5, indicating a more neutral or less positive emotional tone.

As such, these visualizations further support the findings from the acoustic feature evaluations, which show that gender bias in lyrics can significantly influence the emotional and rhythmic aspects of AI-generated music.

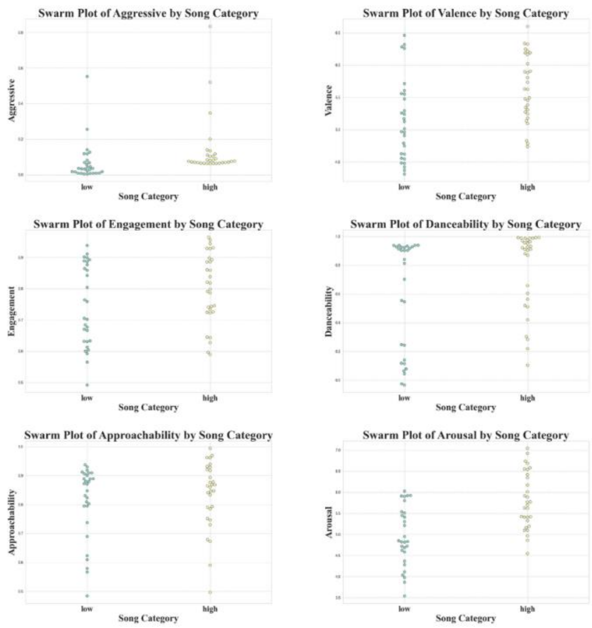

Figure 12: Distributions of six acoustic features between songs from lowest biased group and highest biased group

5. Discussion and Conclusions

This study set out to explore how gender bias in song lyrics might influence AI-generated music, focusing on various acoustic features of the compositions. By analyzing songs from the Billboard Hot 100 using a Transformer based model, I categorized songs into the most biased and least biased groups and generated new music using Suno AI based on these songs.

The findings revealed statistically significant differences between the two groups in several acoustic features. Notably, Valence and Arousal showed the most significant disparities, indicating that songs generated from highly biased lyrics tend to have higher emotional intensity and arousal levels. Danceability also differed significantly, knowing that the genres and styles are varied in the dataset, suggesting that gender bias in lyrics could influence the rhythmic and physical aspects of AI-generated music. Mood Aggressiveness and Engagement showed less pronounced but still notable differences, further underscoring the impact of gender bias on the musical output.

These results contribute to the growing body of literature on bias in AI systems, particularly in the domain of generative AI for music. They highlight the potential for AI models to perpetuate and even amplify human biases present in their training data, extending these biases into domains like music generation, which have been less examined. The study underscores the importance of considering non-textual outputs when assessing the impact of bias in generative models, suggesting that biases in textual inputs can have complex and far-reaching effects on other media forms produced by AI.

The findings support the broader theoretical framework that views AI as both a reflection and a reproduction of societal norms embedded within its training datasets. In the context of music generation, this implies that AI tools not only replicate existing cultural biases but may also reinforce them through widespread dissemination of biased content. This reinforces the need for a deeper understanding of how biases in AI-generated media can influence cultural narratives and societal perceptions.

From a practical standpoint, these findings have significant implications for artists, producers, and developers using AI-based music generation tools. It serves as a cautionary reminder of the potential unintended consequences of integrating AI into creative processes. Recognizing that AI-generated music can carry and propagate biases from the source material is crucial for making informed decisions in the creative industry.

Moreover, this study suggests the need for more robust bias detection and mitigation strategies in developing AI generative models. As AI continues to play a larger role in content creation, ensuring that these systems produce fair and equitable outputs becomes increasingly important, especially in culturally influential industries like music.

Given the findings, it is clear that bias mitigation strategies are essential to address the potential for AI systems to perpetuate gender bias. Several strategies could be employed to reduce the impact of bias in AI-generated music:

1. Data Curation and Augmentation: Carefully curating and augmenting training datasets to ensure balanced gender representation and content diversity can help mitigate embedded biases. Including a wider variety of music genres, cultures, and time periods can reduce the dominance of any single perspective.

2. Bias Detection and Correction: Integrating real-time bias detection mechanisms into the AI music generation process can help identify and correct biased outputs before they are finalized. Techniques like adversarial debiasing can ensure the generated music aligns with fairness criteria.

3.Algorithmic Transparency and Interpretability: Enhancing the transparency and interpretability of AI models allows developers and users to understand how decisions are made and where biases might be introduced. This transparency can lead to ethical guidelines or standards for AI use in music generation.

4.Human-in-the-Loop(HITL)Approaches: Incorporating human oversight in the AI music generation process can help reduce bias. Human reviewers can evaluate outputs for potential biases and make adjustments before release, combining human judgment with AI efficiency.

5.Fairness Constraints in Model Training: Implementing fairness constraints during model training can reduce biased outcomes. Penalizing the model for producing outputs that exhibit gender bias encourages the generation of more balanced content.

This study opens several avenues for future research. Exploring other types of biases, such as racial or cultural biases in AI-generated music, could provide further insights. Additionally, investigating the long-term effects of consuming biased AI-generated music on listeners' attitudes and behaviors is another important area.

Developing more sophisticated bias mitigation techniques that can be integrated directly into the AI music generation process is also crucial. Future studies could focus on the effectiveness of different debiasing algorithms or explore multimodal approaches that combine text, audio, and visual data to reduce bias.

As AI continues to evolve, establishing industry standards and ethical guidelines for its use in creative industries is essential. These standards should emphasize fairness and inclusivity, ensuring that technology enhances rather than detracts from cultural diversity and social equity.

References

[1] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. Advances in neural information processing systems, 27, 2014.

[2] Mark S. Cole S. What is generative ai?, 2024. https://www.ibm.com/topics/generative-ai.

[3] Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. Hierarchical text-conditional image generation with clip latents. arXiv preprint arXiv:2204.06125, 1(2):3, 2022.

[4] Open AI. Generative models,2016. https://openai.com/index/generative-models/.

[5] Bloomberg. Generative ai to become a $1.3 trillion market by 2032, research finds, 2017. https://www.bloomberg.com/company/press/generative-ai-to-become-a-1-3-trillion-market-by-2032-research-ffnds/.

[6] John Koetsier. Generative ai generation gap: 70% of gen z use it while gen x, boomers do not get it, forbes,2017. https://www.forbes.com/sites/johnkoetsier/2023/09/09/generative-ai-generation-gap-70-of-gen-z-useit-while-gen-x-boomers-dont-get-it/.

[7] Ed Lauder. Ai will power 95% of customer interactions by 2025, ai business, 2017. https://aibusiness.com/automation/ai-will-power-95-of-customer-interactions-by-2025.

[8] Aylin Caliskan, Joanna J Bryson, and Arvind Narayanan. Semantics derived automatically from language corpora contain human-like biases. Science, 356(6334):183–186, 2017.

[9] Tolga Bolukbasi, Kai-Wei Chang, James Y Zou, Venkatesh Saligrama, and Adam T Kalai. Man is to computer programmer as woman is to homemaker? debiasing word embeddings. Advances in neural information processing systems, 29, 2016.

[10] Saffya Umoja Noble. Algorithms of oppression: How search engines reinforce racism. In Algorithms of oppression. New York university press, 2018.

[11] Joy Buolamwini and Timnit Gebru. Gender shades: Intersectional accuracy disparities in commercial gender classiffcation. In Conference on fairness, accountability and transparency, pages 77–91. PMLR, 2018.

[12] H. Wiltshire. Revealing ai bias: Reinforcing harmful workplace stereotypes. rippl., 2024. https://rippl.work/blog/research-ai-bias-in-workplace-stereotypes/.

[13] Adi Robertson. Google apologizes for 'missing the mark' after gemini generated racially diverse nazis, 2024.https://www.theverge.com/2024/2/21/24079371/google-ai-gemini-generative-inaccurate-historical.

[14] Douglas Guilbeault, Sole`ne Delecourt, Tasker Hull, Bhargav Srinivasa Desikan, Mark Chu, and Ethan Nadler.Online images amplify gender bias. Nature, 626(8001):1049–1055, 2024.

[15] Ruha Benjamin. Race after technology: Abolitionist tools for the new Jim code. John Wiley & Sons, 2019.

[16] Yueyue Zhu, Jared Baca, Banafsheh Rekabdar, and Reza Rawassizadeh. A survey of ai music generation tools and models. arXiv preprint arXiv:2308.12982, 2023.

[17] Anna M Gorska and Dariusz Jemielniak. The invisible women: uncovering gender bias in ai-generated images of professionals. Feminist Media Studies, 23(8):4370–4375, 2023.

[18] Federico Bianchi, Pratyusha Kalluri, Esin Durmus, Faisal Ladhak, Myra Cheng, Debora Nozza, Tatsunori Hashimoto, Dan Jurafsky, James Zou, and Aylin Caliskan. Easily accessible text-to-image generation ampliffes demographic stereotypes at large scale. In Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency, pages 1493–1504, 2023.

[19] Jordan Vice, Naveed Akhtar, Richard Hartley, and Ajmal Mian. Quantifying bias in text-to-image generative models. arXiv preprint arXiv:2312.13053, 2023.

[20] Mi Zhou, Vibhanshu Abhishek, Timothy Derdenger, Jaymo Kim, and Kannan Srinivasan. Bias in generative ai. arxiv. arXiv preprint arXiv:2403.02726, 2024.

[21] Xiao Fang, Shangkun Che, Minjia Mao, Hongzhe Zhang, Ming Zhao, and Xiaohang Zhao. Bias of ai-generated content: an examination of news produced by large language models. Scientiffc Reports,14(1):5224, 2024.

[22] Jochen Hartmann, Jasper Schwenzow, and Maximilian Witte. The political ideology of conversational ai: Converging evidence on chatgpt's pro-environmental, left-libertarian orientation. arXiv preprint arXiv:2301.01768, 2023.

[23] John J Hanna, Abdi D Wakene, Christoph U Lehmann, and Richard J Medford. Assessing racial and ethnic bias in text generation for healthcare-related tasks by chatgpt1. MedRxiv, 2023.

[24] Shicheng Xu, Danyang Hou, Liang Pang, Jingcheng Deng, Jun Xu, Huawei Shen, and Xueqi Cheng. Invisible relevance bias: Text-image retrieval models prefer ai-generated images. In Proceedings of the 47th international ACM SIGIR conference on research and development in information retrieval, pages 208–217, 2024.

[25] Sourojit Ghosh and Aylin Caliskan. 'person'== light-skinned, western man, and sexualization of women of color: Stereotypes in stable diffusion. arXiv preprint arXiv:2310.19981, 2023.

[26] David Rozado. The political biases of chatgpt. Social Sciences, 12(3):148, 2023.

[27] Fabio Motoki, Valdemar Pinho Neto, and Victor Rodrigues. More human than human: measuring chatgpt political bias. Public Choice, 198(1):3–23, 2024.

[28] Yixin Wan, Arjun Subramonian, Anaelia Ovalle, Zongyu Lin, Ashima Suvarna, Christina Chance, Hritik Bansal, Rebecca Pattichis, and Kai-Wei Chang. Survey of bias in text-to-image generation: Deffnition, evaluation, and mitigation. arXiv preprint arXiv:2404.01030, 2024.

[29] Ranjita Naik and Besmira Nushi. Social biases through the text-to-image generation lens. In Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society, pages 786–808, 2023.

[30] Yan Tao, Olga Viberg, Ryan S Baker, and Rene´ F Kizilcec. Cultural bias and cultural alignment of large language models. arXiv preprint arXiv, 2311, 2024.

[31] Hao-Wen Dong, Ke Chen, Julian McAuley, and Taylor Berg-Kirkpatrick. Muspy: A toolkit for symbolic music generation. arXiv preprint arXiv:2008.01951, 2020.

[32] John Biles et al. Genjam: A genetic algorithm for generating jazz solos. In ICMC, volume 94, pages 131–137.Ann Arbor, MI, 1994.

[33] Shaina Raza, Deepak John Reji, and Chen Ding. Dbias: detecting biases and ensuring fairness in news articles. International Journal of Data Science and Analytics, 17(1):39–59, 2024.

[34] Timo Spinde, Lada Rudnitckaia, Kanishka Sinha, Felix Hamborg, Bela Gipp, and Karsten Donnay. Mbic–a media bias annotation dataset including annotator characteristics. arXiv preprint arXiv:2105.11910, 2021.

[35] Dmitry Bogdanov, Nicolas Wack, Emilia Go´mez Gutie´rrez, Sankalp Gulati, Herrera Boyer, Oscar Mayor, Gerard Roma Trepat, Justin Salamon, Jose´ Ricardo Zapata Gonza´lez, Xavier Serra, et al. Essentia: An audio analysis library for music information retrieval. In Britto A, Gouyon F, Dixon S, editors. 14th Conference of the International Society for Music Information Retrieval (ISMIR); 2013 Nov 4-8; Curitiba, Brazil.[place unknown]: ISMIR; 2013. p. 493-8. International Society for Music Information Retrieval (ISMIR), 2013.

[36] [36] Anna Alajanki, Yi-Hsuan Yang, and Mohammad Soleymani. Benchmarking music emotion recognition systems. PloS one, pages 835–838, 2016.

[37] Jordi Pons and Xavier Serra. musicnn: Pre-trained convolutional neural networks for music audio tagging. arXiv preprint arXiv:1909.06654, 2019.

[38] Pablo Alonso-Jime´nez, Xavier Serra, and Dmitry Bogdanov. Music representation learning based on editorial metadata from discogs. In Rao P, Murthy H, Srinivasamurthy A, Bittner R, Caro Repetto R, Goto M, Serra X, Miron M, editors. Proceedings of the 23nd International Society for Music Information Retrieval Conference (ISMIR 2022); 2022 Dec 4-8; Bengaluru, India.[Canada]: International Society for Music Information Retrieval; 2022. p. 825-33. International Society for Music Information Retrieval (ISMIR), 2022.

[39] Cyril Laurier, Owen Meyers, Joan Serra, Martin Blech, and Perfecto Herrera. Music mood annotator design and integration. In 2009 Seventh International Workshop on Content-Based Multimedia Indexing, pages 156–161. IEEE, 2009.

[40] UNICEF. (2022). Gender action plan, 2022-2025: A vision for lasting, transformative change. Retrieved from https://www.unicef.org