|  |
| --- |
| University of Vermont College of Engineering and Mathematical Sciences |
| Gender Bias in Spotify Recommendation Algorithms |
| Advisor: Lisa Dion |

|  |
| --- |
| Sydney White  9-30-2023 |

# Introduction

As algorithms become more widely used through the adoption of artificial intelligence, algorithmic bias has become more and more prevalent. These algorithms and their implicit bias amplify the bias present already in daily life and can lead to disadvantageous outcomes for oppressed groups. Specifically, recommendation algorithms which govern much of how we use technology and consume media—from short videos on social media to browsing movie titles to streaming a song or podcast—escalate the inequities already present in our society. This project aims to analyze gender bias in music recommendation algorithms, given the existing bias and gender disparity in the music industry.

Specifically, this project aims to answer the following question: Is there gender bias in Spotify’s recommendation algorithm for users’ “Discover Weekly” playlist, and if so, what factors can moderate its effects? This will be accomplished by generating a variety of Spotify users with different constructed identities and listening histories and analyzing their recommended output for 10 weeks. Each user’s proportion of artists recommended by gender can be compared for significant differences and indications of bias.

Ultimately, this research aims to identify factors that can reduce the effects of algorithmic bias, specifically gender bias in recommendation algorithms. It will contribute to a broader understanding of the gender disparity in the music industry, algorithmic bias, and how these two forms of oppression may amplify each other.

# Background

Women in Music

It is no secret that the music industry, like many others, is dominated by men. The extent is difficult to measure and analyze due to the creative and individual nature of music. Since 2012, the USC Annenberg Inclusion Initiative has attempted to capture these inequities by analyzing inclusion on the Billboard Hot 100 Year-End Chart. In 2022, there were a total of 160 artists credited on the 100 songs featured on the Hot 100 Year-End Chart—69.4% men, 30% women, and 0.6% gender non-binary (Smith, 2023, 4). This is a 7.3 percentage point improvement since 2012 for women. They also analyzed songwriters and producers. Of the 451 songwriters on the Chart in 2022, 14% were women, and only 3 credits belonged to gender non-binary songwriters (Smith, 2023, 12). There was a total of 231 producers credited on the Chart in 2022, 3.4% of whom were women, and only one of whom was gender non-binary (Smith, 2023, 16). It is notably that all of these non-binary metrics—artists, songwriters, and producers—belong to the same nonbinary artist, Sam Smith (Smith, 2023). The USC Annenberg report captures the inequities in the music industry; however it falls short in explaining their causes and only addresses the most popular mainstream music. It largely ignores the way music consumption has become an individual experience through the rise of streaming.

Italian researchers attempted to use streaming data from Spotify to assess gender discrimination among Italian singers. The study used the number of listeners on Spotify as a measure of success for each artist. They ultimately concluded there is no gender discrimination among Italian singers as there was no significant difference in number of listeners by gender, despite the average number of listeners for male singers being consistently higher than that of female singers (Nappo, 2019). Their analysis reveals a difficulty in measuring gender bias, despite the existing gender disparity in the music industry, especially using Spotify metrics. By choosing to compare an equal number of male and female artists, the study fails to account for the forces that make a user more likely to listen to a male artist from industry bias to user interface. Furthermore, number of listeners is an unstable metric for comparison as not all listeners are created equal (or contributing to an artist’s monetary success in the same way). One listener could have listened to an artist’s song once in a 28-day period, while another could have streamed the artist’s entire discography, and they each count as one listener. Due to this discrepancy, it is entirely possible for male and female artists to have similar numbers of listeners, but vastly different numbers of streams and therefore success.

While the USC Annenberg report addresses popular music, interviews from 17 “female digital musicians” describe the experience of being a self-taught musician performing online without a record label. The study focused on musicians using YouTube, Soundcloud, and Vine, however Spotify is another online platform where independent artists can promote their work. The musicians cite expectations of their physical appearance as a barrier to success. They see the differences in portrayal of male and female musicians in the music industry and rewarding of female sexualized body parts as reinforcing oppressive societal expectations about appearance for female musicians, even those who are not famous (Choi, 2016). This study captures the individual experience of being a female musician promoting work online including barriers to success related to the gender disparity in the music industry and gender-based oppression as a whole.

Gender

All the previously mentioned studies convolute sex and gender, excluding other gender identities and expressions. In her work “Queering feminist technology studies”, Landström suggests that heteronormativity influences feministic research on gender and technology. She calls for a theoretical shift in framework to account for technology’s coproduction of gender to avoid empirical research that relapses into old patterns of masculine and feminine binary (Landström, 2007).

Many studies on the classification of gender or technology and gender fall into this archaic thinking. For example, a study from Verizon, Yahoo, and Worcester Polytechnic Institute attempted to create a machine learning model to predict a user’s gender based on their username. They only allowed for two outcomes: male or female. They found a character-based logistic regression approach was most accurate for gender classification, though there are problems with all attempted gender classification models (Hu, 2021). Their work demonstrates that name is not enough information to make accurate gender predictions as well as the shortcomings of binary gender classification.

Gender classification is often attempted when a user’s gender is not supplied. Even as sites have expanded past the gender binary in account creation, the gender binary is reinforced elsewhere behind the scenes. An immerse ethnographic study of the 10 most popular English speaking social media sites revealed that while many sites are expanding gender options outside of the binary, demographic information is still heavily used in advertising. They stress the social implication of gender in social media and technology as they become ingrained in our society, especially for identity curation and performance (Rena, 2016).

Gender and Algorithmic Bias

This ignorance of gender identities as well as existing gender bias breeds bias in our algorithms which perpetuates the heteronormativity, ignorance, and oppression which contribute to it. Algorithmic bias extends far beyond the context and application of the algorithm itself and is already apparent in daily life. Perez’s “Invisible Women” details how the gender data gap, the systemic and historic lack of data collected on or about women, perpetuates the oppression that created it. Perez warns this biased input data leaves the impression that men are the default and majority. Ultimately, these machines are not only reflecting but also amplifying our biases (Perez, 2021).

Recommendation algorithms are particularly suited to amplify these effects. YouTube’s recommendation algorithm has already been shown to favor feminized content, creating a vicious cycle that rewards oppressive norms. YouTube systematically promoted feminized material by rewarding and promoting feminized content in beauty vlog recommendations. When failing to account for potential bias, developers create algorithms with unintended side effects which amplify existing societal problems such as discrimination, gender bias, and gendered societal expectations (Bishop, 2018). If recommendation algorithms are amplifying gender oppression on YouTube, similar algorithms may be amplifying it on Spotify.

These unintended effects of biased algorithms have unintended effects themselves. Researchers at Stanford found women receiving stereotypically “feminine” career recommendations had lower estimates of leadership ability than those who received traditionally “masculine” career recommendations and believed the recommendations were based on internal characteristics. An understanding of these systems, algorithms with implicit bias, moderates its effect on sense of self but cannot remove them entirely (French, 2018). Given the sexualization of female musicians and gender disparity in the industry, algorithmic bias could be perpetuating this disparity on a psychological level in those exposed to its effects.

Spotify

Spotify algorithms have been analyzed in a variety of ways in the past, though little is known about the gender discrepancy in their output. It is suggested that Spotify uses content-based and collaborative filtering to generate track representations. Content-based filtering includes the content of each track, such as artist name, songwriting credits, and genre, but also metrics Spotify calculates based on these attributes such as danceability as well as a semantic meaning estimated with natural language processing models. Collaborative filtering describes the track’s connection to other tracks and what other users with similar tastes are listening to. Each user on Spotify also has their own “taste profile” or Spotify’s understanding of what the user likes based on their listening history (Pastukhov, 2022).

While Spotify reveals a lot to users and artists about how its algorithms work, it is currently unknown whether artist gender is meaningfully included in the content-based filtering, and it is also unknown whether there is a data gap in music from female-identifying musicians based on the near infinite amount of content on Spotify. As such, it is difficult to know if there is gender bias in Spotify’s recommendation algorithms and where it is coming from.

These recommendation algorithms are particularly important in long-term user retention as long-term diversity of listening is associated with long-term retention. Researchers found that listeners given personalized recommendations based on their listening history had increased streams but decreased diversity compared to those given recommendations based on podcasts popular among users in their demographic group. They also stress the engagement-diversity trade-off: personalized recommendations drive immediate consumption, but long-term diversity is necessary for long-term user retention (Holtz, 2020). While this study focused on podcasts, the engagement-diversity trade-off also applies to music streamed and demonstrates the dangers both to the user and Spotify of personalized recommendation algorithms decreasing diversity.

A study of Spotify’s visual user interface as well as recommended songs attempted to understand how Spotify recommendation algorithms present gender when presenting music. They analyze the listening experiences of 80 young adults in Moscow and Stockholm using three common Spotify recommendations—Related Artists, Discover Weekly, and Browse. Related Artist recommendations are artists recommended based on a current artist the user is viewing. These recommendations were found to often have artists of the same gender and demographic group as the original. Browse algorithms are general playlists available for large communities of users and were the most diverse. The Discover playlist is a playlist of 30 songs generated weekly for each user individually and was less demographically cohesive, though still recommended music similar in genre, gender, and race. (Werner, 2020).

The study confirms that the Discover playlist is mostly unique to each user but did not address what other factors might determine the recommendations made there. Discover is made up of personalized recommendations which have been shown to be integral to the engagement-diversity trade-off and user satisfaction. As such, the diversity in these user playlists is particularly important. This particular playlist, Spotify’s “Discover Weekly”, though, has been shown to generate very little attachment between artists and fans for those who were not previously musically engaged and only relatively low attachment for those who were highly engaged (Leisewitz, 2022). Its significance and yet its lack of impact for the artists included confuse the overall role of the Discover playlist and question the accuracy of and satisfaction with its algorithmic recommendations.

Researchers at Linkoping University attempted to understand user’s gender’s impact on Spotify’s “black box” algorithm. They performed 288 bot experiments within four genres—rock, gospel, rnb/hip-hop, and dance-electronic—with each bot registering as a male or female user and listening to only the top 10 songs out of the top 100 songs of their given genre. The majority of bot users were given the same recommended artists regardless of their account’s gender, though gender-skewed recommendations were found in rock as well as rnb/hip-hop and dance/electronic. Gender classification of artists in this study was performed through text mining of artists’ information for pronouns, names, and photographs (Eriksson, 2017). This experiment concludes that account gender is not a heavily weighted factor in Spotify’s recommendation algorithm but does not further analyze these “gender-skewed” recommendations by category.

# Goals

My goal is to analyze Spotify’s recommendation algorithm for Discover Weekly—the most personalized of Spotify’s recommendation algorithms—for gender bias. Spotify’s algorithms are difficult to analyze for bias as its inputs and their weights are largely unknown, however by using a variety of accounts with a variety of identities and musical preferences, I will identify larger trends in the Discover Weekly algorithm as a whole and gain an understanding of where gender bias may be more or less apparent. In doing so, I hope to contribute to a general understanding of Spotify’s recommendation algorithm and their potential biases rooted in gender-based oppression.

Specifically, I intend to

1. Develop a method to analyze gender bias in recommendation algorithms through use of the Spotify Web API
2. Develop a gender classification model allowing for identities outside of the gender binary based on an individual’s preferred pronouns.
3. Apply these methods to the Discover Weekly recommendations of a variety of users to assess for gender bias.
4. Understand the factors, such as genre or user preference, which may moderate the effects of this bias.

# Timeline

The methodology of this project follows from the goals. First, a method to analyze gender bias in recommendation algorithms will be developed. This method will scan a user’s Discovery Weekly playlist and categorize each song by the artists’ gender identity. Originally, this will allow for four groups: individual female-identifying artist, individual male-identifying artist, individual gender non-binary artist, and groups. This categorization will be made using pronouns extracted from the voluntary biography supplied by each musical artist on Spotify. This last category, groups, will then have to be analyzed in a more complex way to understand the gender identities of each individual in the group to recategorize each of these entries to one of three final groups: female-identifying individual or group with at least one female-identifying member, exclusively male-identifying group or individual, gender non-binary individual or group with at least one gender-nonbinary individual. The first and third group have potential to overlap which will be addressed if required.

The counts of the gender identities of each artist on a user’s Discover Weekly playlist (generated by a Spotify recommendation algorithm) will allow for a comparison of proportions of songs by artists’ gender. If these proportions are statistically significantly different, then there will be compelling evidence for gender bias in Spotify’s recommendation algorithm. Each user’s proportions can then be compared to reverse engineer Spotify’s recommendation algorithm and understand the factors that contribute to difference in proportion. For example, if one user listens exclusively to country music and another listens exclusively to rap with otherwise identical demographic information and listening behavior, statistically significant differences in proportion of artists’ gender can be attributed to user’s preferred genre. Similarly, if two users share identical listening histories with different demographic inputs (such as gender identity), statistically significant differences in gender proportions can be attributed to the user’s self-selected gender.

All of this work is to be completed by April 1st with specific timeline as follows:

September – October 1st:

Gather advising committee: Lisa Dion (Computer Science) and Kate Nolfi (Philosophy)

Research Spotify Web API

Submit thesis proposal

October 1st – November 1st:

Gain thesis approval

Begin to code mechanisms for Discover Weekly gender by artist scan, including gender classification model based on pronouns in artists’ self-given biography

November 1st – December 1st:

Based on shortcomings or opportunities in scan design, design experiment and control groups (genre, user’s gender, listening patterns, etc)

December 4th – February 5th:

Perform experiment every week for 10 weeks—each Discover Weekly playlist generated by the recommendation algorithm has 30 songs, this gives 300 observations per user over 10 weeks as the recommendation algorithm updates the playlist weekly

Write Experimental Procedure for thesis paper

February 6th – March 1st:

Analyze results of proportion of artist gender identities

Attempt to find factors involved in Spotify’s black-box algorithm

Additional testing as needed

Write results for thesis paper

March 1st – April 1st:

Write abstract and background

Write discussion

Thesis Format—pre-defense final version April 1st

Thesis defense due April 15th

Revised Final Version due April 30th

# References

Bishop, S. (2018). Anxiety, panic and self-optimization: Inequalities and the YouTube algorithm. Convergence: The International Journal of Research into New Media Technologies, 24(1), 69–84. <https://doi.org/10.1177/1354856517736978>

Choi, G. Y. (2016). “Who Run the Music? Girls!” Examining the Construction of Female Digital Musicians’ Online Presence. Popular Music and Society, 40(4), 474–487. <https://doi.org/10.1080/03007766.2016.1174419>

Eriksson, M. C., & Johansson, A. (2017). Tracking gendered streams. Culture Unbound, 9(2), 163–183. <https://doi.org/10.3384/cu.2000.1525.1792163>

French, M. R. (2018). Algorithmic mirrors: An examination of how personalized recommendations can shape self-perceptions and reinforce gender stereotypes (Order No. 28115243). Available from ProQuest Dissertations & Theses A&I. (2436884514). Retrieved from <https://login.ezproxy.uvm.edu/login?url=https://www.proquest.com/dissertations-theses/algorithmic-mirrors-examination-how-personalized/docview/2436884514/se-2>

Holtz, D., Carterette, B., Chandar, P., Nazari, Z., Cramer, H., & Aral, S. (2020). The Engagement-Diversity Connection: Evidence from a Field Experiment on Spotify. SSRN Electronic Journal. <https://doi.org/10.2139/ssrn.3555927>

Hu, Y., Hu, C., Tran, T., Kasturi, T., Joseph, E., & Gillingham, M. (2021). What’s in a name? – gender classification of names with character based machine learning models. Data Mining and Knowledge Discovery, 35(4), 1537–1563. <https://doi.org/10.1007/s10618-021-00748-6>

Landström, C. (2007). Queering Feminist Technology Studies. Feminist Theory, 8(1), 7–26. <https://doi.org/10.1177/1464700107074193>

Leisewitz, A., & Musgrave, G. (2022). Does Spotify create attachment? Culture Unbound, 14(1), 75–100. <https://doi.org/10.3384/cu.3384>

Nappo, F., Letterese, E. M., Fusco, M., & Schimperna, F. (2019). The Impact of Gender Capital in the Music Industry. Academic Conferences International Limited.

Pastukhov, D. (2022, February 9). *How Spotify’s algorithm works? A Complete Guide to spotify recommendation system [2022]: Music tomorrow blog*. How Spotify’s Algorithm Works? A Complete Guide to Spotify Recommendation System [2022] | Music Tomorrow Blog. https://www.music-tomorrow.com/blog/how-spotify-recommendation-system-works-a-complete-guide-2022

Perez, C. C. (2021). *Invisible Women: Data bias in a world designed for men*. Abrams Press.

Rena, B., & Haimson, O. L. (2016). Baking Gender Into Social Media Design: How Platforms Shape Categories for Users and Advertisers. Social Media + Society, 2(4) https://doi.org/10.1177/2056305116672486

Smith, S. L., Pieper, K., Hernandez, K., & Wheeler, S. (2023). (publication). *Inclusion in the Recording Studio? : Gender & Race/Ethnicity of Artists, Songwriters & Producers across 1,100 Popular Songs from 2012 to 2022*. Retrieved September 17, 2023, from https://assets.uscannenberg.org/docs/aii-inclusion-recording-studio-jan2023.pdf.

Werner, A. (2020). Organizing music, organizing gender: algorithmic culture and Spotify recommendations. Popular Communication, 18:1, 78-90. [https://doi.org/[10.1080/15405702.2020.1715980](https://doi.org/10.1080/15405702.2020.1715980)](https://doi.org/10.1108/oir-05-2017-0153)