# Experimental Design

|  |  |
| --- | --- |
| Bias on the musician’s gender | Bias on the user’s gender |
| Sources:   * Gendered data gap * Gender disparity in the music industry * And societal ideas about gender and gender-based oppression which have created and contributed to the factors above | Sources:   * User retention-diversity tradeoff: make recommendations that keep the user engaged * Spotify’s subscription-based business model * Depending on findings here, also societal ideas about gender and gender-based oppression which have contributed to whatever these discrepancies in recommendations show |
| Evidence:   * Repeated and consistent underrepresentation of artists of specific gender groups across all findings | Evidence:   * Two users of different genders but the same in all other ways (including input data) with different recommended output |
| Design:   * Depends on definition of “algorithmic bias”  1. “algorithmic bias” as significantly worse than existing bias in the world outside of the algorithm—then must compare to current standards 2. “algorithmic bias” as any bias in the algorithm, regardless of whether it is the same as the bias in the “outside world” or not—then can simply compare to 50-50 or expected equivalent proportions  * Want to take all 30 Spotify users and analyze each week’s proportion of artists by gender | Design:   * Create five users for every possible scenario or genre being analyzed—one for each option in account creation * Analyze their recommended output for differences despite identical input data |
| Effects:   * Immediate impact on success (both monetary and otherwise) of the artists being disproportionately recommended less than others * Positive feedback loop which contributes to the gendered data gap and disparity which creates it * Less immediate impact on the user * Still some impact on the user’s sense of self in the way that recommendation impacts an individual (learned behaviors and attitudes due to lack of artists matching their gender identity) | Effects:   * Immediate impact on the user as opposed to the artist * The user’s sense of self could be altered based on recommendations given to them * In the same way that a user’s sense of self is altered based on lack of representation * Or completely different ways based on what they are presented and “supposed” to like and the ways popular culture constructs our identities |

## Research Question

Is there gender bias in Spotify’s recommendation algorithm for users’ “Discover Weekly” playlist, and if so, what factors can moderate its effects?

1. Determine if there is gender bias (of either type)
2. Given differences in user’s recommendations, determine what factors are responsible for these changes—what factors of the input data are associated with significant changes in output

## Kinds of Bias

1. Purely algorithmic bias acting on musician’s gender
2. Built-in bias based on user’s gender impacting the algorithm

Naturally, there are other kinds of bias at play here as well, which are the sources of the bias we are seeing in technology. Biases such as differences in user experience, gender disparity in the music industry, and more will be addressed in the background and assumed.

## Sources of the Bias

1. Purely algorithmic bias acting on musician’s gender
   1. Comes from the gendered data gap and gender disparity in the music industry
   2. Deeper roots in gender-based oppression
2. Systematic bias on the user’s gender impacting the algorithm
   1. Comes from a desire for monetary success and user retention in an effort to make the best recommendations for a user
   2. Comes from business model
   3. Comes from traditional gender expectations and roles
      1. Which ultimately comes from gender-based oppression and its sources as well

## Design

1. In order to detect purely algorithmic bias, we must demonstrate that Spotify is consistently recommending more artists of one gender than another, and that this difference is significantly different from the true population of musicians.
   1. This will be more challenging than analyzing the second kind of bias. Bias of the basis of user’s gender is easily identified in a simulated, designed experiment; however algorithmic bias, likely due to the systemic gendered data gap and gender disparity in the music industry, is more difficult to capture.
   2. I think the bulk of the philosophical work and data ethics comes here. The way that this study is conducted depends heavily on a definition of algorithmic bias, and if algorithmic bias that reflects societal biases is damaging or is to be expected. If harmful algorithmic bias is only bias that is significantly “worse” than existing societal bias, it begs a larger question about how to measure the existing societal bias which is out of the scope of this study.
   3. My current idea is to claim that a strong, consistent underrepresentation of artists of one gender would be evidence of algorithmic bias, and compare it to even proportion in an “ideal” world (equal not just)
   4. There are obviously a lot of issues with this approach, and lots of assumptions that it is making.
2. In order to detect systematic bias on the user’s gender, we must demonstrate that Spotify is consistently making different recommendations to users of different genders when fed the same input data.
   1. Gender options: man, woman, non-binary, something else, prefer not to say

A screenshot of a black screen

Description automatically generated

## Outcomes

1. Outcomes that would suggest algorithmic bias on artist’s gender
   1. Repeated underrepresentation of artists of gender groups across all findings
2. Outcomes that would suggest algorithmic bias on the user’s gender
   1. Two users of different genders with same input data
      1. Analyze for differences in output
      2. Every track has:
         1. Album, **artist**, available countries, duration, **explicit** (or not), content restrictions, name, track number, **popularity, acousticness, danceability, energy, instrumentalness, key, liveness, loudness, mode, speechiness, tempo, time signature, valence**
         2. Popularity : score (0-100) calculated by a Spotify algorithm involving total numbers and how recent they were
         3. Acousticness: confidence measure 0.0 to 1.0 of whether a track is acoustic or not
         4. Danceability: measure of 0.0 to 1.0 of how suitable a track is for dancing based on tempo, rhythmic stability, beat strength, and overall regularity
         5. Energy: measure from 0.0 to 1.0 of perceptual measure of intensity and activity including dynamic range, perceived loudness, timbre, onset rate, and general entropy
         6. Instrumentalness: confidence measure 0.0 to 1.0 of whether a track is purely instrumental (contains no vocals)
         7. Liveness: measure of 0.0 to 1.0 based on the presence of an audience in the recording (probability the track was performed live)
         8. Speechiness: measure of 0.0 to 1.0 of the presence of spoken word in a track
         9. Valence: measure from 0.0 to 1.0 describing musical positiveness conveyed in the track (with higher scores being happier)

## Effects

1. Purely algorithmic bias on the artist’s gender
   1. Has the most immediate impact on the artists who are disproportionately being recommended less than artists of other genders
      1. Success—monetary and otherwise
      2. Which perpetuates the same bias that feeds it
   2. Has a wider impact on society as a whole by continuing the gender disparity and gendered data gap
   3. Has a less immediate impact on the user whose sense of self could be altered or constructed based on the lack of artists matching their gender identity presented to them
      1. Overall less personal effects (?)
2. Systematic bias on the basis of the user’s gender
   1. Most immediate effect on the user whose sense of self will be altered based on what kinds of recommendations are made to them and what they feel they are “supposed” to listen to, like, etc…
   2. The differences in recommendations by user’s gender will inform some of these wider implications
      1. Such as if women are receiving more sad songs or more love songs
   3. Generally, Spotify making different recommendations based on the gender of the user impacts the user and their experience on the app and within themselves.
   4. This has the potential to have similar effects to the algorithmic bias above or effects that are completely different

* The alignment problem

All of the variables:

* Time of day
* Day of week
* Month