



Data Ethics



Search



Home



Library

The Ethics of Spotify's Machine Learning Algorithm for the Personalization of Song Recommendations

Shray Dewan





Table of Contents



01

Executive
Summary



2:00 min

02

Context &
Problem



6:00 min

03

Ethical Analysis



6:00 min

04

Guidelines



6:00 min



Data Ethics



Search



Home



Library

01



Executive Summary



1 Executive Summary



- Spotify has emerged as the **music streaming giant** due to its accurate personalized recommendation algorithm
- Two main actors: **Spotify users** and **artists**
- Amount of data collected and the way algorithm works
→ ethical problems
 - **Data privacy**
 - **Altered interactions** between users and artists



1 Executive Summary (cont'd)



- Recommendations
 - 1. Provide an **opt-out** or **opt-in** option from data tracking
 - 2. Offer more **diverse song recommendations** to users
 - 3. Remove the **monthly listeners feature** from artists' pages



Data Ethics



Search



Home



Library

02



Context & Problem

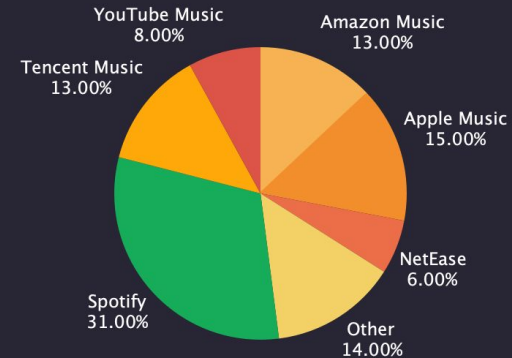


2.1 Music Streaming Industry



- Transition to streaming services
- Music streaming market share breakdown
- Spotify: **500M monthly users** - biggest platform

Global Music Streaming Market Share
2021 Q2

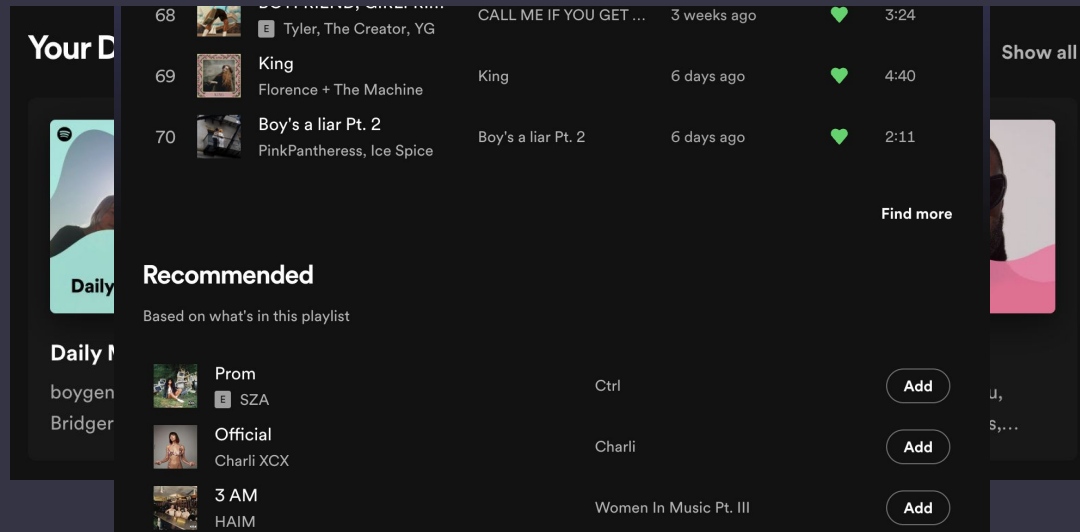




2.2 Spotify



- Known for its **personalized playlists** and **accurate song recommendations**





2.2 Spotify



- Employs artificial intelligence (**machine learning algorithm**) for its song recommendation algorithm
- Collects **1 TB** of user data per day





2.3 Spotify's Algorithm: Breakdown

- Called Bandits for Recommendations as Treatment (BaRT)
 - Multi-armed bandit algorithm
- Objective: Recommend songs as similar to the songs a user likes
- 3 types of predictor variables:
 - Collaborative Filtering (Exploitation)
 - Collaborative Filtering (Exploration)
 - Content-Based Filtering



2.3 Spotify's Algorithm: Variables



Collaborative Filtering (Exploitation)

- User's playlists
- User's skips
- User's favorite songs, artists, and genres
- User's shares
- User's listening behavior (what/when)

Collaborative Filtering (Exploration)

- Similar people's behavior
- Similar songs and artists
- Others' playlists
- Trending songs
- New songs

Content-Based Filtering

- Song's audio analysis
- Song's lyrics
- Song's online critical analysis
- Artist's label
- Song's metadata (title, release date, producers)



2.3 Spotify's Algorithm: Example



Original Playlist



Order by

#		Title	Album	Date	Duration
1		The Predatory Wasp of... Sufjan Stevens	Illinois	Jul 3, 2022	5:23
2		Are You Even Real? James Blake	Are You Even Real?	Aug 4, 2022	3:52
3		GINGER BROCKHAMPTON	GINGER	Sep 7, 2022	3:54
4		Baby Boy Kevin Abstract	ARIZONA BABY	Jul 2, 2022	3:32
5		Reach Out Sufjan Stevens, Angelo De Augustine	A Beginner's Mind	Oct 1, 2022	3:43

2.3 Spotify's Algorithm: Example



Original Playlist



Order by

(0,100)

(0,1)

(0,1)

(-60, 0)

(0,1)

(0,1)

song_name		artist	genre	popularity	danceability	energy	key	loudness	mode	...	liveness	valence	tempo
The Predatory Wasp of the Palisades Is Out to ...		Sufjan Stevens	art pop	51	0.470	0.420	4	-10.883	1	...	0.111	0.307	152.943
Are You Even Real?		James Blake	art pop	51	0.289	0.619	0	-5.970	1	...	0.397	0.197	72.139
GINGER	BROCKHAMPTON	boy band	53	0.748	0.715	2	-5.137	1	...	0.164	0.226	130.087	
Baby Boy		Kevin Abstract	afrofuturism	53	0.372	0.485	5	-8.841	1	...	0.140	0.433	170.417
Reach Out		Sufjan Stevens	art pop	47	0.571	0.311	0	-14.301	1	...	0.110	0.327	82.911
				(47, 53) MED	(0.29, 0.75) LOW-MED	(0.31, 0.71) MED	(-14.30, -5.14) LOW-MED				(0.11, 0.40) LOW	(72, 170) MED	
											(0.20, 0.43) LOW		



2.3 Spotify's Algorithm: Example



Enhanced Songs



Order by

#		Title	Album	Date	Duration
1		PDLIF Bon Iver	PDLIF	May 11, 2023	2:57
2		Fangs Matt Champion	Fangs	May 11, 2023	3:54
3		NO HALO BROCKHAMPTON	GINGER	May 11, 2023	4:20
4		iMi Bon Iver	i,i	May 11, 2023	3:16
5		Mile High (feat. Travis...) James Blake, Travis Scott, Metro Boo...	Assume Form	May 11, 2023	3:14



2.3 Spotify's Algorithm: Example



Enhanced Songs: Why



Order by

#	Title	Possible Recommendation Reason	Prev. Liked
1	 PDLIF Bon Iver	Liked songs by artist in past	
2	 Fangs Matt Champion	Matt Champion is a member of BROCKHAMPTON, like Kevin Abstract	
3	 NO HALO BROCKHAMPTON	Had a different GINGER song in original playlist	
4	 iMi Bon Iver	Liked songs by artist in past	
5	 Mile High (feat. Travis...) James Blake, Travis Scott, Metro Boo...	Had James Blake song in original playlist	



2.3 Spotify's Algorithm: Example

Enhanced Songs: Why



Order by

song_name	artist	genre	popularity	danceability	energy	key	loudness	mode	...	liveness	valence	tempo
PDLIF	Bon Iver	chamber pop	49	0.328	0.461	6	-8.790	1	...	0.1550	0.232	126.015
Fangs	Matt Champion	bedroom pop	56	0.690	0.387	11	-9.116	0	...	0.2520	0.471	83.981
iMi	Bon Iver	chamber pop	52	0.477	0.399	8	-11.544	1	...	0.2030	0.520	133.975
NO HALO	BROCKHAMPTON	boy band	59	0.643	0.573	4	-7.034	1	...	0.0695	0.470	165.192
Mile High (feat. Travis Scott & Metro Boomin)	James Blake	art pop	66	0.754	0.370	6	-9.854	0	...	0.1070	0.228	114.941



2.3 Spotify's Algorithm: Example



Enhanced Songs: Why



Order by



Original Songs Metrics



Enhanced Songs Metrics





2.4 Problems with Algorithm



1. **Privacy** of user data
 - a. Data security, third-party usage
2. **Cold-start problem** - new artists do not have user data
3. Amplification of **biases** in listening history



Data Ethics



Search



Home



Library

03



Ethical Analysis



3.1 Privacy



- Tension: providing a personalized listening experience and **individual privacy**
- Q: How does Spotify handle user listening data, including **storage**, **analysis**, and **selling**?
 - **Storage** of 1 TB/day/user
 - **Analysis** of what and when user listens to
 - **Selling** to third parties of user listening data



3.1 Privacy (cont'd)



- Extension: **Target's** analysis of sales data to pinpoint pregnant customers

“Pole applied his program to **every regular female shopper** in Target's national database and soon had a list of tens of thousands of women who were **most likely pregnant**. If they could entice those women or their husbands to visit Target and buy **baby-related products**, the company's cue-routine-reward calculators could kick in and start pushing them to buy **groceries, bathing suits, toys and clothing**, as well.”¹

¹ Charles Duhigg, “How Companies Learn Your Secrets,” *New York Times*, February 16, 2012, <https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html>.



3.1 Privacy (cont'd)



- **Example:** If you listen to the Calm Labor & Delivery playlist, Spotify may analyze your listening data and pinpoint you as someone who could be pregnant
 - Spotify could then sell this information



Public Playlist

Calm Labor & Delivery

Soft pop/rock to take the edge off.



Spotify • 65,272 likes • 79 songs, 4 hr 56 min



3.2 Classification



- Q: How does Spotify's algorithm resemble and fail to resemble a **classification system**?
- Classification system - “set of boxes (metaphorical or literal) into which things can be put to then do some kind of work;” three properties:
 - 1. “There are **consistent**, unique classificatory **principles** in operation.”
 - 2. “The categories are **mutually exclusive**.”
 - 3. “The system is **complete**.”²

² Geoffrey C. Bowker, and Susan Leigh Star, “Introduction: To Classify is Human,” in *Sorting Things Out: Classification and Its Consequences* (The MIT Press, 2000), 10-1.



3.2 Classification (cont'd)



- No fixed categories but can think of an ordinal scale of categories based on similarity to user song preferences
- Example for classification of song:
 - 7 - very similar
 - 6 - similar
 - 5 - somewhat similar
 - 4 - neither similar nor different
 - 3 - somewhat different
 - 2 - different
 - 1 - very different



3.2 Classification (cont'd)



- Does Spotify's algorithm meet the principles of a classification system?
 - 3. Complete system → Yes
 - Ability to analyze every song on Spotify's database
 - 1. Consistent principles → No
 - Cold-start problem: new artists' songs are only based on content-based filtering, not collaborative filtering
 - 2. Mutually exclusive → No
 - Some songs can have elements of similarity and dissimilarity to user's music taste



3.3 Music “Echo Chambers”



- Q: How does Spotify’s algorithm create music “echo chambers”?
- Echo chambers - “a social structure in which:
 - 1. One must subscribe to a certain belief system to be a member.
 - 2. That belief system includes the belief that all non-members are untrustworthy, and all members trustworthy.”³

³ C. Thi Nguyen, “How Twitter Gamifies Communication,” Forthcoming in *Applied Epistemology*, ed. Jennifer Lackey (OUP): 31.



3.3 Music “Echo Chambers” (cont’d)



- Algorithm only feeds users “very similar” songs (positive feedback loop)
- Easy for users to get stuck listening to the same music and thinking only their music is ‘good’ (Condition #1)
 - Fosters belief that others without same music interest have ‘bad’ taste (Condition #2)

3.3 Music “Echo Chambers” (cont’d)

- Amplification of biases in “echo chambers”
- Listening to and having recommendations for similar music → increase biases in music niche





3.4 Gamification for Artists



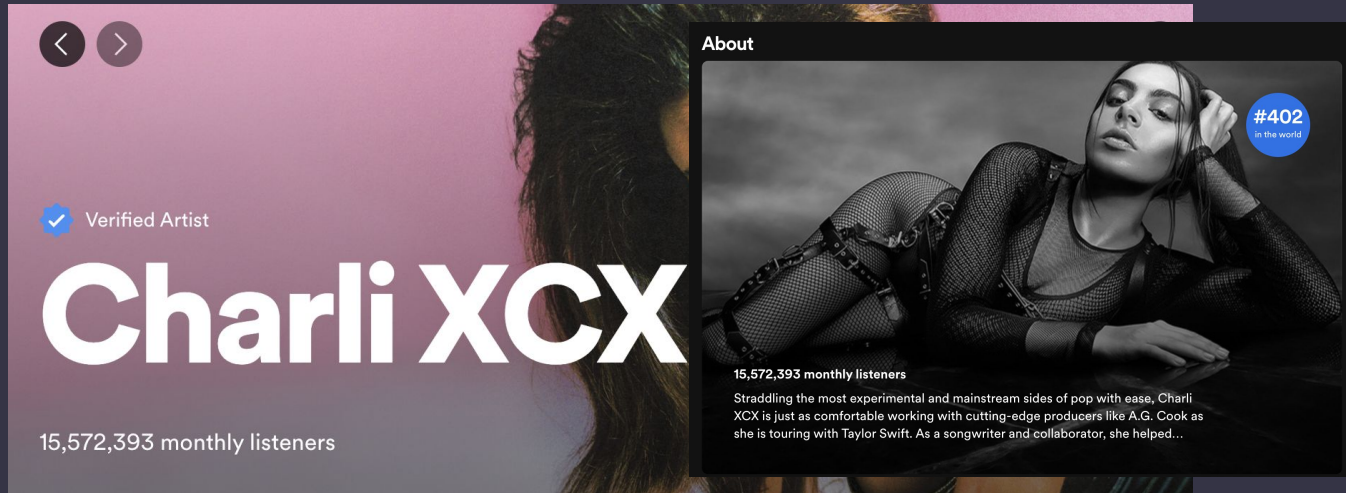
- Q: How does Spotify's algorithm **gamify popularity** for artists?
- Gamification - "increases our motivation by changing the nature of the activity"
 - "Often, the goals of ordinary activity are **rich and subtle**."
 - "When we gamify these activities, we change those goals to make them **artificially clear**."⁴

⁴ Nguyen, "How Twitter Gamifies Communication," 3.



3.4 Gamification for Artists

- **Artist popularity:** traditionally difficult-to-measure value
- Spotify gamifies artist popularity with **monthly listeners** and **artist rank**

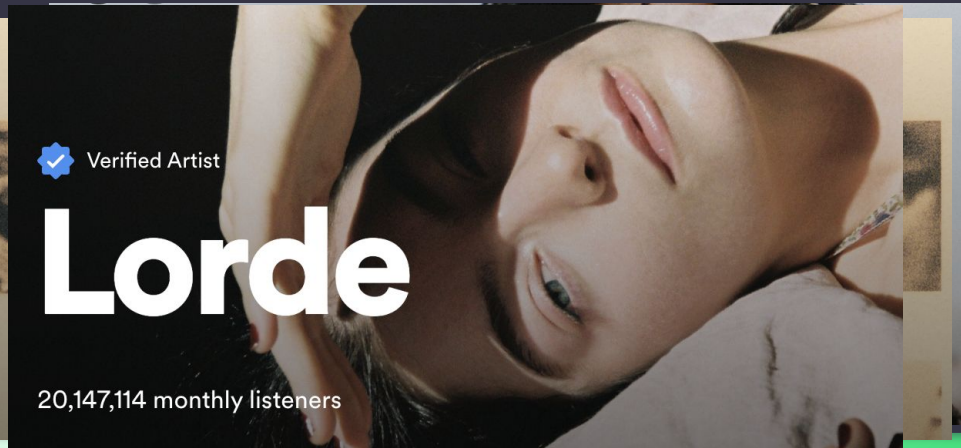




3.4 Gamification for Artists



- **Monthly listeners:** not the most accurate measure of artist popularity
 - Based only on song listens this month
- Popularity: Ice Spice > J. Cole? Frank Ocean? Lorde?





3.4 Gamification for Artists



- Has led to value capture for artists
- Value capture - “occurs when:
 - 1. Our natural values are rich, subtle, and hard-to-express.
 - 2. We are placed in a social or institutional setting which presents simplified, typically quantified, versions of our values back to ourselves.
 - 3. The simplified versions take over in our motivation and deliberation.”⁵

⁵ Nguyen, “How Twitter Gamifies Communication,” 21.



3.4 Gamification for Artists



- Artists have adopted new strategies to gain “popularity” (value capture)
 - 1. Change in focus: **albums** → **singles**
 - 2. Release singles more **frequently**
 - 3. Release music that falls into pre-established music **niches** (music “echo chambers”)

The End of an Era: The Death of the Album and Its Unintended Effects



3.5 Transparency over Accountability

- Q: Does Spotify prioritize transparency over accountability?
- **Transparency** of algorithm through releasing and discussing metrics
- Limit of transparency: “**disconnected from power**”
 - “If transparency has no meaningful effects, then the idea of transparency can lose its purpose.”⁵
- Unethical practices continue despite ethical concerns
- **Accountability** necessary

⁵ Mike Ananny and Crawford, Kate, “Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability,” *new media & society* 20, no. 3 (2018): 978.



3.6 Ethical Questions Summary



1. How does Spotify handle user listening data, including storage, analysis, and selling?
2. How does Spotify's algorithm resemble and fail to resemble a classification system?
3. How does Spotify's algorithm create music "echo chambers"?
4. How does Spotify's algorithm gamify popularity for artists?
5. Does Spotify prioritize transparency over accountability?



Data Ethics



Search



Home



Library

04



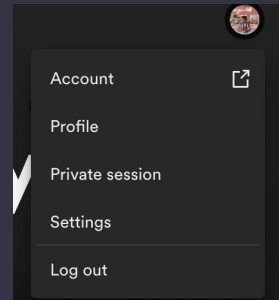
Guidelines



4.1 Privacy



- **Objective:** improve data privacy for Spotify users
- Spotify should provide an **opt-out**, or better yet, an **opt-in** from data tracking
 - Tradeoff: reduced personalization
 - Would benefit users who do not want their data potentially stored, analyzed, and sold

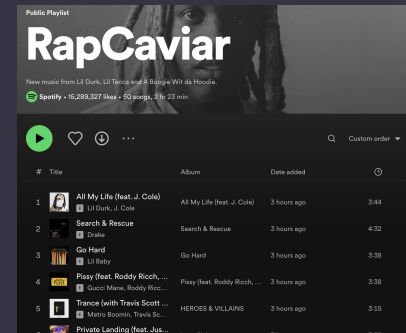




4.2 Music “Echo Chambers”



- **Objective:** reduce frequency of music “echo chambers”
- Spotify should provide more **unexpected, diverse song suggestions** that are only somewhat related to user’s music preferences
 - Consider employing if user reaches a music “echo chamber” threshold





4.3 Gamification for Artists



- **Objective:** reduce level of gamification for artists
- Spotify should **hide the monthly listeners features**
- Spotify should only have its artist rank feature, if anything
 - Incorporate album success in rank





Data Ethics



Search



Home



Library

05



References



5.1 References



Ananny, Mike, and Crawford, Kate. "Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability." *new media & society* 20, no. 3 (2018): 973-89.

Bowker, Geoffrey C., and Susan Leigh Star. "Introduction: To Classify is Human." In *Sorting Things Out: Classification and Its Consequences*, 1-28. The MIT Press, 2000.

Duhigg, Charles. "How Companies Learn Your Secrets." *New York Times*, February 16, 2012.
<https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html>.

krn6. "The End of an Era: The Death of the Album and Its Unintended Effects." *Gnovis Journal*, February 28, 2014.

Marius, Hucker. "Uncovering How the Spotify Algorithm Works." *Towards Data Science*, November 23, 2021.

Michell, Tay. "Spotify's Algorithm: Helping Or Hurting Musicians?" *Arts Management & Technology Laboratory*, April 28, 2022.

Nguyen, C. Thi. "How Twitter Gamifies Communication." Forthcoming in *Applied Epistemology*, ed. Jennifer Lackey (OUP): 1-40.

Platinga, Bo. "What do Spotify's audio features tell us about this year's Eurovision Song Contest?" *Medium*, April 28, 2018.

Spotify for Developers. "Get Track's Audio Analysis." Accessed May 11, 2023.
<https://developer.spotify.com/documentation/web-api/reference/get-audio-analysis>.

Tirumala, Sameer. "APIs Zero to Hero: ft. Spotify." *Medium*, September 27, 2022.

Wall Street Journal, "Spotify Breaks Down the Mapping Tech Behind Its Algorithm." April 17, 2023.