



Assessing and Addressing Algorithmic Bias - But Before We Get There...

Jean Garcia-Gathright
Aaron Springer
Henriette Cramer

2nd FATREC Workshop:
Responsible Recommendation
October 6, 2018
Vancouver, BC

Algorithmic fairness was anticipated at least 20 years ago!

Friedman & Nissenbaum, '96

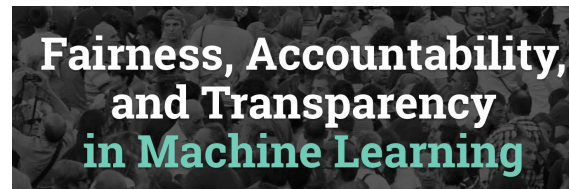
“[W]e use the term bias to refer to computer systems that systematically and unfairly discriminate against certain individuals or groups of individuals in favor of others.”

Friedman, B. and Nissenbaum, H., 1996. Bias in computer systems. *ACM Transactions on Information Systems (TOIS)*, 14(3), pp.330-347.

```
123DECBUEBRC<<
 23DEC  SUN  BUE/2-3      BRC/¥0
1AR  1684  Z7 Y7 Q7 V7 W7*ABPBRC  115P  335P 73W S 0 XF DCA /E
      B7 N7 H7 K7 M0 L0 00 S7 G4 *A
2LA/** 4348 Y9 B9 H9 K9 M9 ABPBRC  110P  335P 320 S 0 S DCA /E
      L9 V9 S9 N0 Q0 00 X0 *A
3AR/AU 2696 Z2 Y7 Q7 V7 W7*ABPBRC  230P  450P E90 S 0 JS DCA /E
      B7 N7 H7 K7 M7 L7 05 S7 G7 *A
4LA/** 4344 Y9 B9 H9 K9 M9 ABPBRC  1100A 125P 320 S 0 JS DCA /E
      L9 V9 S9 N0 Q0 00 X0 *A
5AR  1682  Z7 Y7 Q7 V7 W7*ABPBRC  1005A 1225P 73W S 0 DCA /E
      B7 N7 H7 K7 M0 L0 00 S7 G7 *A
6LA/** 4350 Y9 B9 H9 K9 M9 ABPBRC  345P  610P 320 S 0 DCA /E
      L9 V9 S9 N0 Q0 00 X0 *A
* - FOR ADDITIONAL CLASSES ENTER 1*C
* - FOR AIR EXTRAS INCLUDING PAID SEATS ENTER 1*A.
126DECBUEBRC~
 26DEC  WED  BUE/2-3      BRC/¥0
1AR  1684  Z7 Y7 Q7 V7*ABPBRC  115P  335P 73W S 0 XF DCA /E
      W7 B7 N7 H7 K7 M7 L7 00 S7 G7 *A
2LA/** 4346 Y9 B9 H9 K9 ABPBRC  1220P  245P 320 S 0 XJS DCA /E
      M9 L9 V9 S9 N9 Q9 09 X0 *A
3LA/** 4342 Y9 B9 H9 K9 ABPBRC  1020A 1245P 320 S 0 W DCA /E
      M0 L0 V0 S0 N0 Q0 00 X0 *A
4AR  1682  Z7 Y7 Q7 V7*ABPBRC  1005A 1225P 73W S 0 DCA /E
      W7 B7 N7 H7 K7 M0 L0 00 S7 G0 *A
5LA/** 4350 Y9 B9 H9 K9 ABPBRC  345P  610P 320 S 0 DCA /E
      M9 L9 V9 S9 N9 Q9 09 X0 *A
6AR  1688  Z7 Y7 Q7 V7*ABPBRC  535P  755P 73W S 0 DCA /E
      W7 B7 N7 H7 K7 M7 L7 00 S7 G7 *A
* - FOR ADDITIONAL CLASSES ENTER 1*C
* - FOR AIR EXTRAS INCLUDING PAID SEATS ENTER 1*A.
```

SABRE, the original impetus for considering computational bias

Research communities are growing and becoming more visible



How do we translate research into practice?



Outline

- Establishing a common language and framework
- Addressing domain-specific challenges: case study in Voice
- Organizational work: pragmatic challenges

Establishing a common language and framework



Three entry points for bias

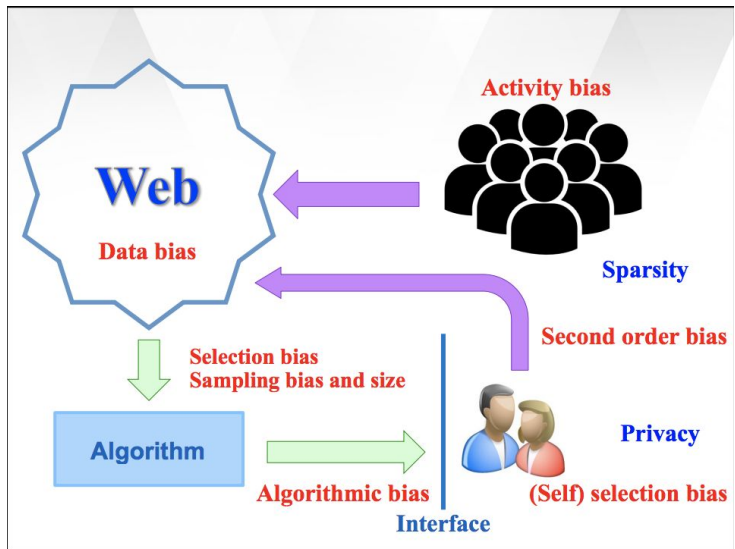




Biases in social data (Olteanu et al.)

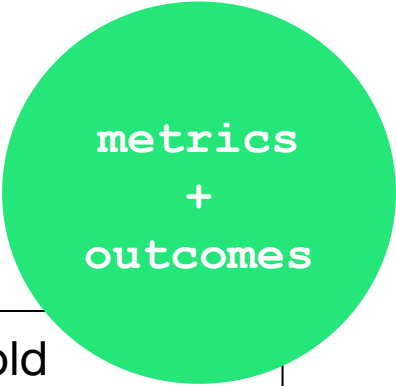
- Population bias
- Behavioral bias
- Content production bias
- Linking bias
- Temporal variations
- Redundancy
- Functional biases
- Data collection biases
- Data processing biases
- Biases in evaluation and interpretation of findings

Second order effects: Data -> algorithm -> more data (Baeza-Yates)



Baeza-Yates, R. Bias on the Web.
Communications of the ACM, June 2018, Vol. 61 No. 6, Pages 54-61.
<https://cacm.acm.org/magazines/2018/6/228035-bias-on-the-web/fulltext>

Different types of outcome biases



metrics
+
outcomes

“[W]e use the term bias to refer to computer systems that systematically and unfairly discriminate against certain individuals or groups of individuals in favor of others.”

Friedman & Nissenbaum (1996)

Friedman, B. and Nissenbaum, H., 1996. Bias in computer systems. *ACM Transactions on Information Systems (TOIS)*, 14(3), pp.330-347.

Harms of allocation withhold opportunity or resources from certain groups

Harms of representation reinforce subordination along the lines of identity / stereotypes

Kate Crawford, “The Trouble With Bias” keynote at NIPS 2017

https://www.youtube.com/watch?v=fMym_BKWQzk

Helping teams think concretely about bias in their products is the first step.



- Type of bias
- Outcome of bias
- Risks and benefits
- Next steps

Addressing domain-specific challenges: case study in Voice

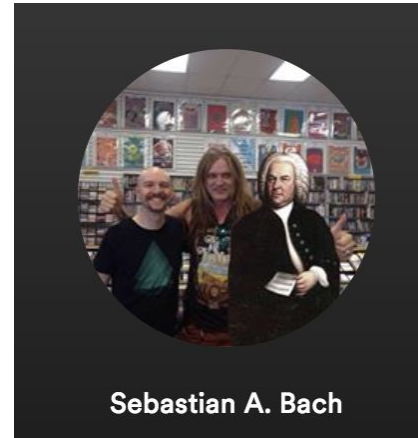
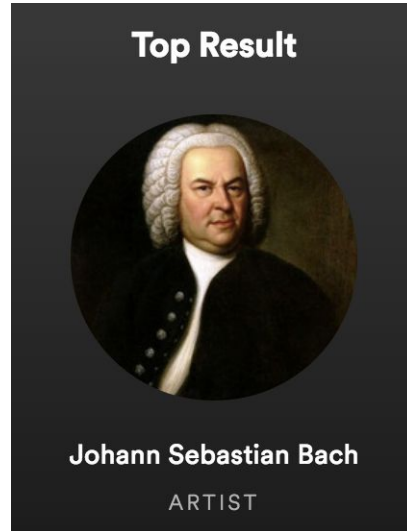
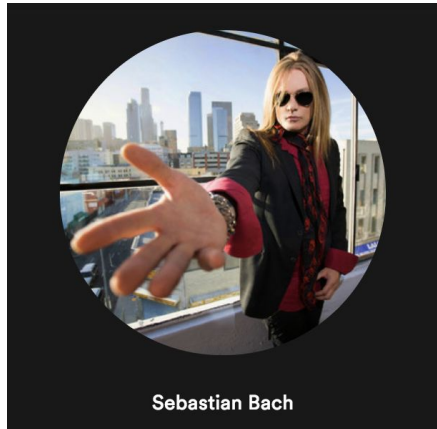


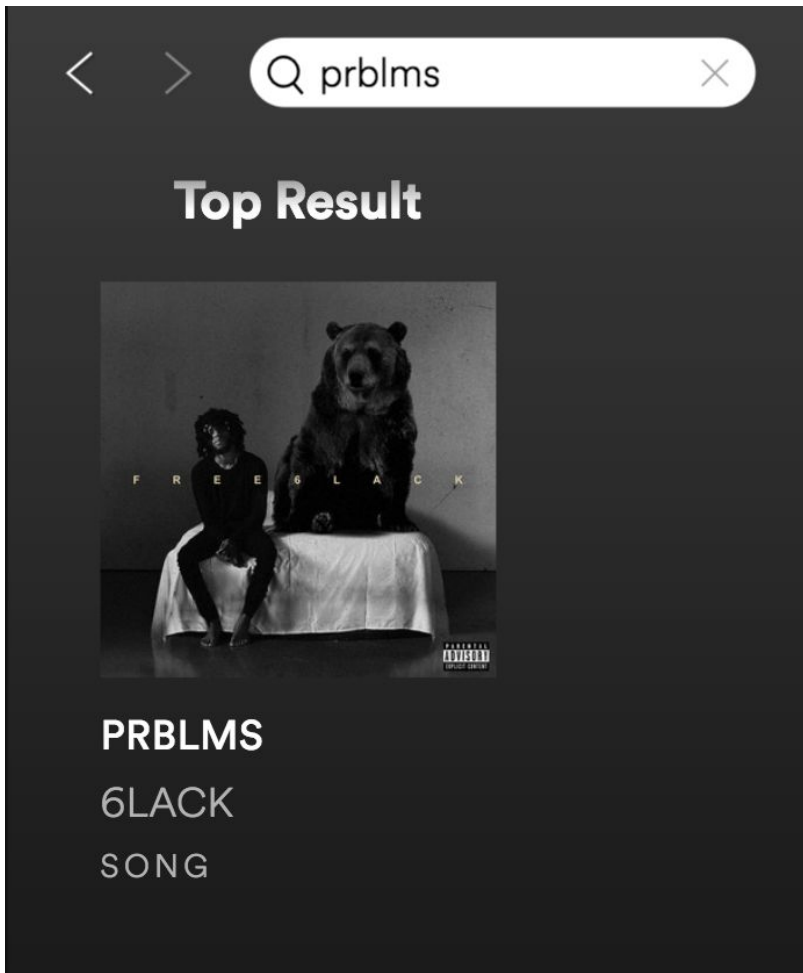
Voice Experiences

“Play my
Discover Weekly”



Voice amplifies what's on top





What becomes inaccessible when using voice?

“Play PRBLMS”

“Play Problems by Black”

“Play P.R.B.L.M.S. by six lack”



bear ✓

@6LACK

Follow



if y'all catch anybody pronouncing it six lack
or 6-black, enlighten em 🏠

5:48 PM - 21 May 2016

English dialects

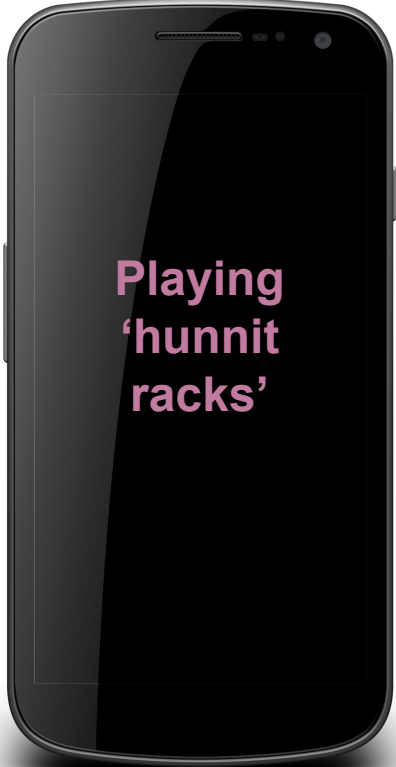
Play you da
baddest



Playing
'YouTube
Baddest'

Expressive spelling

Play 100it racks



Playing
'hunnt
racks'

Non-english & code switching

Play Dile Que Tu
Me Quieres



Playing
'Delicate
Tony
Curious'

Organizational work: pragmatic challenges

What constitutes a 'fair' outcome is not obvious.



Different stakeholders can have different perspectives on 'fairness'

Decision-maker: of those I've labeled high-risk, how many will recidivate?

Predictive value

Defendant: what's the probability I'll be incorrectly classified high-risk?

False positive rate

Society [think hiring rather than criminal justice]: is the selected set demographically balanced?

Demography

Did not recidivate	TN	<u>FP</u>
Recidivated	<u>FN</u>	TP
	Labeled low-risk	Labeled high-risk



Arvind Narayanan: 21 fairness definitions and their politics

<https://www.youtube.com/channel/UCO19zyFNtkbcTQwERVVZB0Q>

Challenges in assessing 'fairness'.

Some content gaps & biases are **intentional**:

- New music playlists: recency bias

Some content gaps & biases can be argued to be **unfair**:

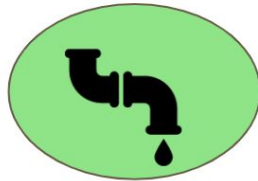
- Under-index of certain genres over others

Teams need to be aligned on priorities and interdependencies.

Priorities and interdependencies



Mitigation efforts must compete with other prioritized deliverables



Established products may have upstream and downstream dependencies



It's most effective to appeal to the desire to make a better product



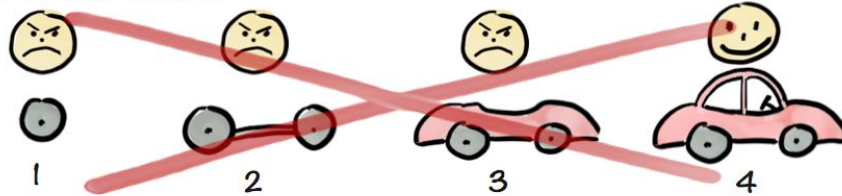
Must get organizational support through education and evangelism

Mitigation efforts should fit into ways-of-working in product.

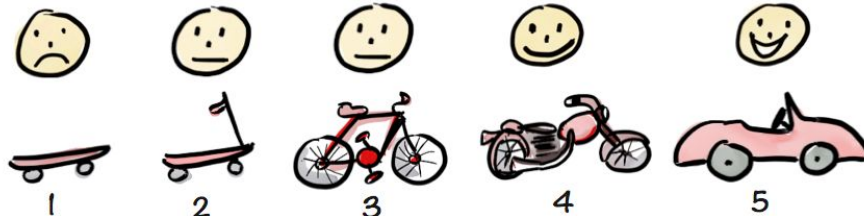
Minimum viable products and continuous improvement

- What would an MVP for algorithmic bias assessment and mitigation look like?

Not like this....



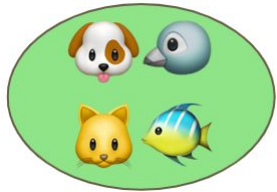
Like this!



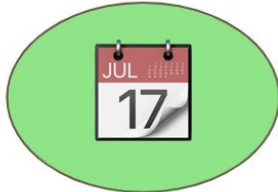
Address technical debt through cultural changes.



Address technical debt through cultural changes



More inclusive product development through diverse teams



Integrate algorithmic bias assessment into everyday workflow



Summary

- Leverage existing literature to establish a common language for product teams about bias
- Address domain-specific challenges
- Define desired outcomes; align strategy with ways-of-working

Any dataset or algorithmic outcome is 'biased'

*has characteristics we can influence

jean@spotify.com

@scinoise

