AI, why you ain't fair? : Understanding AI Bias

Jaydeep Borkar

Savitribai Phule Pune University

Hey! What is this poster about?

It's on AI Bias…

Oh great! But what's that?

AI systems have been reported for their unfair and biased decisions against a specific community/group of people on the basis of a lot of factors. They also haven't been robust and reliable.

Okay. What factors?

Gender, Race, Color, Geographical location, tone of voice are a few of them

I see. That's interesting. But how do they get biased?

Biased training data, incomplete data, stereotypical beliefs of person handling the data, interactions. Yellow panel and green panel for more

AI systems are only as good as their training data. So they get biased due to the biased datasets

Hmm, how do datasets get biased then?

As real-world datasets are handled by humans, sometimes bias enters the dataset unknowingly. See the yellow panel for more

Okay, but does this happen on purpose?

- Bias can also enter due to certain stereotypical beliefs while handling/making the dataset.
- Sometimes bias enters in spite of the best of intentions.

POTENTIAL HARMS FROM ALGORITHMIC DESCISION-MAKING

INDIVIDUAL HARMS		COLLECTIVE SOCIAL HARMS
ILLEGAL DISCRIMINATION	UNFAIR PRACTICES	
HIRING		LOSS OF OPPORTUNITY
EMPLOYMENT		
INSURANCE & SOCIAL BENEFITS		
HOUSING		
EDUCATION		
CREDIT		ECONOMIC LOSS
DIFFERENTIAL PRICES OF GOODS		
LOSS OF LIBERTY		SOCIAL STIGMATIZATION
INCREASED SURVELLIANCE		
STEREOTYPE REINFORCEMENT		
DIGNATORY HARMS		

Chart Contents Courtesy of Megan Smith, Former CTO of the United States

Joy Buolamwini, MIT Media Lab

Bias in Datasets

what we have

what we want to recognize

(Timnit Gebru, MLSS 2019)

Type of Bias

1. **ALGORITHMIC BIAS:** feeding biased data to the model. Ex gender biased data.
2. **DATA LIMITATIONS:** Sample bias? **3. FROM INTERACTIONS**
4. **CONFIRMATION BIAS:** presenting only the information that aligns with beliefs
5. **EMERGENT BIAS:** similar to interactive bias, happens via interaction. Also due to changing societal knowledge, population, cultural values, new set of users.

(The Ethical Implications of AI)

LIMITATIONS

Lidarradar.com

Triton

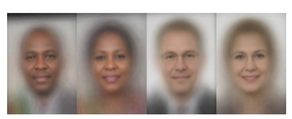
Examples?!

Yeah sure. Please see the red panel

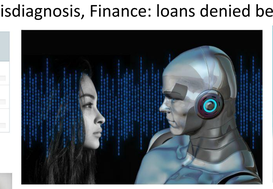
Examples

- Facial recognition software made no mistakes when identifying the gender of men, but mistook women (of color) for men.
- AI bias in hiring/recruitment, Bias in NLU/NLP.
- Healthcare: can lead to misdiagnosis, Finance: loans denied because of race

Gender Classifier	Darker Male	Darker Female	Lighter Male	Lighter Female	Largest Gap
Microsoft	94.0%	79.2%	100%	98.3%	20.8%
FACE++	99.3%	65.5%	99.2%	94.0%	33.8%
IBM	88.0%	65.3%	99.7%	92.9%	34.4%



Joy Buolamwini, MIT Media Lab



Recruiting Times

Forbes

Raindrop74/ Shutterstock

Laura Dauglas

But still, should we really care about this?!

YES! AI has been getting intelligent enough to solve humanity's most critical problems, it's important for it to be fair and reliable.

Yeah. So can we do anything about it?

Yes :) Please see blue panel for more

Key takeaways?

Achieving accuracy in AI systems is important, but it's equally important for the AI to be fair, reliable, and unbiased.

Great! How can we reach out to you for further discussion or stay in touch?

Twitter @JaydeepBorkar

jaijborkar@gmail.com

jaydeepborkar.github.io

Where can I learn more about this topic?

Visit this link: https://cutt.ly/ai-bias
Or scan this QR code:

REFERENCES

- The Ethical Implications of AI
- Joy Buolamwini, Timnit Gebru. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. Conference on Fairness, Accountability, and Transparency, 2018.
- Bellamy et al., AI FAIRNESS 360: An Extensible Toolkit for Detecting, Understanding, and Mitigating Unwanted Algorithmic Bias. 2018.
- Calmon et al., Optimized Pre-Processing for Discrimination Prevention. NIPS, 2017.
- Gebru et al., Datasheets for Datasets. 2019

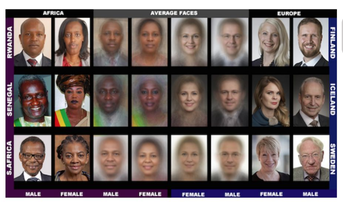
Solutions

Ilana Pelzman-Kern, MIT Media Lab

Making datasets as diverse as possible.

Pilot Parliaments Benchmark (PPB) Dataset

So how do we make the dataset diverse?



Pilot Parliaments Benchmark

Involvement of people from different background, not just similar categories

Joy Buolamwini, MIT Media Lab

Also, there are more ways to mitigate AI Bias…

Various algorithms: IBM's Optimized Pre-Processing, Post-Processing and In-Processing algorithms, etc…

Calmon et al., NIPS 2017

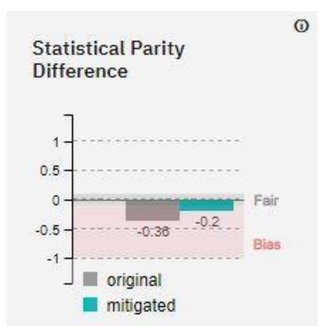
(Gebru et al., 2019)

Datasheets for Datasets

Bellamy et al., 2018

Fairness metrices, toolkits, testing techniques to check fairness

Mitigating Bias in Natural Language Processing



AIF 360 by IBM

Various organizations should step up to conduct high-quality research on ethical implications and bias in AI

AI Ethics education in curriculum, laws to restrict use of unfair AI