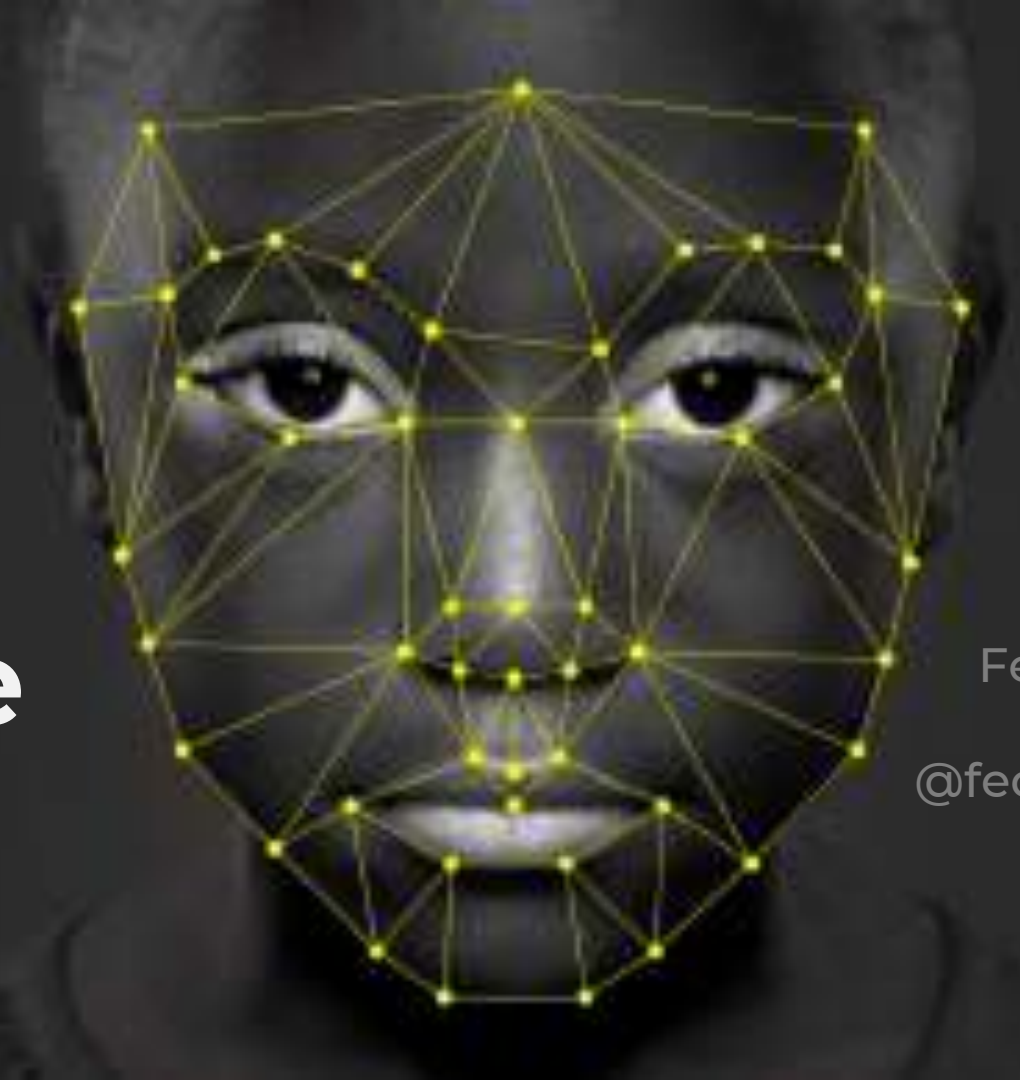
Avoiding bias in predictive analytics.



Federica Pelzel
@federicca

“There’s software used across the country to predict future criminals. And it’s biased against blacks.”

https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Why this Matters Now.

“Film stock's failures to capture dark skin aren't a technical issue, they're a choice”

Film "could have been designed initially with more sensitivity to the continuum of yellow, brown and reddish skin tones but the design process would have to be motivated by a recognition of the need for extended range." Back then there was little motivation to acknowledge, let alone cater to a market beyond white consumers.

https://www.buzzfeednews.com/article/syreetamcfadden/teaching-the-camera-to-see-my-skin#.iqNB3KwAva

A 2018 report from *Wired* shows, nearly three years on and Google hasn't really fixed anything. The company has simply blocked its image recognition algorithms from identifying gorillas altogether

https://www.theverge.com/2018/1/12/16882408/google-racist-gorillas-photo-recognition-algorithm-ai

Facial recognition algorithms made by Microsoft, IBM and Face++ were more likely to misidentify the gender of black women than white men.

When the person in the photo is a white man, the software is right **99% of the time.**

Gender was misidentified in up to 12 percent of darker-skinned males in a set of 318 photos.

Gender was misidentified in 35 percent of darker-skinned females in a set of 271 photos.

https://www.aclu.org/blog/privacy-technology/surveillance-technologies/amazons-face-recognition-falsely-matched-28

Amazon’s recruiting system taught itself that male candidates were preferable. It penalized resumes that included the word “women’s,” as in “women’s chess club captain.” And it downgraded graduates of two all-women’s colleges.

The technology also favored candidates who described themselves using verbs more commonly found on male engineers’ resumes, such as “executed” and “captured,”.

By **2021**, research firm IDC predicts, organizations will spend

$52.2 billion annually

on A.I.-related products—and economists and analysts believe they'll realize many billions more in savings and gains from that investment.

PwC estimates that A.I. could contribute up to

$15.7 trillion in 2030

to the global economy, more than the combined output of China and India today.

How did We get Here?

The simple answer:

If DATA isn’t fair or just,

OUTPUTS won’t be fair or just.

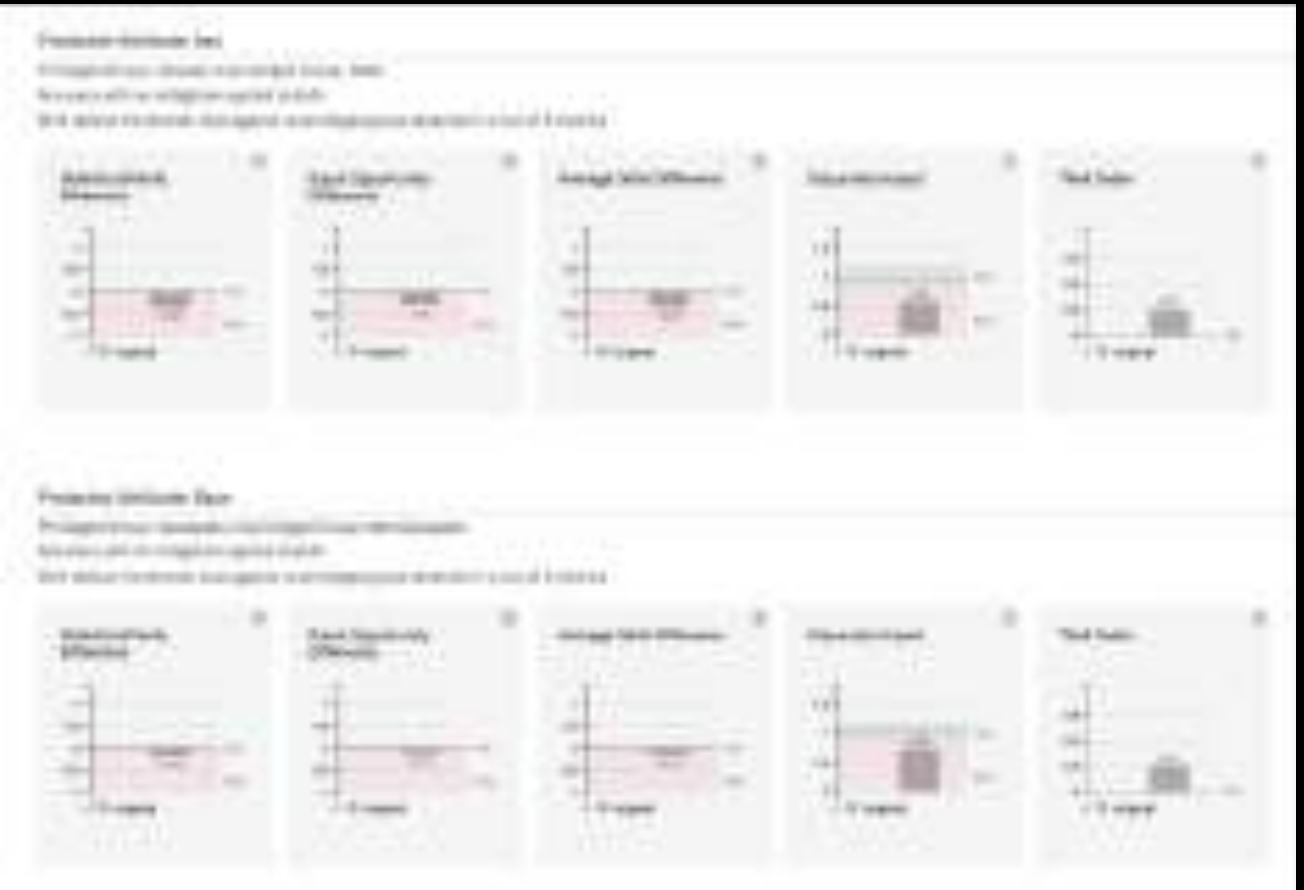

WHEN YOU DESIGN AN ALGORITHM TO BRING BALANCE TO THE FORCE
BUT FEED IT DATA CONTAMINATED WITH BIAS
YOU HAVE BECOME THE VERY THING
YOU SWORE TO DESTROY

What's Being done Today?

Initiatives launched in 2018

05/01/18	facebook	Facebook says it has a tool to detect bias in its artificial intelligence	Quartz
05/25/18	Microsoft	Microsoft is creating an oracle for catching biased AI algorithms	MIT Technology Review
05/31/18	pymetrics	Pymetrics open-sources Audit AI, an algorithm bias detection tool	VentureBeat
06/07/18	Google	Google Education Guide to Responsible AI Practices – Fairness	Google
06/08/18	accenture	Accenture wants to beat unfair AI with a professional toolkit	TechCrunch

Automated algorithm audits



The AI Fairness 360 Python package includes a comprehensive set of metrics for datasets and models to test for biases, explanations for these metrics, and algorithms to mitigate bias in datasets and models.

It's fully **open source**, and can be found in

http://aif360.mybluemix.net/

“Fairness is a multifaceted, context-dependent social construct that defies simple definition.”

AI Fairness 360 considers **over 70 fairness metrics**, and still barely scratches the surface

Disrupt traditionally biased systems

Uses neuroscience and more than 150 pre-identified biases, combined with high performance indicators, to create custom models that provide bias free and efficient recruitment.

Re-thinks credit score in the US as the sole measure of credit worthiness, and has seen that they can approve more than double the people for loans, while maintaining similar or equal loss rates as traditional banks.

Identify bias and take action

Organizations like **AlgorithmWatch** and **The Algorithmic Justice League** founded by Joy Buolamwini (watch her amazing TED talk!) are striving to help evaluate and identify bias in existing algorithms by providing education and training materials, as well as a collaborative and inclusive space for people to report bias in algorithms, and help solve these issues as a community.

Policy and Regulation

GDPR

- Forbids Profiling
- Promotes right to an explanation
- Against Bias and Discrimination

So...
How do I Prevent this?

1. Diversity

2. Awareness

3. Regulation

4. Transparency

5. Testing and QA

Thank you!

Federica Pelzel

@federicca
fedepelzel@gmail.com