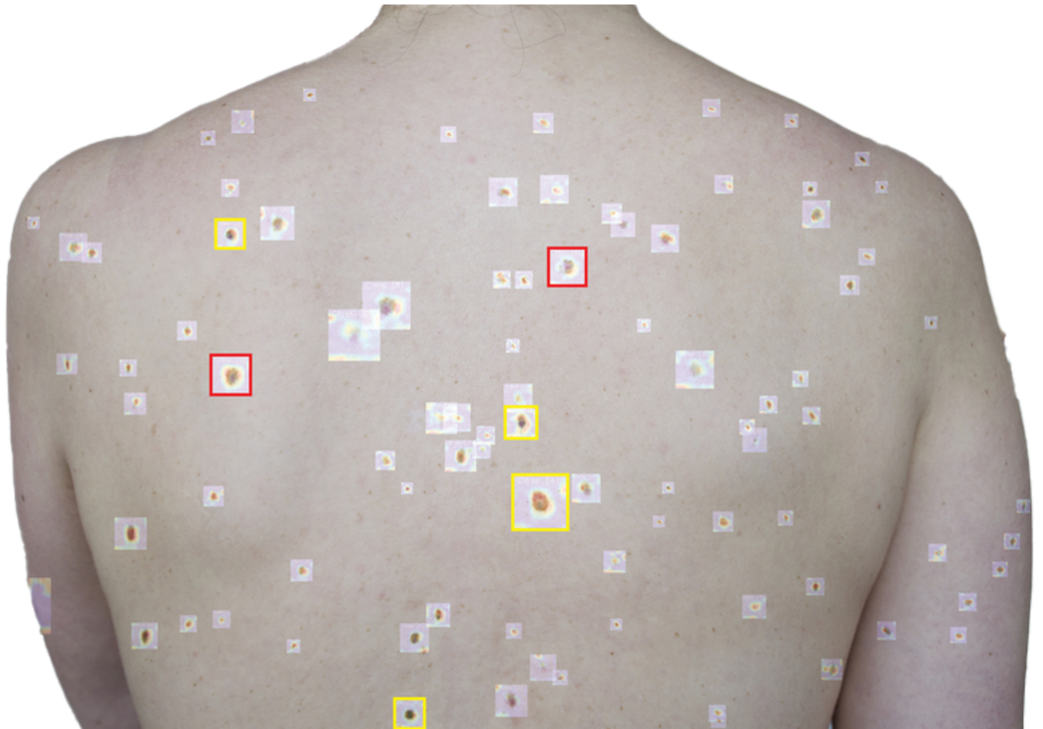
Fairness case

Fact: AI image analyzing for cancer performs better on whiter skin
*Contrast and visibility difference / Training data difference

Paradigmatic AI healthcare case, AI better for some groups than others

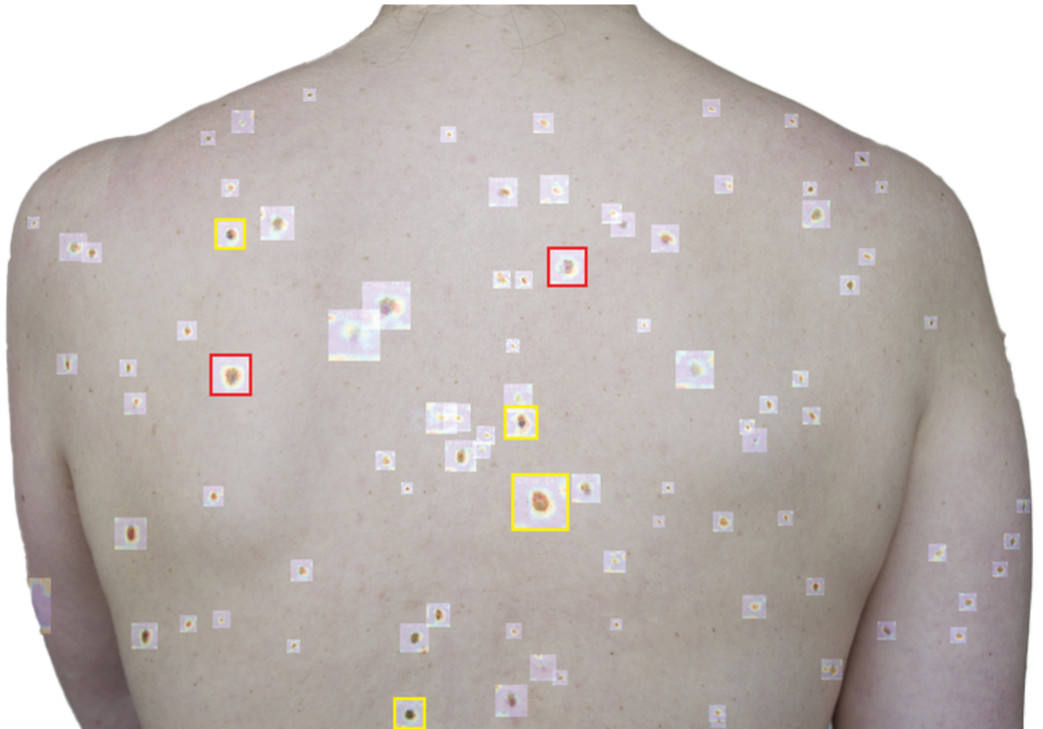
Is this a problem? If so, is there a solution?

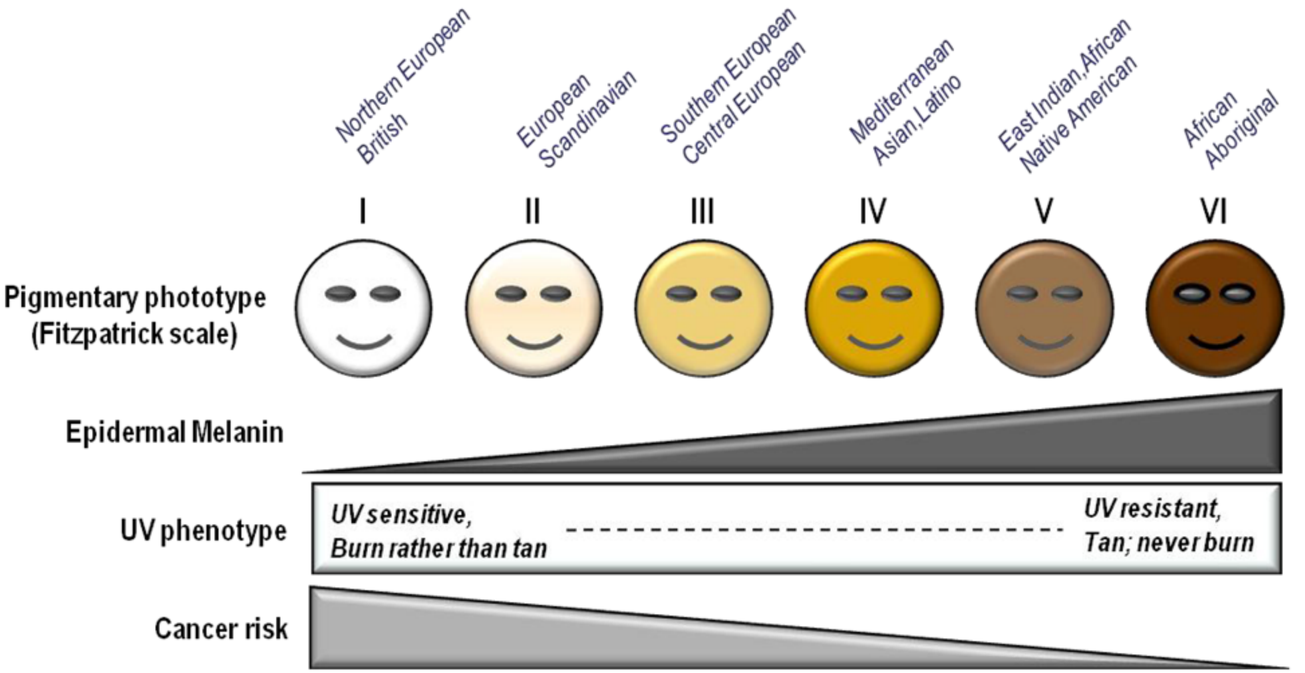


Fairness case

Fact: AI image analyzing for cancer performs better on whiter skin
*Contrast and visibility difference / Training data difference

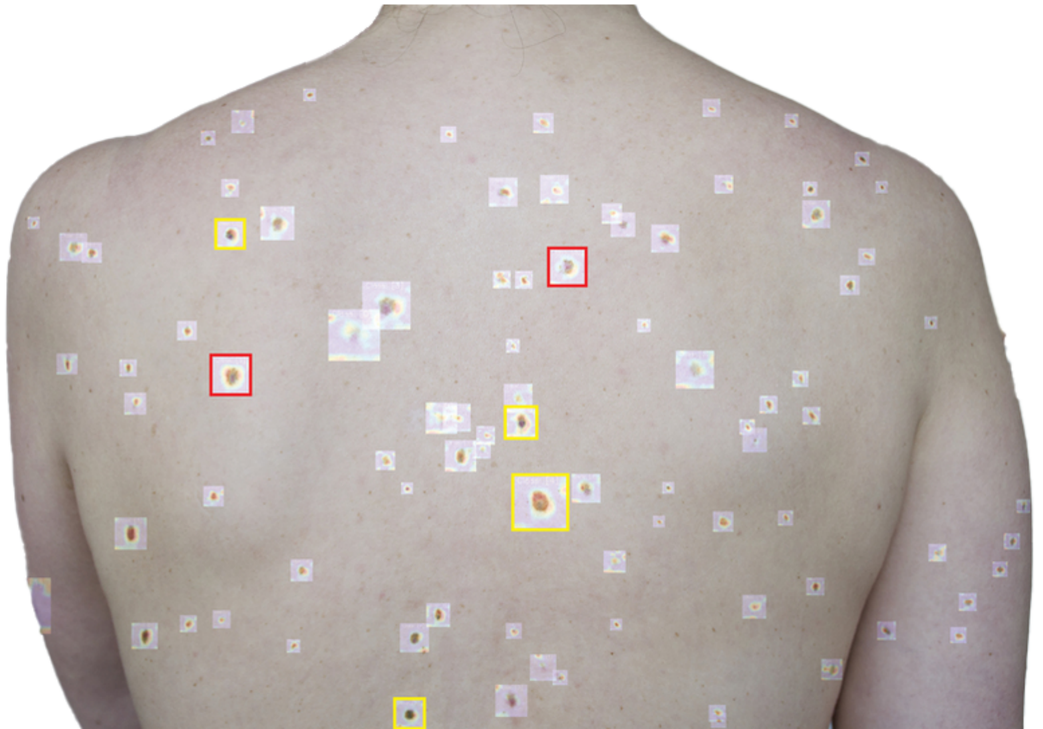
Complication: Cancer risk rises with whiter skin

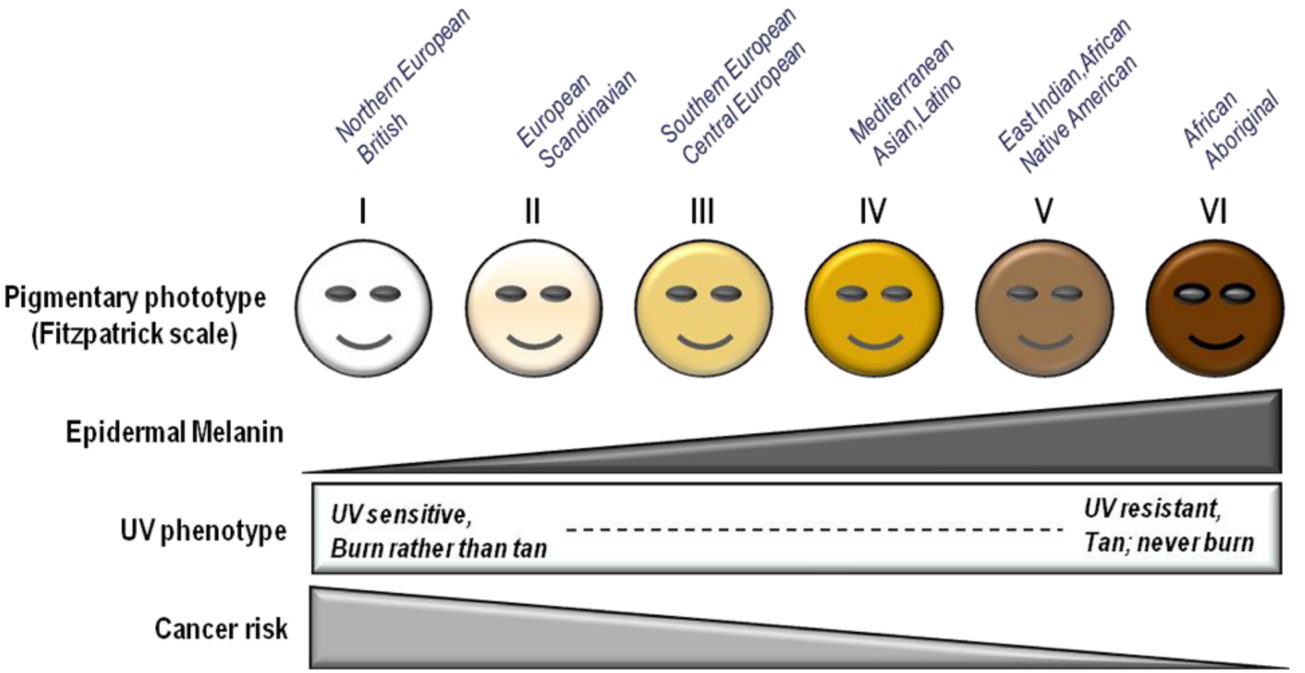




Fairness case

Question: Should disproportionate resources be dedicated to the spectrum's darker end to ensure equal AI accuracy across skin tones?
*Which resources? How much?





Fairness case



Stop *all* image analyzing until equal AI accuracy across skin tones

Yes? No? Is this fair? Is this possible?





Fairness Experiments

What is the *value* of fairness?

How can it be described, or measured?

Definition fairness

1. ChatGPT

2. Definitions
   - Philosophical
   - In the literature…
   - Within domains
   - Historical



"Please depict the concept of fairness in AI," ChatGPT 2023

Definition fairness

Philosophical

Treat equals equally,
and unequals proportionately unequally,
within the decision domain

- Aristotle, *Nicomachean Ethics*, Book V, Chpt 3

Basketball players
Lawyers
Managers...?

Blaker et al. "The height leadership advantage in men and women: Testing evolutionary psychology predictions about the perceptions of tall leaders."



Case: Fairness academic literature

Fairness?

ple, algorithms are vulnerable to biases that render their decisions "unfair" [6, 117]. In the context of decision-making, fairness is the *absence of any prejudice or favoritism toward an individual or group based on their inherent or acquired characteristics.* Thus, an unfair algorithm is one whose decisions are skewed toward a particular group of people. A canonical example comes from a tool used by courts in the United States to make pretrial detention and release decisions. The software,

ACM Reference format:
Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. 2021. A Survey on Bias and Fairness in Machine Learning. *ACM Comput. Surv.* 54, 6, Article 115 (July 2021), 35 pages.
https://doi.org/10.1145/3457607

Case: Fairness academic literature

Inequality - not equality - as engine of fairness

ple, algorithms are vulnerable to biases that render their decisions "unfair" [6, 117]. In the context of decision-making, fairness is the *absence of any prejudice or favoritism toward an individual or group based on their inherent or acquired characteristics.* Thus, an unfair algorithm is one whose decisions are skewed toward a particular group of people. A canonical example comes from a tool used by courts in the United States to make pretrial detention and release decisions. The software,

ACM Reference format:
Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. 2021. A Survey on Bias and Fairness in Machine Learning. *ACM Comput. Surv.* 54, 6, Article 115 (July 2021), 35 pages.
https://doi.org/10.1145/3457607

Case: Fairness academic literature, a plague of error

Google Scholar

mehrabi "a survey on bias and fairness in machine learning"

Articoli

In qualsiasi momento
Dal 2024
Dal 2023
Dal 2020
Intervallo specifico...

Ordina per pertinenza
Ordina per data

Qualsiasi lingua
Pagine in Italiano

A survey on bias and fairness in machine learning

N Mehrabi, F Morstatter, N Saxena, K Lerman, A Galstyan

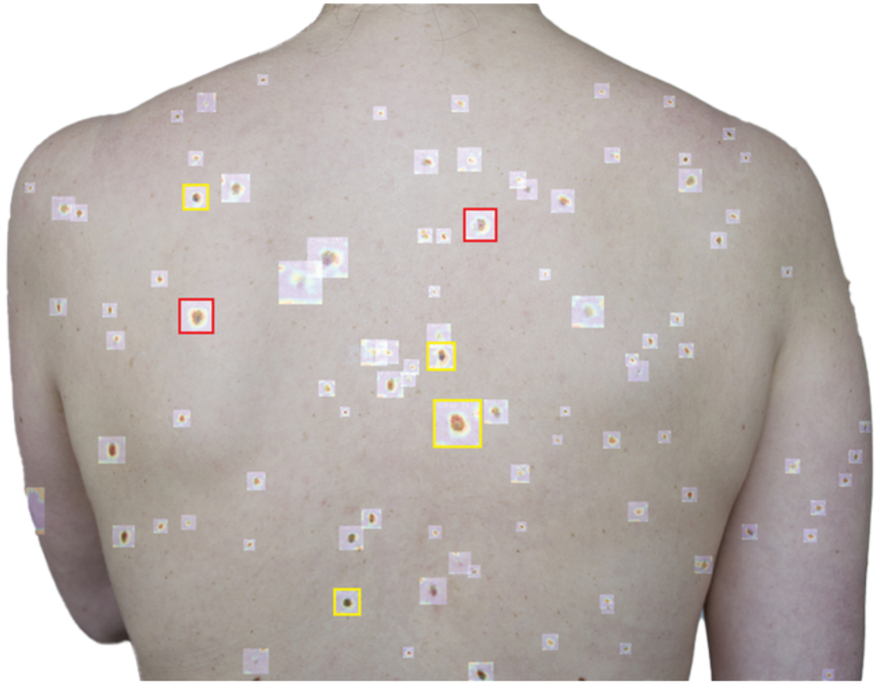
ACM computing surveys (CSUR), 2021 • dl.acm.org

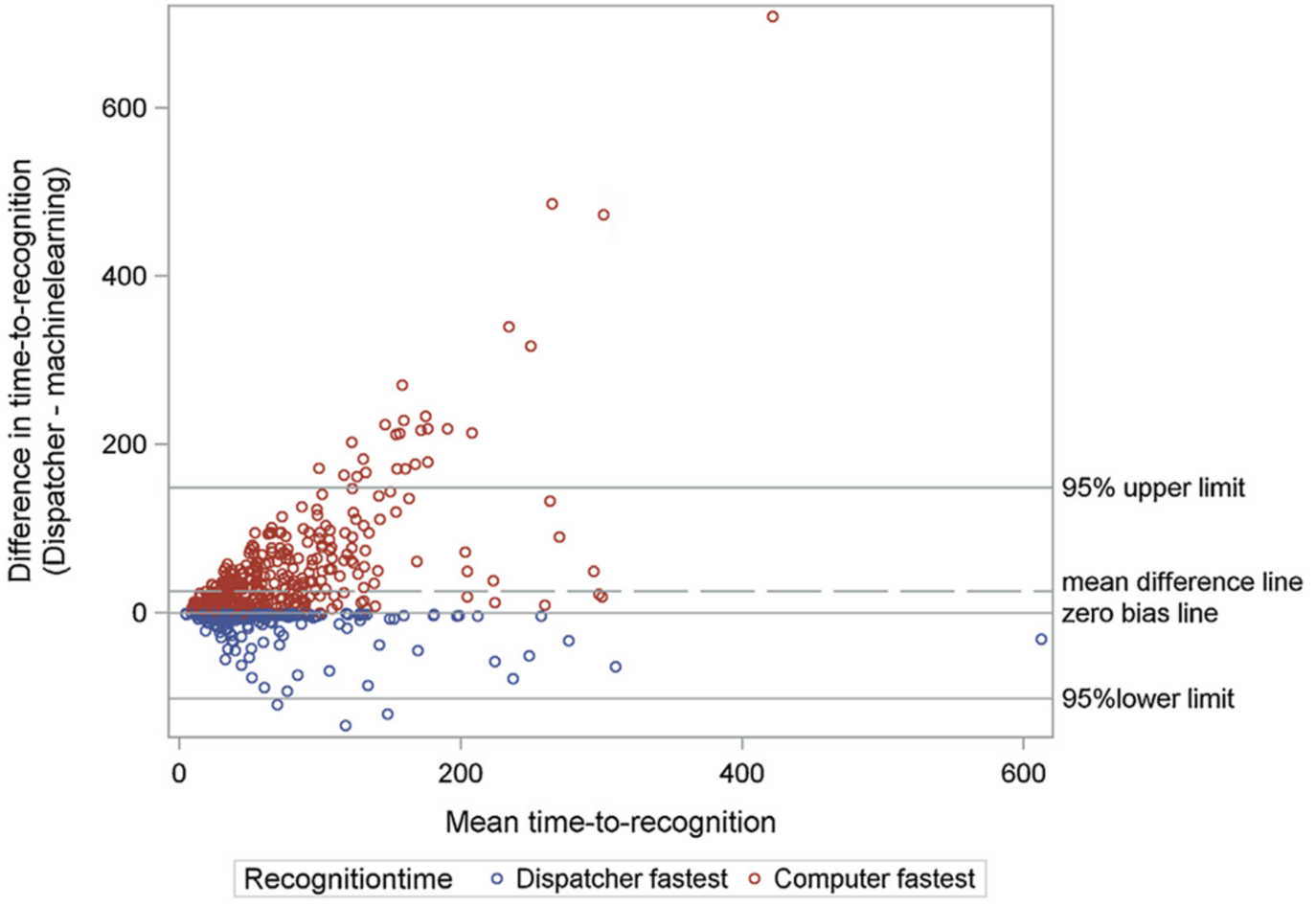
With the widespread use of artificial intelligence (AI) systems and applications in our everyday lives, accounting for fairness has gained significant importance in designing and engineering of such systems. AI systems can be used in many sensitive environments to make important and life-changing decisions; thus, it is crucial to ensure that these decisions do not reflect discriminatory behavior toward certain groups or populations. More recently some work has been developed in traditional machine learning and deep

MOSTRA ALTRO

☆ Salva Cita Citato da 3657 Articoli correlati Tutte e 10 le versioni Web of Science: 926 »

Fairness within decision domains





Skin color within domain of image analysis, but outside domain of emergency calls

Definition fairness

*Historical

"An eye for an eye and a tooth for a tooth..."

Basis for retributive justice, *Lex Talionis*

Outline

AI analysis skin lesions

What is fairness?

Equals treated equally, and unequals proportionately unequally within decision domain

Hiring algorithms

Individual versus group fairness: Group

Individual versus group fairness: Individual

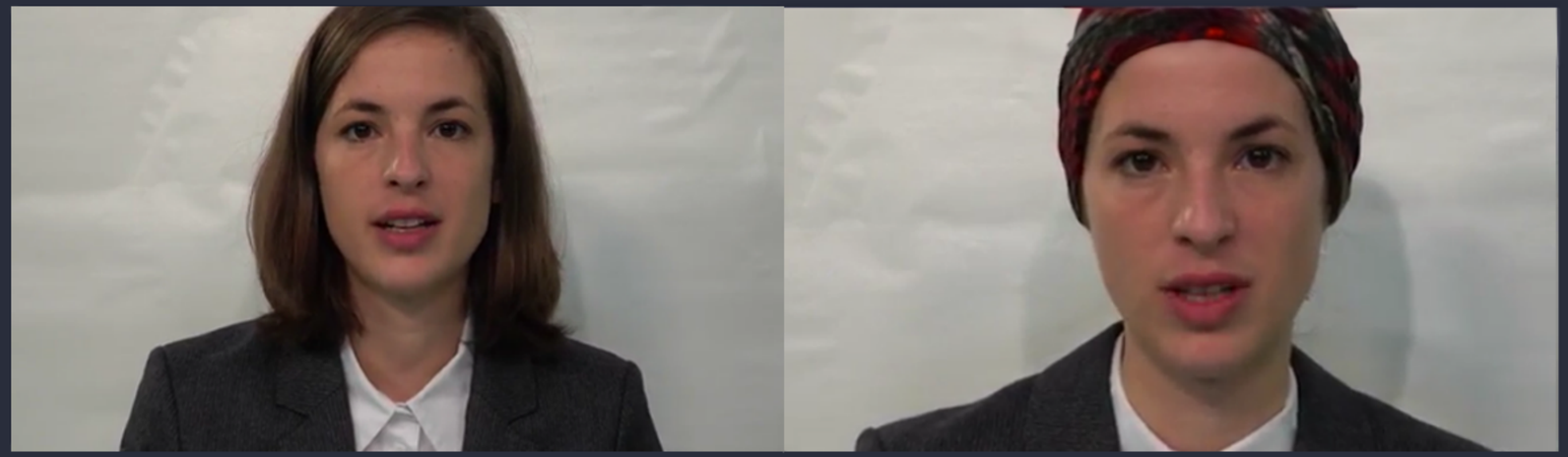
Fairness and privacy dilemma

Representational versus distributive fairness

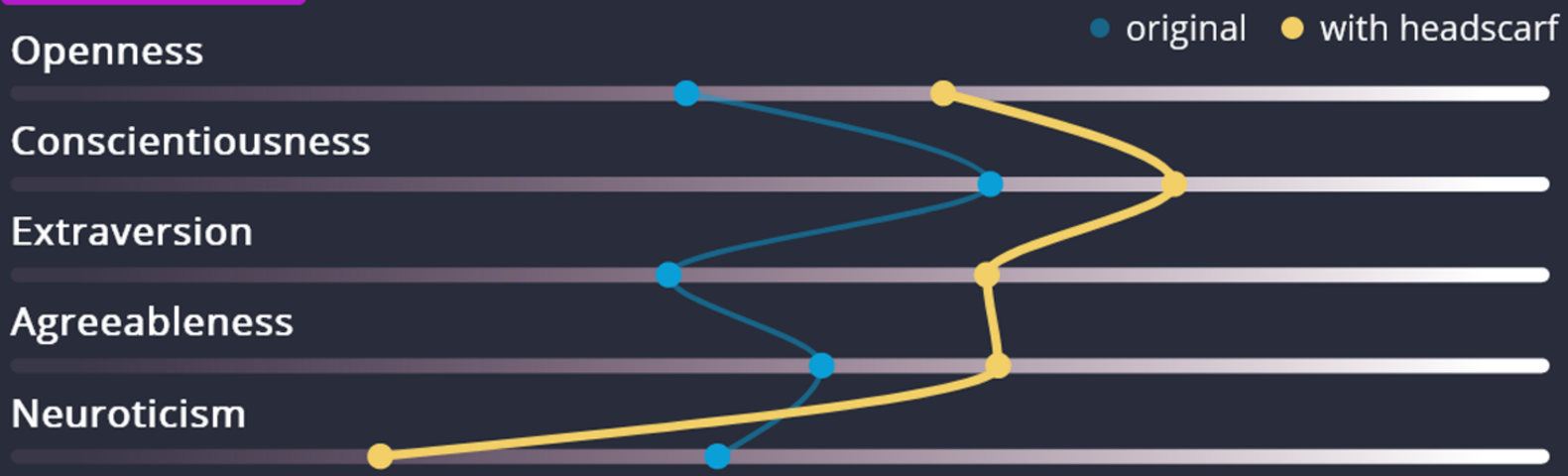
Blindness and insight

Hiring algorithm fairness debate #1

HEADSCARF



OCEAN RESULTS



Process:
Applicants stage and submit their own video

Experiment:
Actors repeat the same presentation in diverse contexts

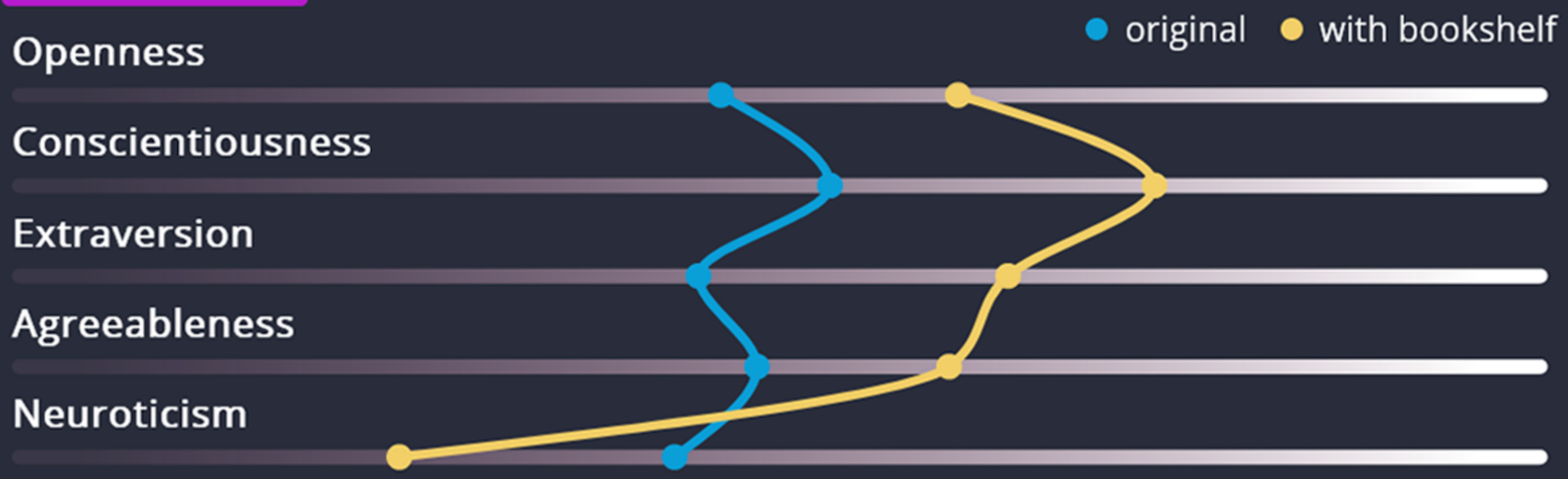
Source: German newsmedia report on hiring software
https://interaktiv.br.de/ki-bewerbung/en/

Hiring algorithm fairness debate #1

BACKGROUND



OCEAN RESULTS



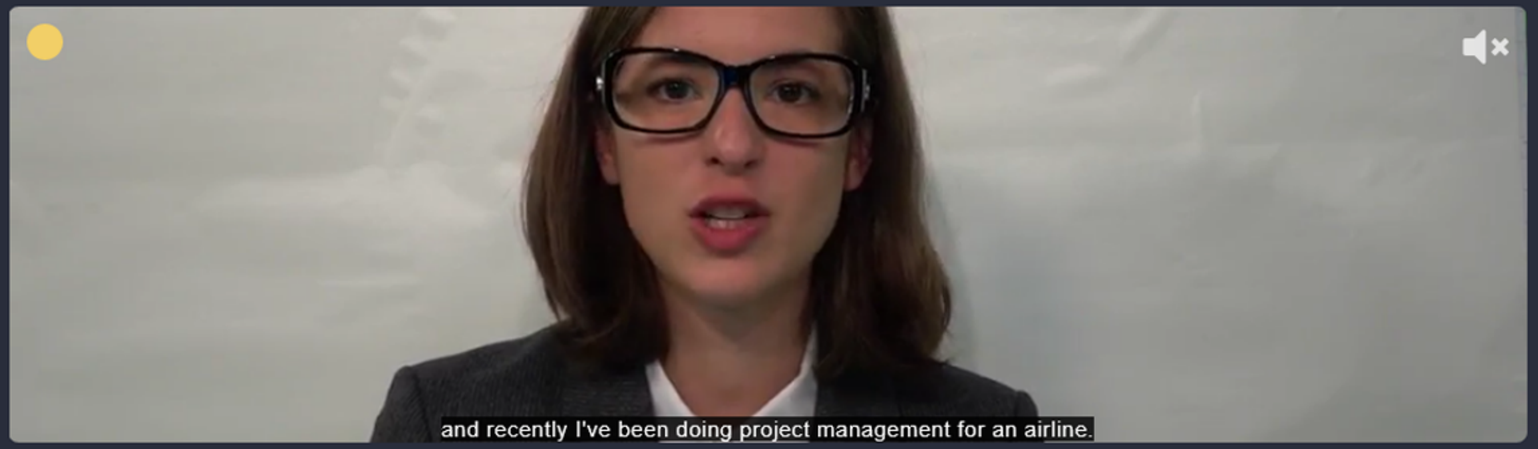
"Retorio's AI was trained using videos of more than 12,000 people of different ages, gender and ethnic backgrounds.

2,500 people rated how they perceived them in terms of the personality dimensions based on the Big Five model.

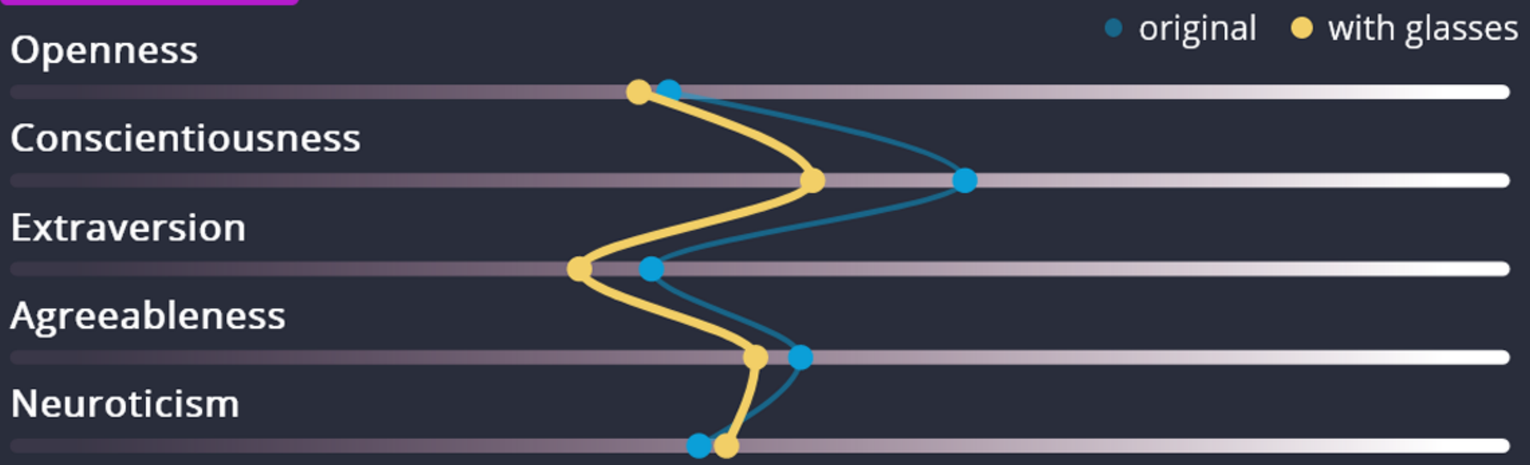
According to the the start-up the AI's assessments have an accuracy of 90 percent compared to those of a group of human observers."

Case note #1: Hiring algorithm fairness

GLASSES



OCEAN RESULTS



Are glasses different?

(Do different glasses have different effects)

How could the AI designer manage the glasses effect?

Case note #2: Hiring algorithm fairness

"The fundamental problem with face recognition by machine learning is that we never know exactly which pattern in an image these machines are responding to."

- Computer science professor Katharina Zweig

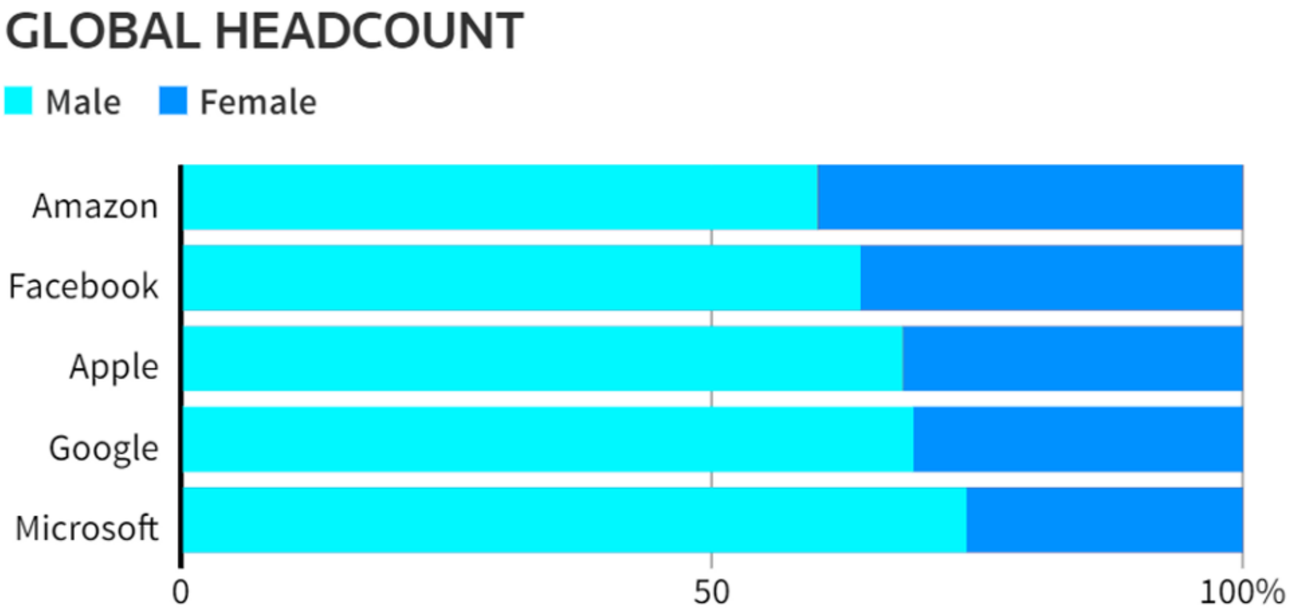
When *humans* judge the compentence of a person, do we always know what influences our judgement?

→



Fairness hiring algorithm debate #2: Amazon & domain spillover

Top U.S. tech companies have yet to close the gender gap in hiring, a disparity most pronounced among technical staff such as software developers where men far outnumber women. Amazon's experimental recruiting engine followed the same pattern, learning to penalize resumes including the word "women's" until the company discovered the problem.

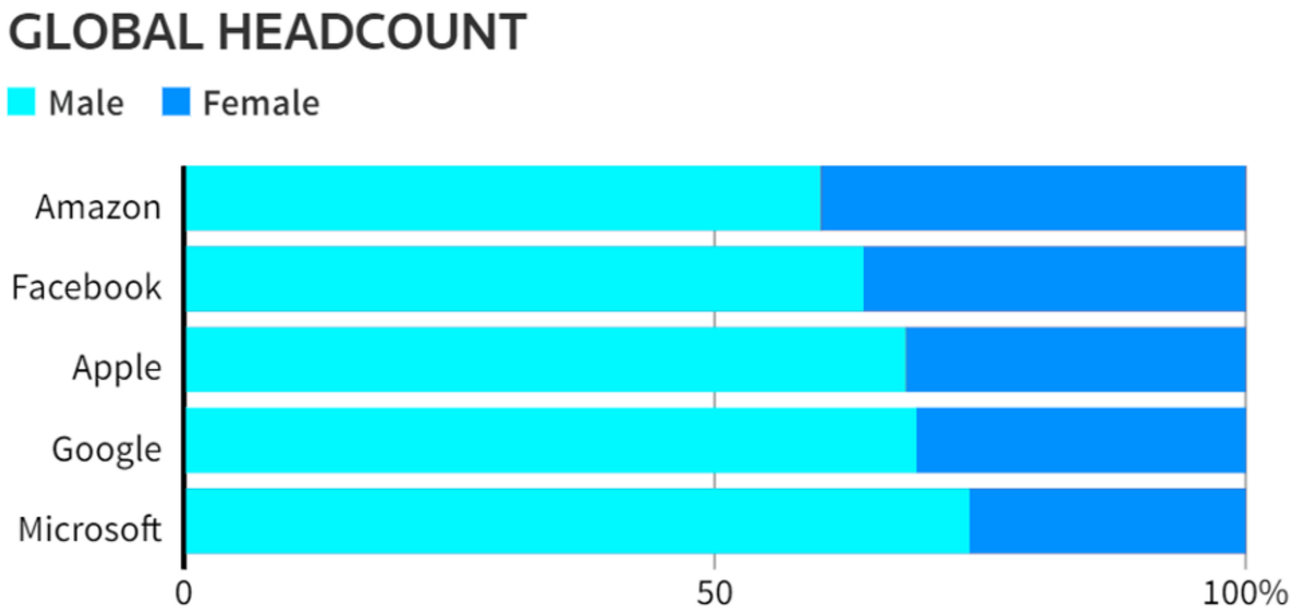


Software task:
Identify characteristics of successful current employees, then filter applicants for same traits

Past data bias --> future prediction bias

Fairness hiring algorithm debate #2: Amazon & domain spillover

Top U.S. tech companies have yet to close the gender gap in hiring, a disparity most pronounced among technical staff such as software developers where men far outnumber women. Amazon's experimental recruiting engine followed the same pattern, learning to penalize resumes including the word "women's" until the company discovered the problem.



Replace:
Male --> Tall male
Female --> Short male
Amazon --> Basketball team

Questions: Fairness / hiring / Amazon / domain spillover

1. Is the problem in technology or society? Should developers get involved in re-balancing results to undo the M preference, or is that outside their scope?
   - Given the many M/F proxies, is re-balancing even possible?

2. What happens if the imbalance arises from a statistical anomaly invisible to common sense? **Is this unfair…?**
   - Many employees with last name starting with "W," almost none with "S"

3. Can the machine be used if unbiased results cannot be guaranteed?
   - What biases to what extent will we accept, if any?
   - Are invisible biases fair…?

Outline

AI analysis skin lesions

What is fairness?
- Equals treated equally, and unequals proportionately unequally within domain

Hiring algorithms
- Equals/Unequals, domain spill

Individual versus group fairness: Group

Individual versus group fairness: Individual

Fairness and privacy dilemma

Representational versus distributive fairness

Blindness and insight

Individual versus group fairness (ethics versus law)

Background:

Fairness comparisons between groups, not individuals

Social Justice

- Vocabulary evolving: "Statistical fairness," or "Statistical" versus individual fairness (Kearns, Roth)

Origin of divide:

Group fairness emerges from language of legal protected groups

= flood of statistical comparisons (Gender/race/age/religion etc....)

- Ethics and law distinct in theory, maybe not in practice

Conceptual essence of dilemma:

Can it be that individuals are treated fairly, while collectives of same individuals are treated unfairly?

Individual versus group fairness: 2 cases

ProPublica versus COMPAS

Lemonade

Individual versus group fairness: ProPublica accuses COMPAS



Bernard Parker, left, was rated high risk; Dylan Fugett was rated low risk. (Josh Ritchie for ProPublica)

Machine Bias

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

by Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, ProPublica

May 23, 2016

Individual versus group fairness: ProPublica accuses COMPAS

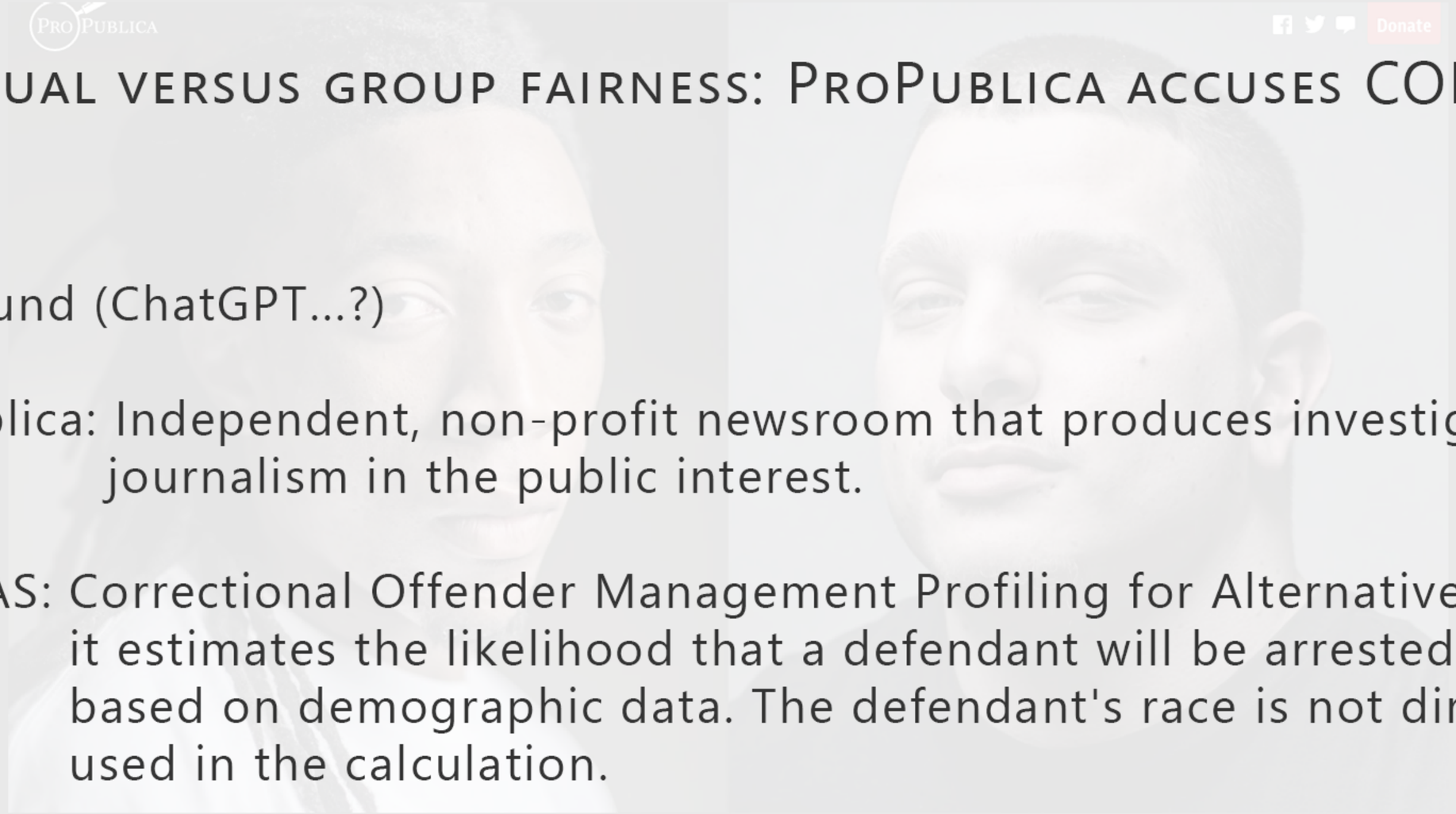
Background (ChatGPT...?)

ProPublica: Independent, non-profit newsroom that produces investigative journalism in the public interest.

COMPAS: Correctional Offender Management Profiling for Alternative Sanctions, it estimates the likelihood that a defendant will be arrested again, based on demographic data. The defendant's race is not directly used in the calculation.

Tool → Used by judges to determine whether bail should be granted.
→ High stakes.

*COMPAS software developed by Northpointe inc.



Bernard Parker, left, was rated high risk. Dylan Fugett was rated low risk. (Josh Ritchie for ProPublica)

Individual versus group fairness: ProPublica accuses COMPAS

[Background, COMPAS data to predict recidivism, selected, typical]

Criminal history including severity of offenses, number of past convictions, time since the last offense.

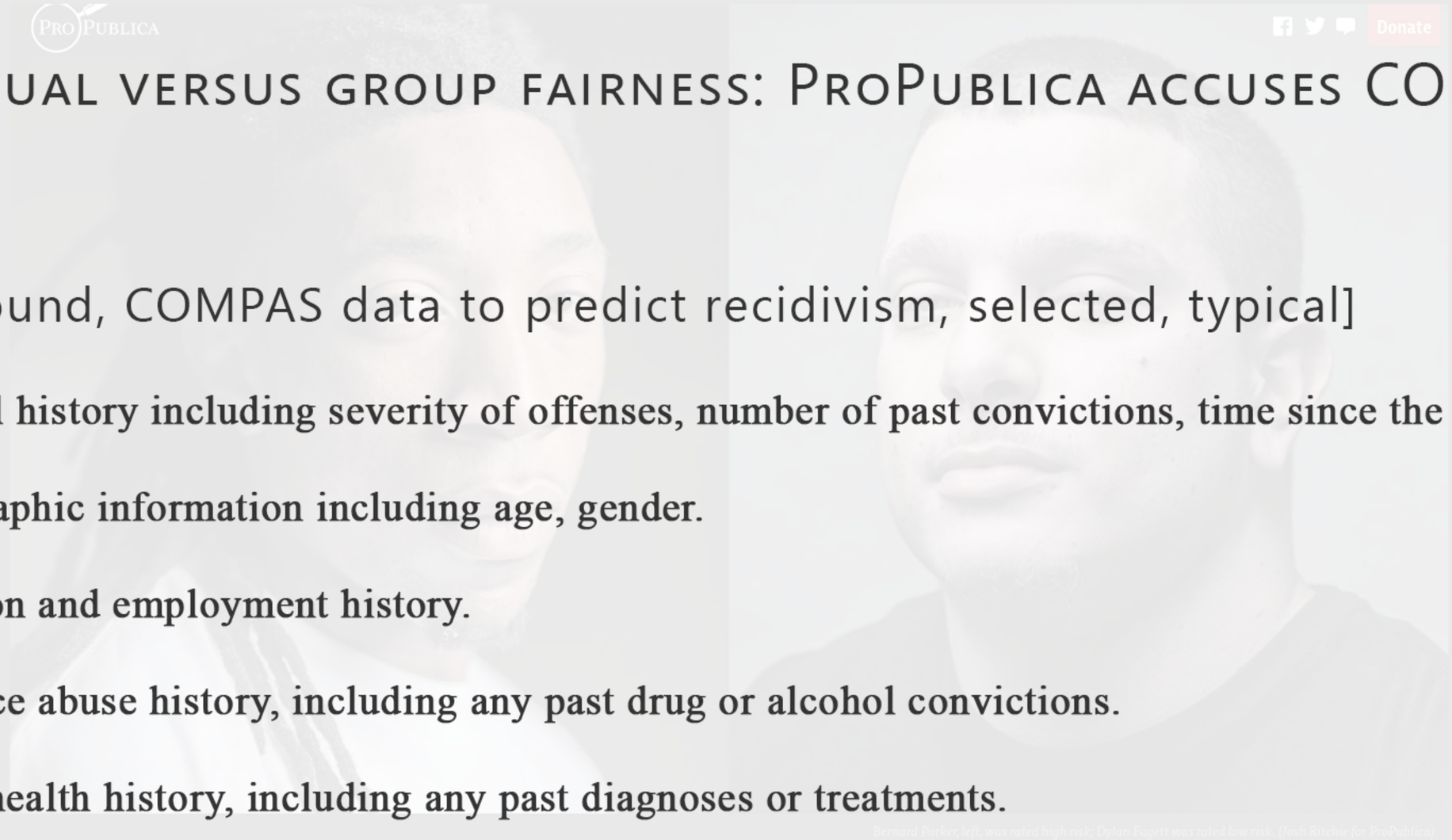
Demographic information including age, gender.

Education and employment history.

Substance abuse history, including any past drug or alcohol convictions.

Mental health history, including any past diagnoses or treatments.

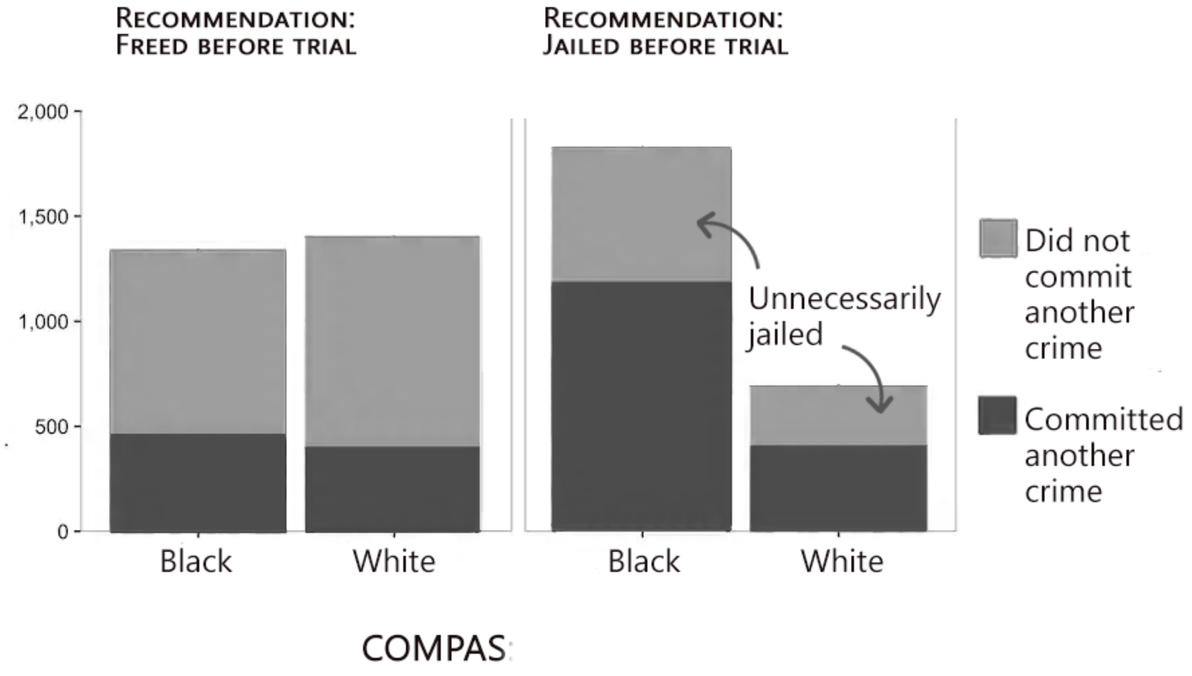
Social history, including family and community support, history of domestic violence and child abuse.



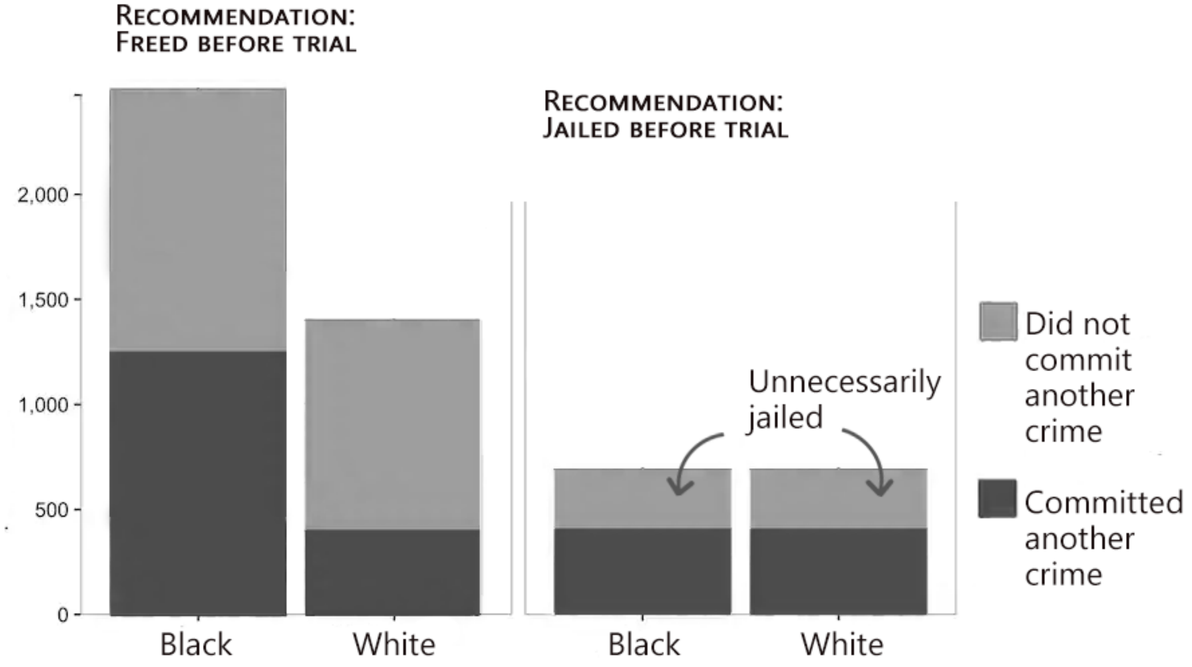
Bernard Parker, left, was rated high risk. Dylan Fugett was rated low risk. (Josh Ritchie for ProPublica)

Fair = Equally accurate predicting recidivism for all individuals, regardless of their race, gender, or other demographic factors. (Score of 7 means some thing for everyone.)

Fair = False positive rate measured against entire population should be same for all demographic groups.



COMPAS

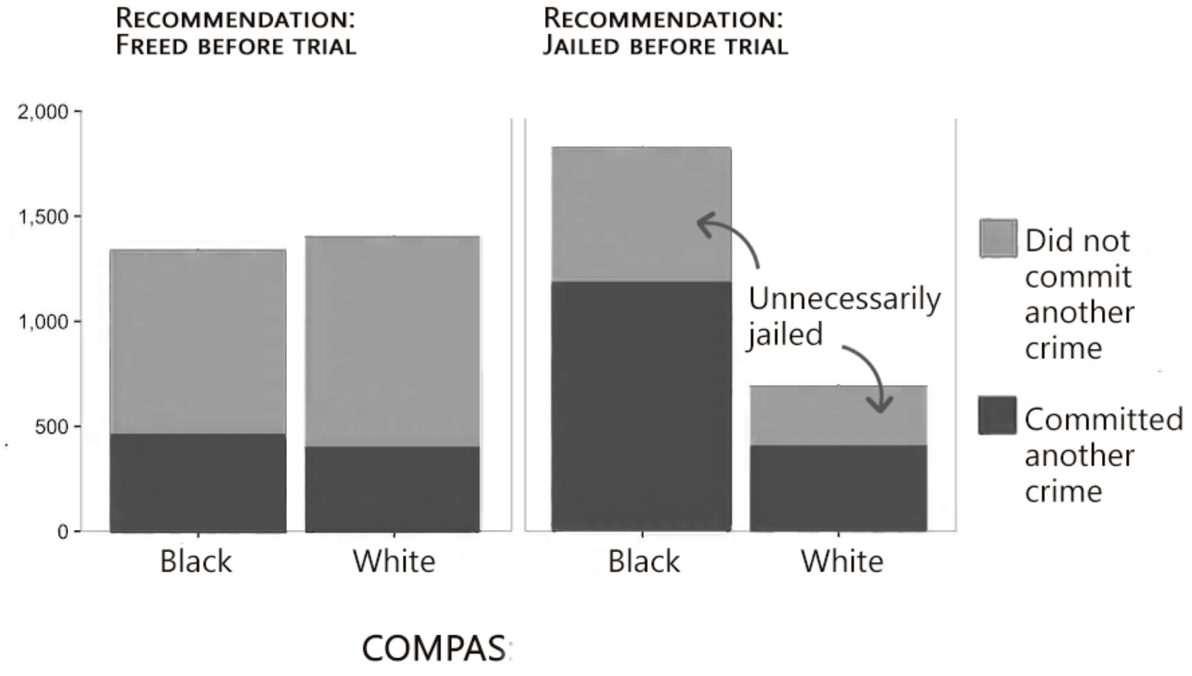


ProPublica

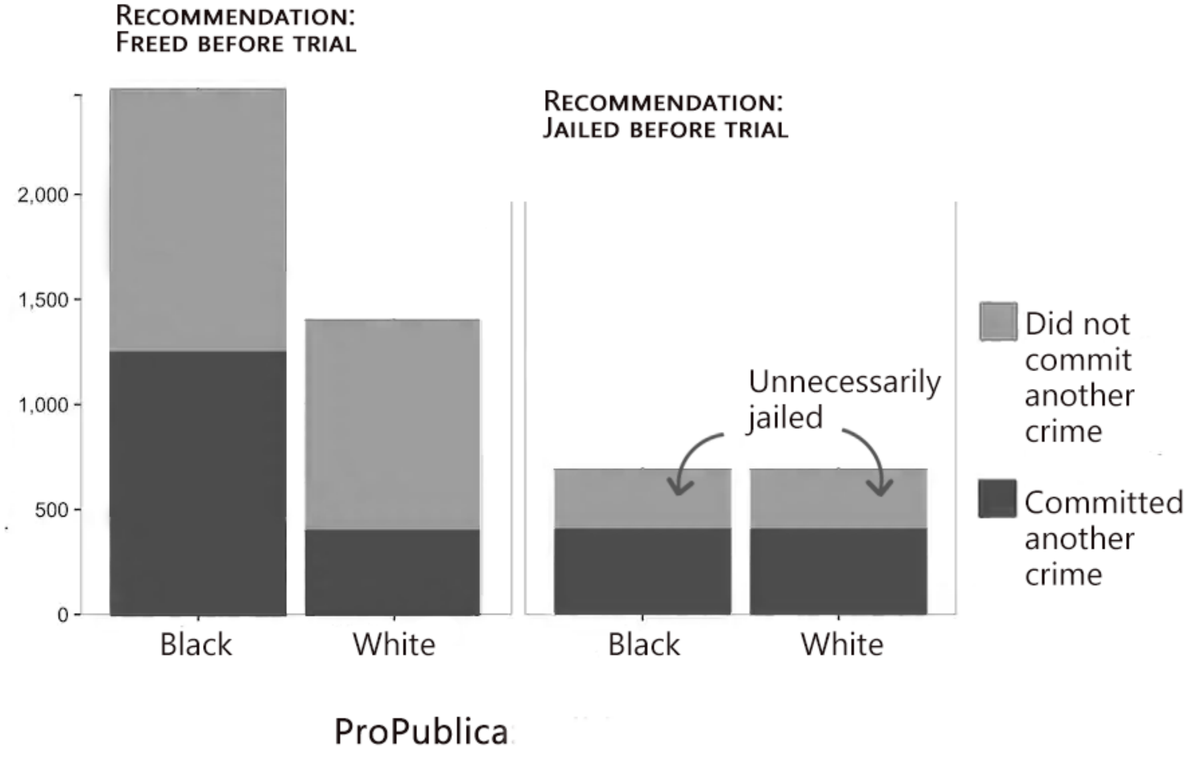
CASE: COMPAS and ProPublica redistribution: ***Which one is fair? For whom?***

Case: Individual versus group fairness, ProPublica / COMPAS

*Can historical racial unfairness affect what counts as fairness today, and is that part of data and algorithms?



COMPAS



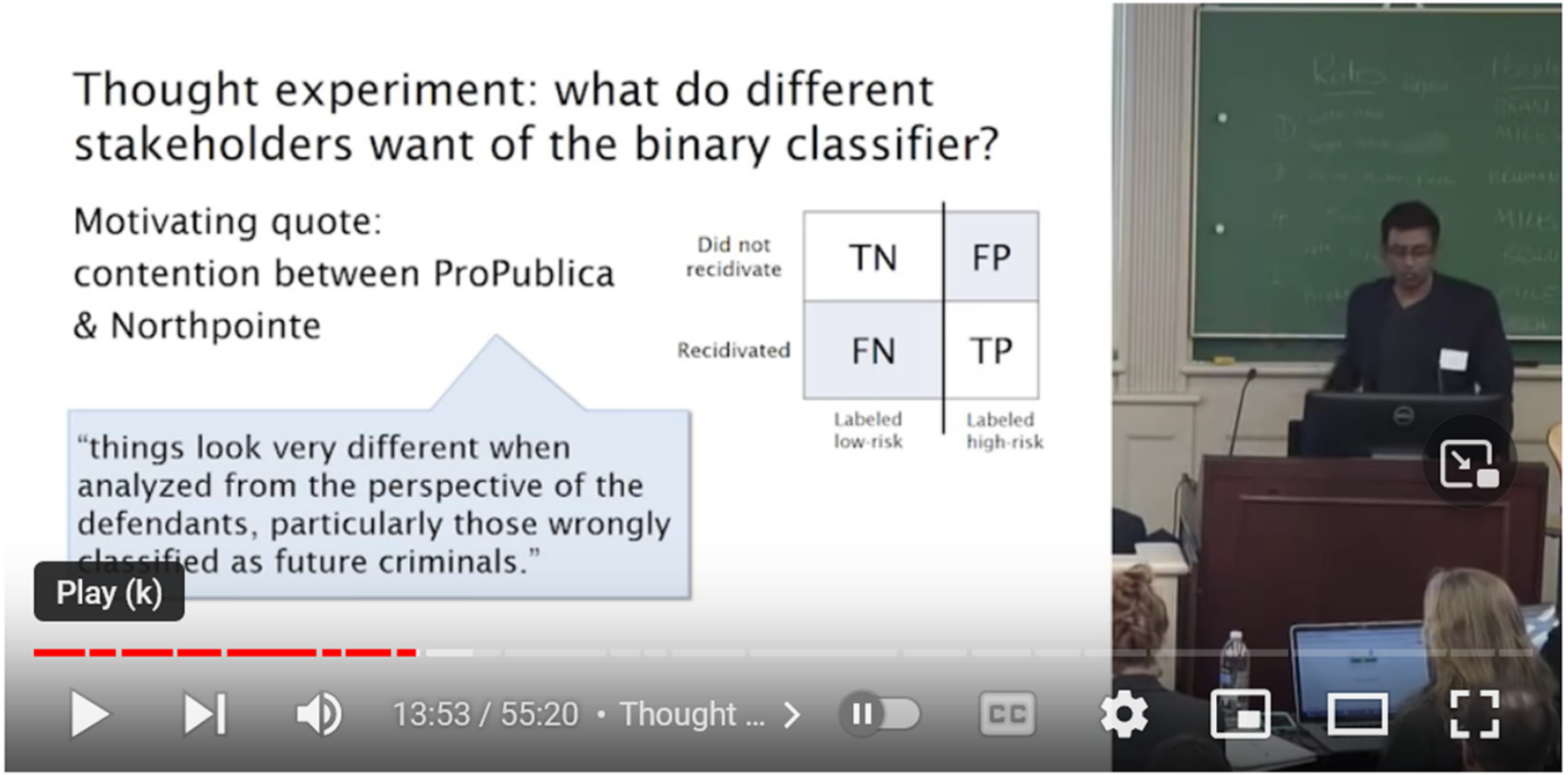
ProPublica

Case: Individual versus group fairness, ProPublica / COMPAS

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

Example, group fairness argument:

Fairness determined at group level, and by the particular group most affected by possible misclassification



Tutorial: 21 fairness definitions and their politics

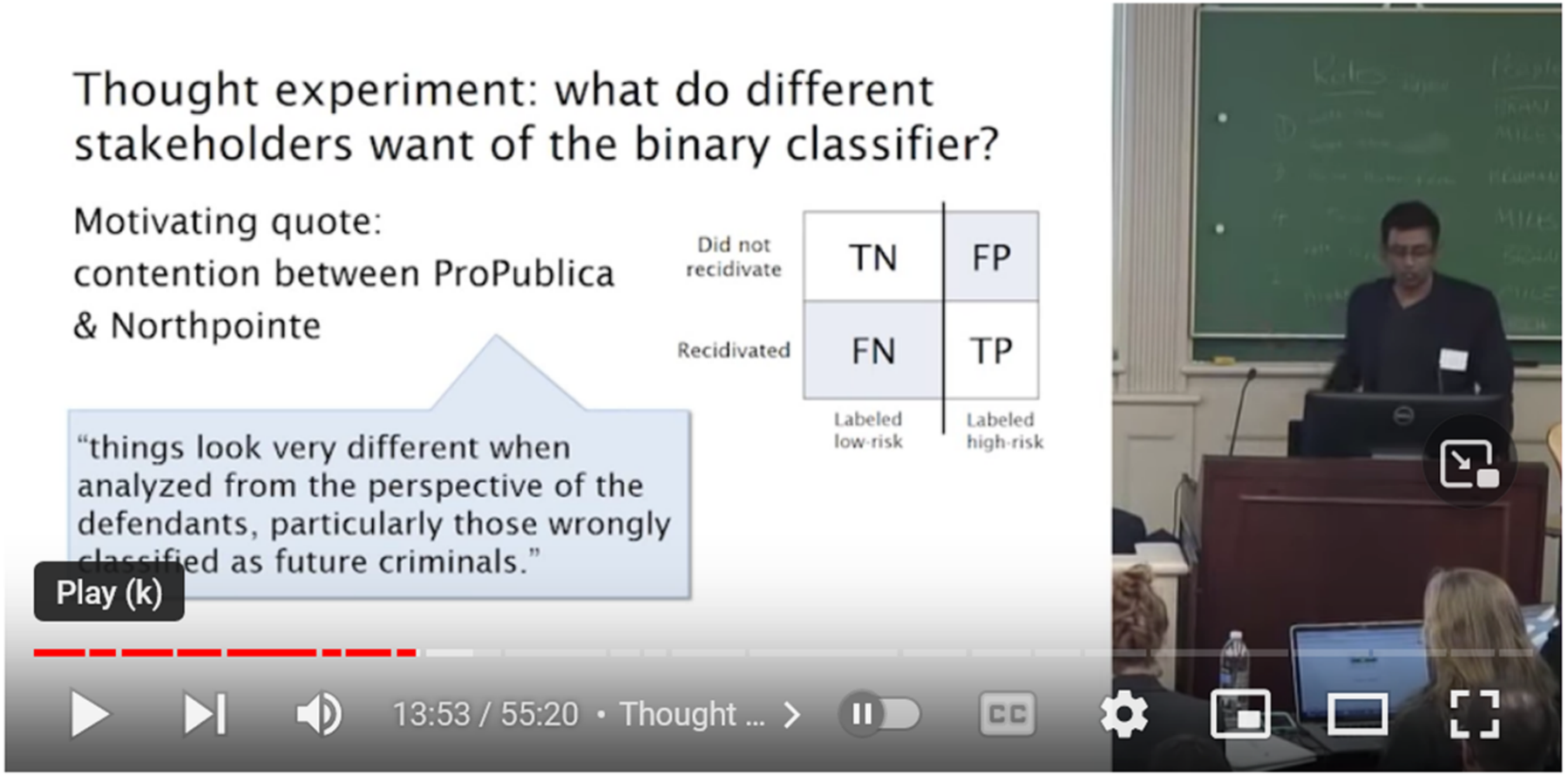
Arv...
382...
Subscribe
DOWNLOAD
393
Share

Case: Individual versus group fairness, ProPublica / COMPAS

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

Counter argument:

Category mistake, idea is not to get one or another perspective, but no perspective



Tutorial: 21 fairness definitions and their politics

Arv...
382...
Subscribe
DOWNLOAD
393
Share

*Group fairness note: Simpson's paradox (Proportion confusion)

In terms of winning admission, are men, or women, or neither treated unfairly?

University of California-Berkeley
Graduate admission figures, fall 1973

Department	Men		Women	
	Applicants	Admitted	Applicants	Admitted
A	***825***	62%	108	82%
B	***560***	63%	25	68%
C	325	37%	***593***	34%
D	417	33%	375	35%
E	191	28%	***393***	24%
F	373	6%	341	7%
Total				

*Group fairness note: Simpson's paradox (Proportion confusion)

In terms of winning admission, are men, or women, or neither treated unfairly?

University of California-Berkeley
Graduate admission figures, fall 1973

Department	Men		Women	
	Applicants	Admitted	Applicants	Admitted
A	***825***	62%	108	82%
B	***560***	63%	25	68%
C	325	37%	***593***	34%
D	417	33%	375	35%
E	191	28%	***393***	24%
F	373	6%	341	7%
Total	2691	45%	1835	30%

Group fairness conclusion

Group fairness can conflict with individual fairness, and be misleading. What about individual fairness?

Outline

AI analysis skin lesions

What is fairness?
- Equals treated equally, and unequals proportionately unequally within domain

Hiring algorithms
- Equals/Unequals, domain spill

Individual versus group fairness: Group

Individual versus group fairness: Individual

Fairness and privacy dilemma

Representational versus distributive fairness

Blindness and insight

Individual fairness *against* group fairness: Lemonade example

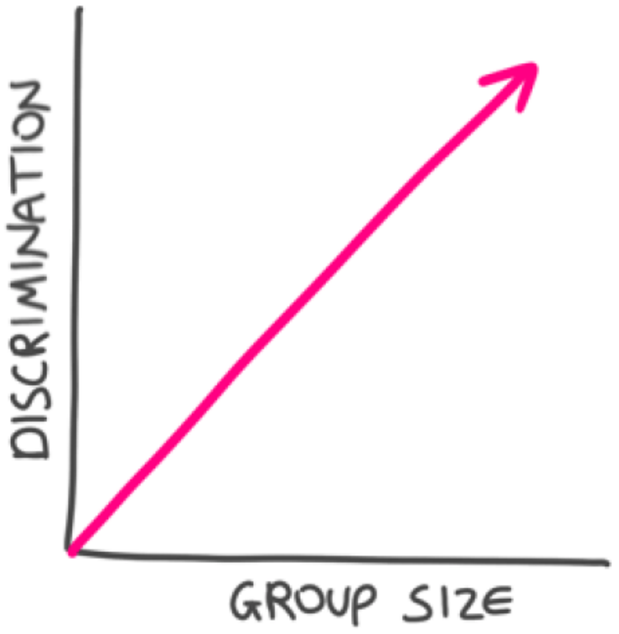
Background



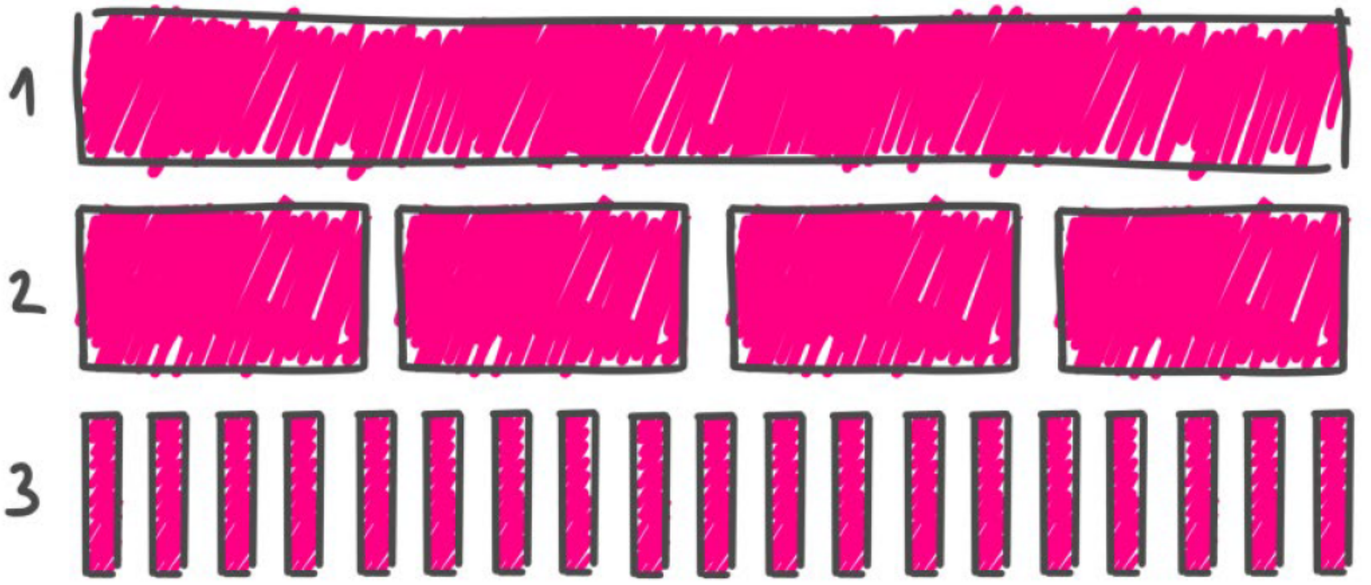
Individual fairness *against* group fairness: Lemonade example

Solve fairness problems with radical individuality

Instead of balancing different collectives, distribute for individual characteristics…



To understand why, let's think about the process of using data to segment – or 'discriminate' – as evolving in 3 phases.



Individual fairness *against* group fairness: Lemonade example

https://www.lemonade.com/blog/ai-can-vanquish-bias/

Argument for individuality:

"Let's say I'm Jewish (I am) and my tradition involves lighting candles every Friday. People like me represent a higher risk of fire than the national average.

If AI charges Jews, on average, more than others for fire insurance, is that unfair?

Yes, if being Jewish resulted in higher premiums. That, though, is a Phase 2 problem.

A Phase 3 algorithm identifies people's proclivity for candle lighting, and charges them more for the risk. It is fair.

*The fact that fondness for candles is concentrated among Jews means that, on average, Jews will pay more. It does not mean that people are charged more for being Jewish.

Case: Individual fairness *against* group fairness: Lemonade

1. Is the Lemonade argument persuasive (If every individual is treated fairly, can a collection of those individuals be treated unfairly?)

2. Lemonade sells house insurance, would the rules for fairness be different for other types of insurance…?

3. Does individual fairness defeat the point of insurance…? (As a vehicle for *sharing* risk…)

Outline

AI analysis skin lesions

What is fairness?
- Equals treated equally, and unequals proportionately unequally within domain

Hiring algorithms
- Equals/Unequals, domain spill

Individual versus group fairness: Group

Individual versus group fairness: Individual

Fairness and privacy dilemma

Representational versus distributive fairness

Blindness and insight

Privacy v Fairness

| 32 | Female | New York, NY | Bachelor's | Single | $60,000 | Travel, Food,
in | 28 | Male | Los Angeles, CA | Master's | Married | $80,000 | Technology,
| 24 | Female | Chicago, IL | High School | Single | $40,000 | Fashion, Musi
| $100,000 | Finance, Readi
| $50,000 | Arts, Cooking,
| 31 | Male | San Francisco, CA | Bachelor's | Married | $90,000 | Startups, Hiki
| 27 | Female | Boston, MA | Master's | Single | $70,000 | Science, Gami
| 33 | Male | Seattle, WA | Bachelor's | Married | $75,000 | Technology, Tr
| Travel, Food, Fitness | High | White | Christian | Heterosexual
| Technology, Gaming, Movies | Moderate | Asian | Jewish | Homos
| Fashion, Music, Outdoor Sports | Low | Black | Muslim | Bisexual
| Finance, Reading, Wine Tasting | High | White | Christian | Heterosexu
| Arts, Cooking, Yoga | Moderate | Hispanic | Buddhist | Heterosexu
| Startups, Hiking, Photography | High | White | Atheist | Heterosexu
| Science, Gaming, Anime | Moderate | Asian | Hindu | Homosexua
| Technology, Travel, Music | High | White | Christian | Heterosexual
| Sports, Shopping, Social Media | Low | Hispanic | Catholic | Bisexual

Require people to reveal their religion to be fair...?

What can be required in name of social fairness?

Privacy v Fairness case: Is ChatGPT right?

You

When AI is used to screen loan applications, what personal information should we demand from loan applicants to ensure fairness? Please list the factors as short bullet points.

ChatGPT

- Income level
- Employment history
- Credit history and scores
- Current debt obligations
- Age (with caution to avoid age discrimination)
- Residency status
- Avoid using race, ethnicity, gender, religion, or marital status to prevent discrimination
- Consider alternative data (e.g., rent payment history, utility payments)
- Ensure transparency about data collection and use
- Audit AI models for biases and unintended consequences
- Ensure accessibility for people from diverse backgrounds
- Adhere to financial regulations and data protection laws

Outline

AI analysis skin lesions

What is fairness?
- Equals treated equally, and unequals proportionately unequally within domain

Hiring algorithms
- Equals/Unequals, domain spill

Individual versus group fairness: Group

Individual versus group fairness: Individual

Fairness and privacy dilemma

Representational versus distributive fairness

Blindness and insight

Representational fairness (versus distributive)

Google CEO search, 2018 (Subsequently adjusted results)

Fairness defined in terms of ***representation*** of opportunity as opposed to ***distribution*** of opportunity

"You Can't Be What You Can't See..."

* Mathematical bias versus ethical fairness



Google ceo

Case: Representational fairness (versus distributive)

"You Can't Be What You Can't See…"



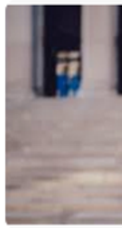

OnBoard CEO vs. Owner: What's the Differen… | Corporate Finance Institute CEO vs. CFO - Definitions, Differences … | Wikipedia Sundar Pichai - Wikip… | University of Lincoln What does CEO stand for? - University … | The Motley Fool Top Black CEOs: Ex… | CU Management By the Numbers: Women as CU CEOs | CU … | LinkedIn Traits that make a successful female CEO | UGA Today Research finds











ecutive Search a Woman Under 55 … | USA Today Meet the Black CEO fighting for grea… | Robb Report How to Becor

What - if anything - should Google do?

Algorithm adjustment options? Where does the adjusting stop? (Where did it start, is there a "neutral"?)

What other values come into play? (What are we willing to exchange (dignity, etc.) to represent fairly?




Business Insider Black CEOs of Fortune 500 Compa… | Corporate - LAWS CEO: What do they do? - LAWS.com











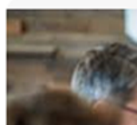
Entrepreneur What Are the Duties of a CEO … | Yahoo Finance CEO vs. President: What's the Difference? | Pix4free Creative Commons ceo Image … | UGA research Female CEOs face subtle bias - UGA … | Time Deloitte CEO Cathy Engelbert on Work… | McKinsey CEO Excellence | Strategy & C… | MIT Technology Rev… Google CEO Sundar… | Hive Life Magazine Chief Executive Office











Certainly! Here is a portrait of a Founding Father of America:




What - if anything - should Google do?

The Gemini disaster







Sure, here is an image of a Viking:



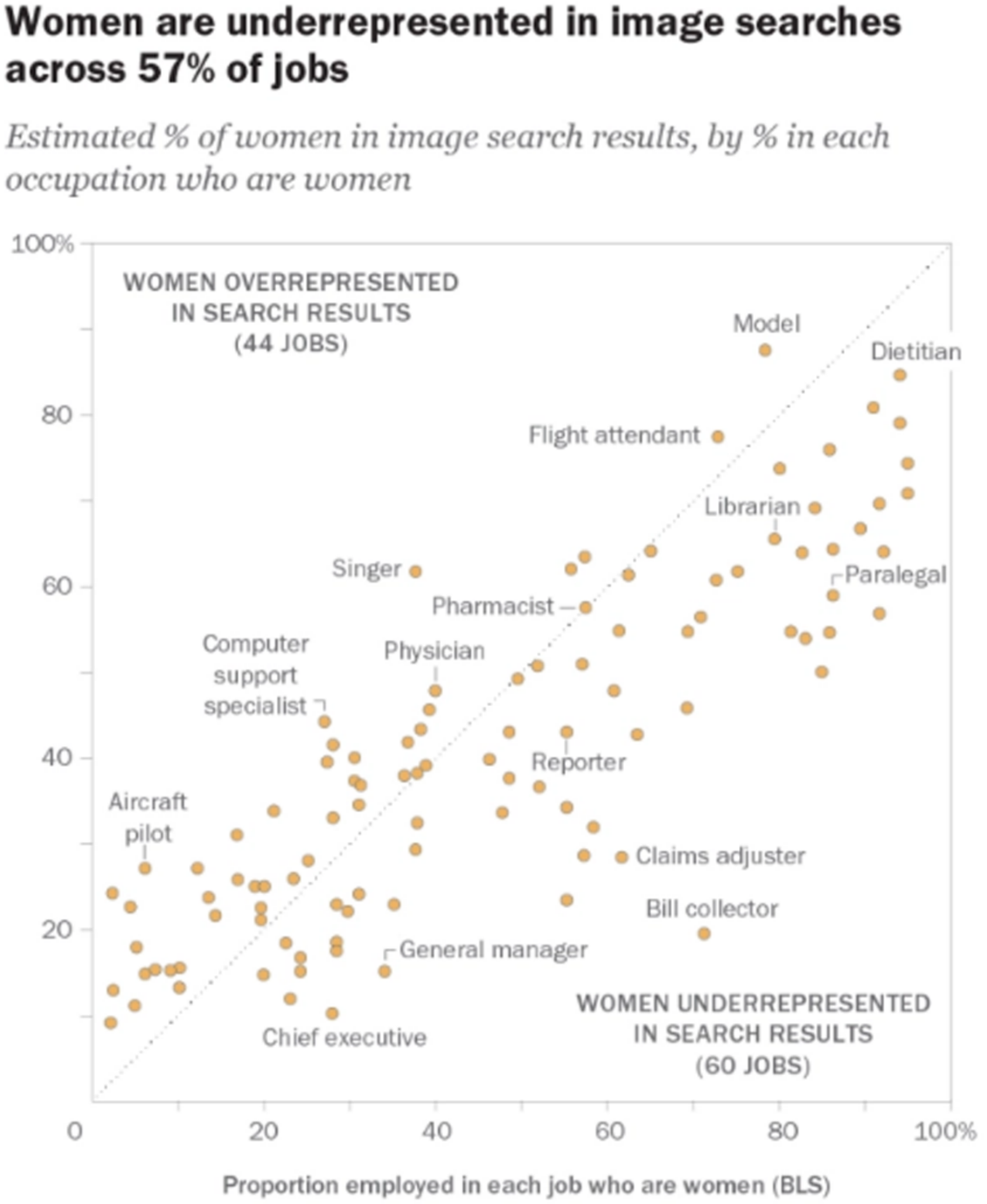

Sure, here is an image of a pope:




Representational fairness (versus distributive)

Extreme case thought experiment:

Should M/F (and races etc.) be **accurately** represented in divergent fields, or **equally** represented in all?

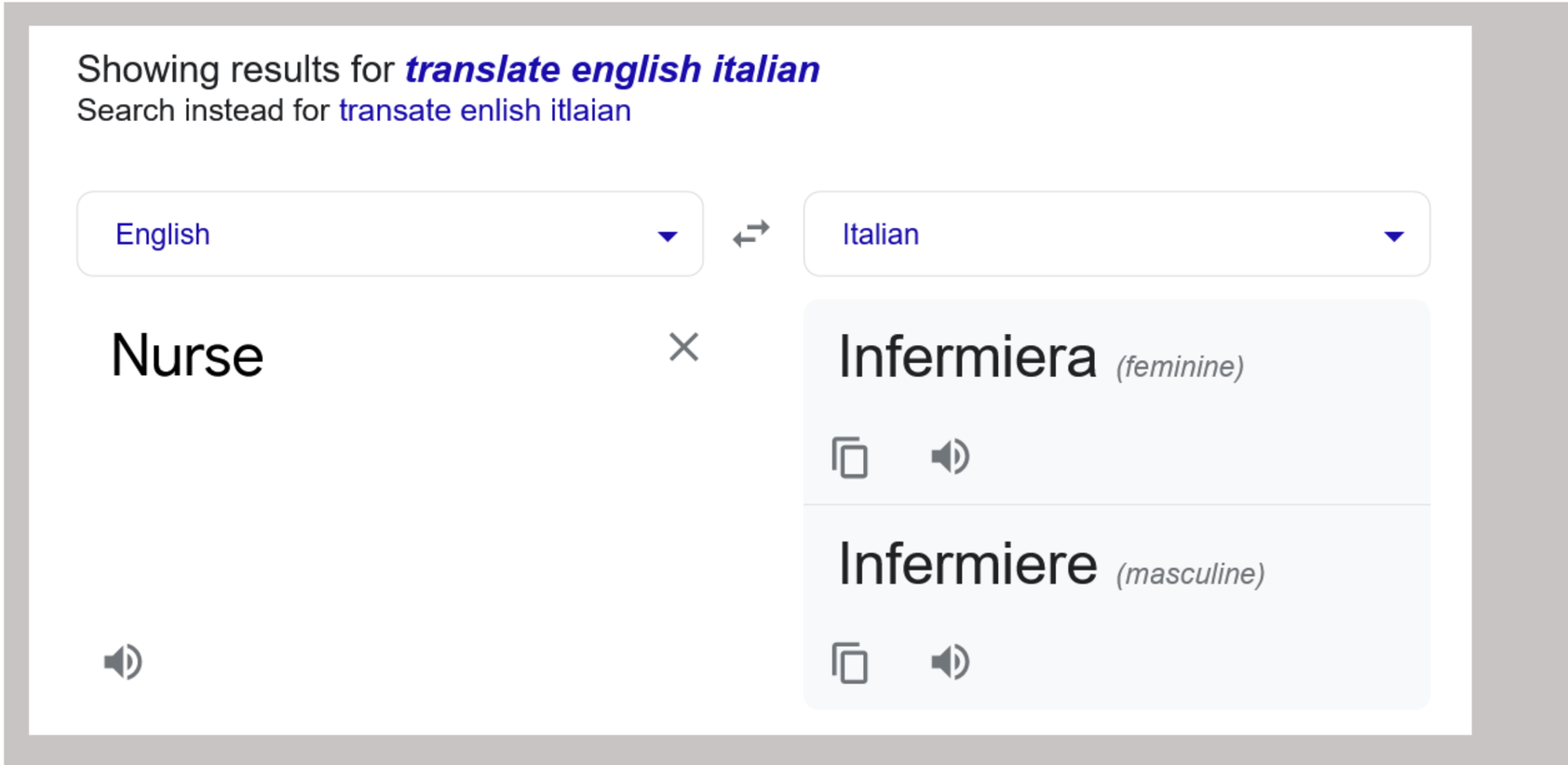


Representational fairness (versus distributive)

"You Can't Be What You Can't See…"

Gendered language translation, from English (non-gendered nouns) to Italian/Spanish/…. where nouns must be assigned a gender

- What to do when translating a sentence?



Outline

AI analysis skin lesions

What is fairness?
    Equals treated equally, and unequals proportionately unequally within domain

Hiring algorithms
    Equals/Unequals, domain spill

Individual versus group fairness: Group

Individual versus group fairness: Individual

Fairness and privacy dilemma

Representational versus distributive fairness: “You Can’t Be What You Can’t See…”

Blindness and insight

Blindness and insight

The theoretical impossibility of fairness - *always starting from unfairness* - because always trying to compensate imbalance of first selection, first data, first application...