CS 312 - Research Methods

Fairness

In AI ethics

Presented by Caroline, Denise & Kailyn

Outline

Individual vs. Group Fairness: Group

Individual Vs. Group Fairness: Individual

Fairness and Privacy

Representational vs. Distributive Fairness

Blindness & Insight

What is Group Fairness?

Group fairness focuses on comparing outcomes across protected groups and grew from legal anti-discrimination frameworks that rely on statistical measures of bias.

- Group fairness ensures that protected groups receive equal treatment are not systematically disadvantaged.
- Group fairness asks whether justice is achieved by equalizing outcomes between social groups rather than only judging how individuals are treated.

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

ProPublica versus COMPAS Case

ProPublica analyzed COMPAS scores used in Broward County, Florida, and found that Black defendants were more likely to be labeled high risk and later not reoffend, while white defendants were more likely to be labeled low risk but later reoffend. COMPAS was accoused by ProPublica for being bias against blacks. Northpointe, the company behind COMPAS, argued that the tool was fair because it had similar overall predictive accuracy across racial groups. The case became a foundational debate because it showed that different definitions of fairness can conflict in practice.

- ProPublica argued that fairness should be evaluated at the group level, especially for the groups that are most harmed by possible misclassification

What is Individual Fairness?

A Focus on the Person, Not the Collective

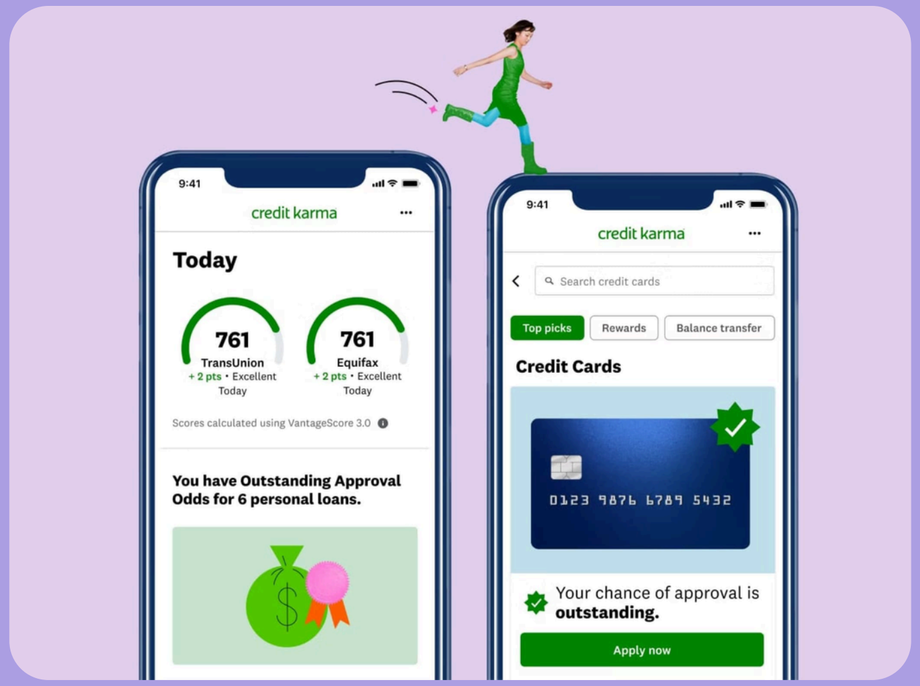
Individual fairness is rooted in the philosophical principle of treating similar people similarly. It focuses on the person's unique data.

- **The Philosophical Side:** "Treat equals equally and unequals unequally." — Aristotle
- **The Goal:** To ensure that an algorithm's decision is based on an individual's specific characteristics and merit rather than their belonging in a “protected” class.

The Question: Is it possible for an AI to be mathematically fair to every person on an individual level, while still perpetuating inequality?

90% Approval Odds

Credit Karma Case



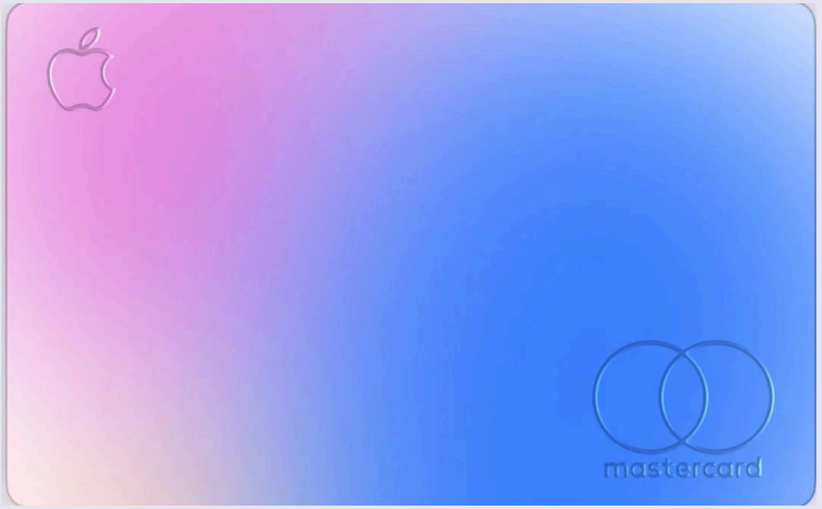
In 2024, the Federal Trade Commission (FTC) finalized a $3 million settlement with Credit Karma after finding that the company used AI-driven "dark patterns" to deceive users. The algorithm provided "individualized" approval odds like "90%" or "Highly Likely" to users who were subsequently denied, causing them unnecessary financial harm through wasted time and lowered credit scores from hard inquiries. This case illustrates that even when AI focuses on an individual's specific history, the absence of transparency can turn "personalized data" into a tool for corporate exploitation.

Source:
https://www.dfs.ny.gov/reports_and_publications/press_releases/pr202103231

Apple Card Case

When Individual Math Fails the Social Test

The 2019 Apple Card case is a perfect example of why individual fairness can be a "trap" in a biased society. When David Heinemeier Hansson went public to criticize Apple's Credit Card failed system, Goldman Sachs argued their algorithm was "gender-blind" because it excluded gender as a variable and looked only at individual credit history. While an investigation found no legal violations, the math ignored that credit scores are already shaped by systemic bias. This proves that treating "equals equally" on paper is socially insufficient if the AI stays blind to the fact that the starting line was never equal to begin with.



Source:
https://www.dfs.ny.gov/reports_and_publications/press_releases/pr202103231

Fairness & Privacy Dilemma

Fairness: is the idea that people will be treated the same

Privacy: the decision to determine what information people know about you

- When we create an AI machine that filters job applications, we want to know that the machine is being fair
- In order to know if the machine is being fair, we have to know who is applying (race, age, gender, and religion)

This creates the Fairness & Privacy Dilemma. What is more important, fairness or privacy?

- If fairness is more important
  - We will need to require applicants to disclose their personal information to determine whether the machine treats them fairly. This will violate people's privacy.
- If privacy is more important
  - People keep their information, and we let the machine choose if person A or person B gets the interview/job. We will never know whether the machine is being fair.
- In the real world, we try to protect both
  - We make it optional for people to disclose their information, and we hope enough people do so to evaluate the machine's fairness.

Representational Fairness (vs Distributive Fairness)

Distributive Fairness: the way social benefits are distributed

Representational Fairness: the representation of an opportunity to get a social good

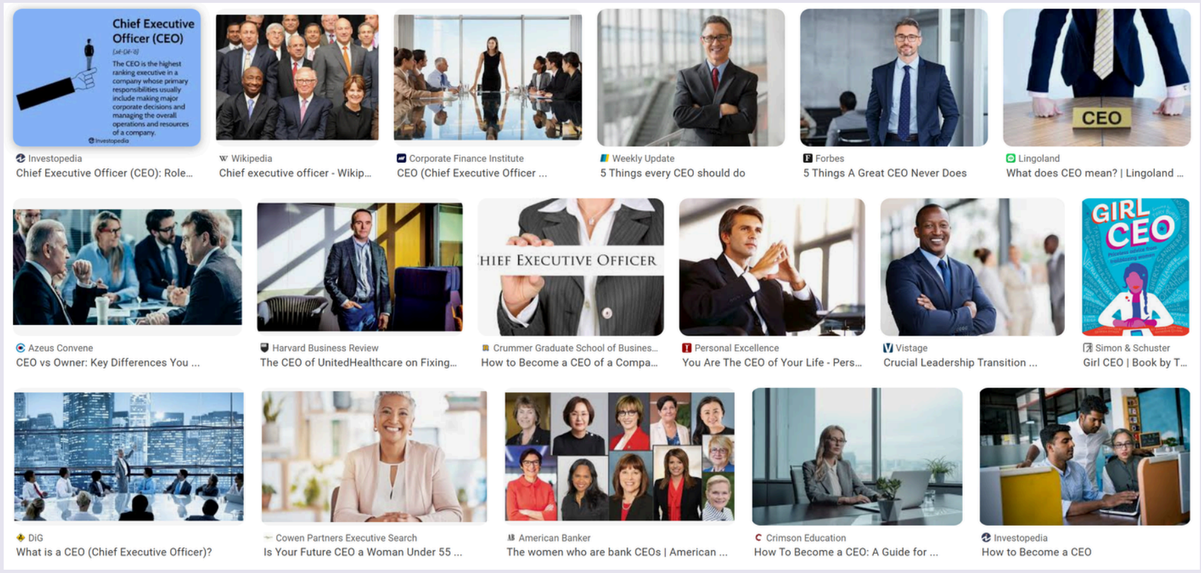
Google CEO

Representational Fairness → the idea that when you search for CEO, you should be able to find everyone (men, women, multiple races, religions, ages)

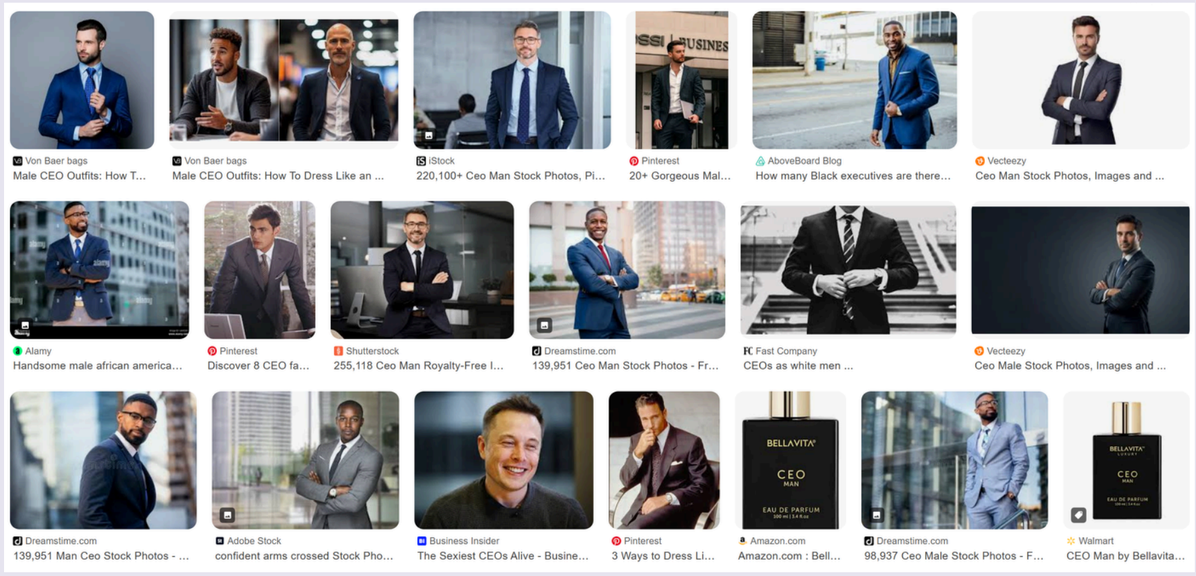
Theoretical Question: Do we want to misrepresent reality to achieve representational fairness?

Prioritize representational fairness

- We would have to make sure that when you search up “CEO” we get images of men, women, people of different ages, races, and religions
- If we choose to prioritize representational fairness, we risk misrepresenting reality

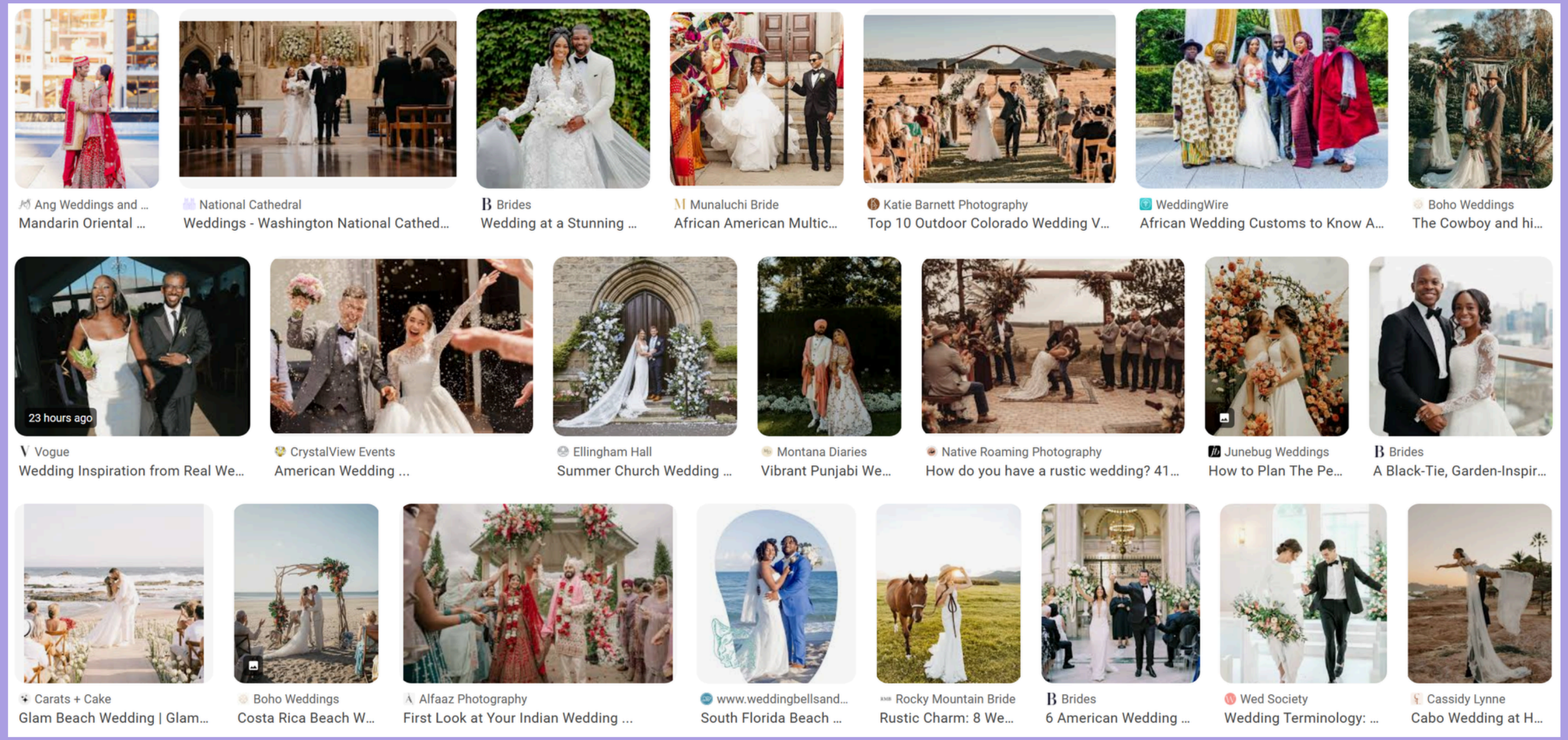


Google CEO 2026



Google CEO 2018

Example of Representational Fairness



Example of Representational Fairness

Men's Health
5 Questions Your Doctor Wis...

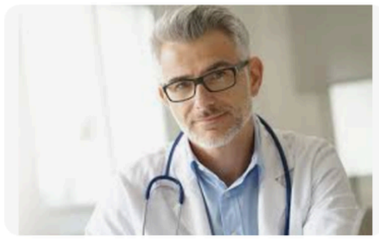
Wellington Regional Medical Center
Wellington Regional Medical Center

Trinity School of Medicine
Trinity School of Medicine

Shutterstock
63,397 Doctor Unifo...

UKNow - University of Kentucky
Feeling nervous talking to your doctor ...

Healthline
Difference Between MD and DO: Cho...



Snibbs
6 Challenges of Being a Doctor | Snibbs

Freepik
Doctor patient clipa...

UAG School of Medicine
4 Reasons to Become a Doctor | UAG...

Barton Associates
Is a Nurse Practitioner a Doctor?

Etsy
Doctor PNG, Cute Doctor Clipart ...

Harvard Health
my hospital care? - Harvard Health

Nanavati Max Hospital
Different Types of Doctors & Their ...

InvestIN Education
How to Become a Doctor in the UK: Guide ...

Scripps Health
Primary Care Doctor ...

Vecteezy
Doctor Stock Photos, Images and ...

Keck Medicine of USC
Medical Doctor and a Nurse Practitioner ...

Blindness and Insight

Why Perfect AI Fairness Is Impossible

Society has centuries of structural bias

That history becomes the data we collect

Data trains our AI models

Choosing not to adjust for group differences is still a choice; it defaults to encoding current inequality.

Most training data reflects who had access, who got arrested, and who got hired. It records the world as it was, not as it should be (and both are true).

Bias in AI is social, economic, and political. No purely mathematical fix can solve what centuries of structural inequality created.

Thank you!