

Data Ethics: Choices and Values

Benjamin Xie, Ph.D.

McCoy Family Center for Ethics in Society - HAI

We use data to inform our decisions

- Evidence-based
- Impartial
- Reliable

QUALITY

5.0

DIFFICULTY

1.0



CS101



AWESOME

May 21st, 2015

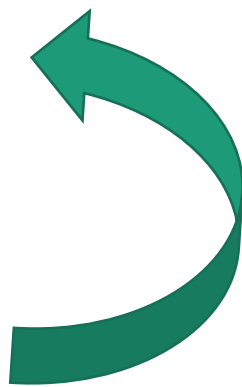
Attendance: **Not Mandatory** Grade: **A** Textbook: **Yes** Online Class: **Yes**

This class was awesome. A beginner like me that has never done anything further than facebook on a computer, Professor . was very clear and easy to listen to. I very much enjoyed the lectures and how easy it was to learn from such a great teacher. Thank you for all that you do



What can we learn from a data set?

- Patterns
- Correlations
- Distributions
- ...



- Choices
- Assumptions
- Values
- Biases



What can we learn
from a data set?

How to interrogate a
data set to ethically
find relevant
elements?



Values in Design

Values in Design

- Design decisions encode values.
- Values are expressive of what we care about
- Values reveal our assumptions about world, people interacting with our designs, and benefiting from them

Values in Design

- **Explicit values:** Values that designers intend their products to embody
- **Collateral values:** values that crop up as side effects of design decisions and the way users interact with them

Explicit Values

Contact-tracing

Health
Safety
Efficiency
Public interest

TECH

Apple and Google will build their coronavirus contact tracing software right into your phone

PUBLISHED TUE, SEP 1 2020-12:00 PM EDT | UPDATED TUE, SEP 1 2020-1:50 PM EDT



WATCH LIVE

Apple and Google partner on COVID-19 contact tracing technology



Collateral Values

- Security?
 - Where is information stored?
 - Encryption?
- Privacy?
 - Who has access to information?
 - Geolocation or bluetooth?
 - What information is accessible to health authorities/ the public?
- Autonomy?
 - Informed consent?

TECH / GOOGLE

Android bug exposed COVID-19 contact tracing logs to preinstalled apps / Preinstalled apps on Android phones could potentially access sensitive data

By [Nicole Wetsman](#)

Apr 27, 2021, 7:20 AM PDT | [0 Comments](#) / [0 New](#)

November 23, 2022

Class Action Filed Against Commonwealth of Massachusetts for Alleged COVID-19 Contact-Tracing Spyware Installation

Kathryn Rattigan

Robinson+Cole Data Privacy + Security Insider

FACT CHECK

fact-checking

[Add Topic +](#)

Fact check: Google did not automatically sign up Android users for COVID-19 tracing app



McKenzie Sadeghi
USA TODAY

Published 7:27 p.m. ET June 14, 2020 | Updated 5:21 p.m. ET July 29, 2020

Explicit Values

Bike-sharing app

Mobility
Health
Sustainability
Inclusion



Collateral Values

Bike-sharing

Who is the default user?
Impact on other mobility?



VALUES IN DESIGN



PROBLEM
FORMULATION



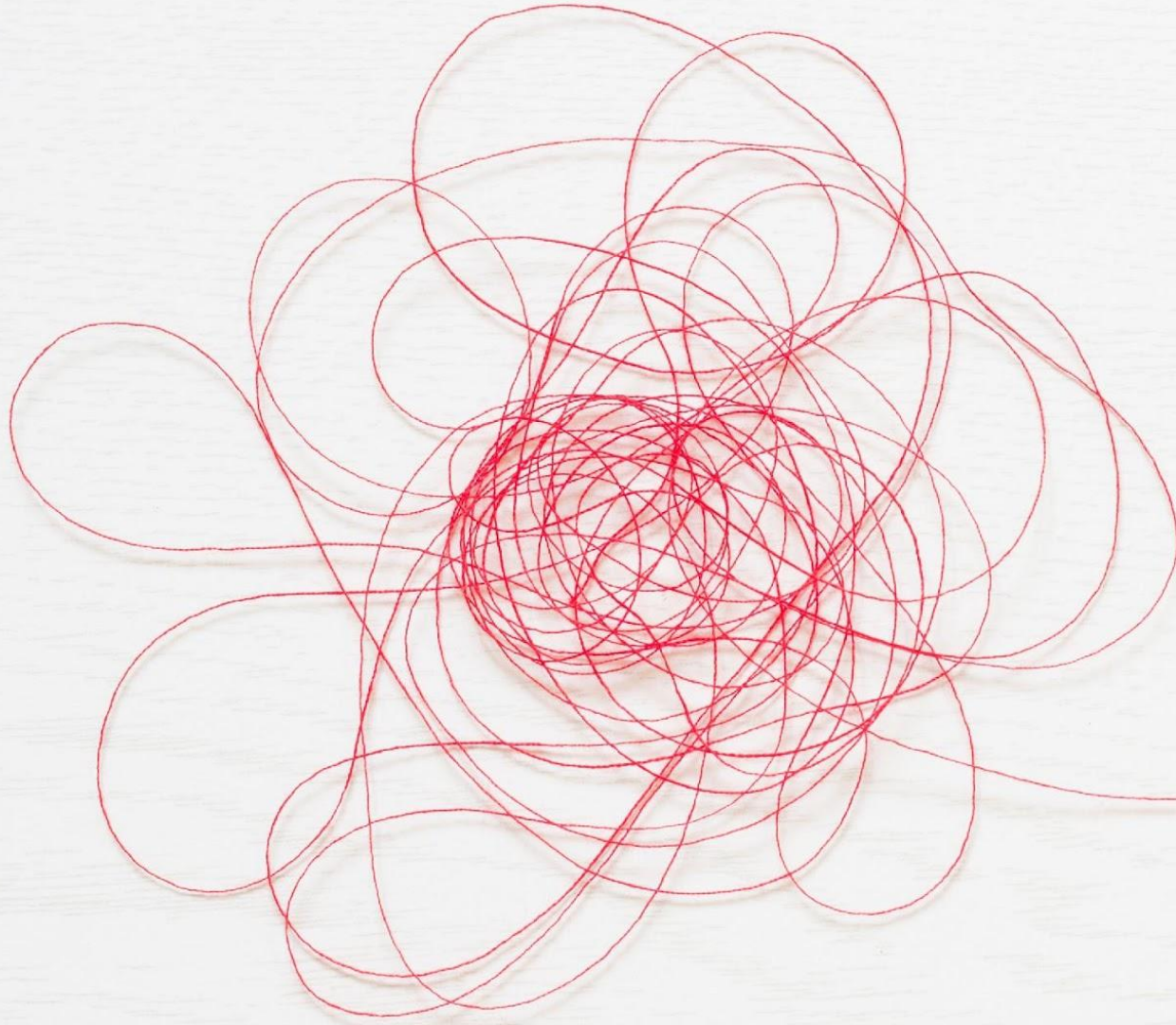
LANGUAGE



BIAS AND
REPRESENTATION

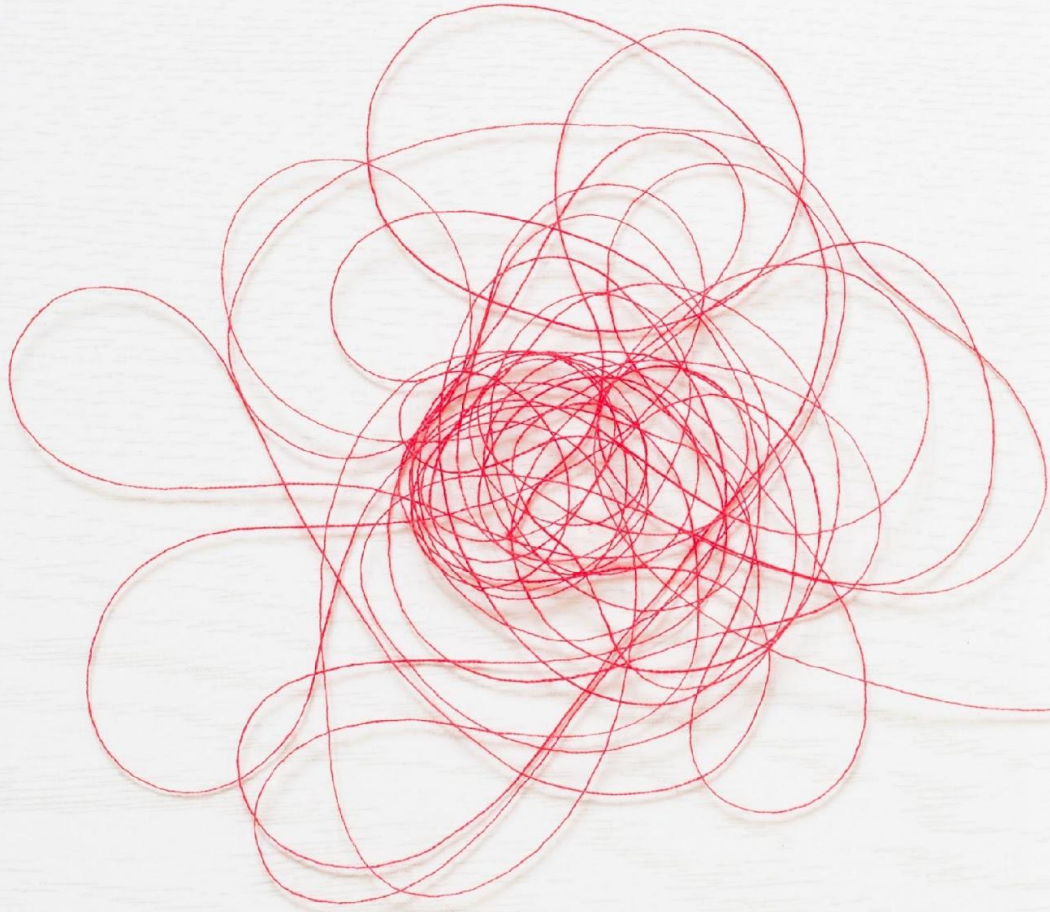


FAIRNESS



Problem Formulation

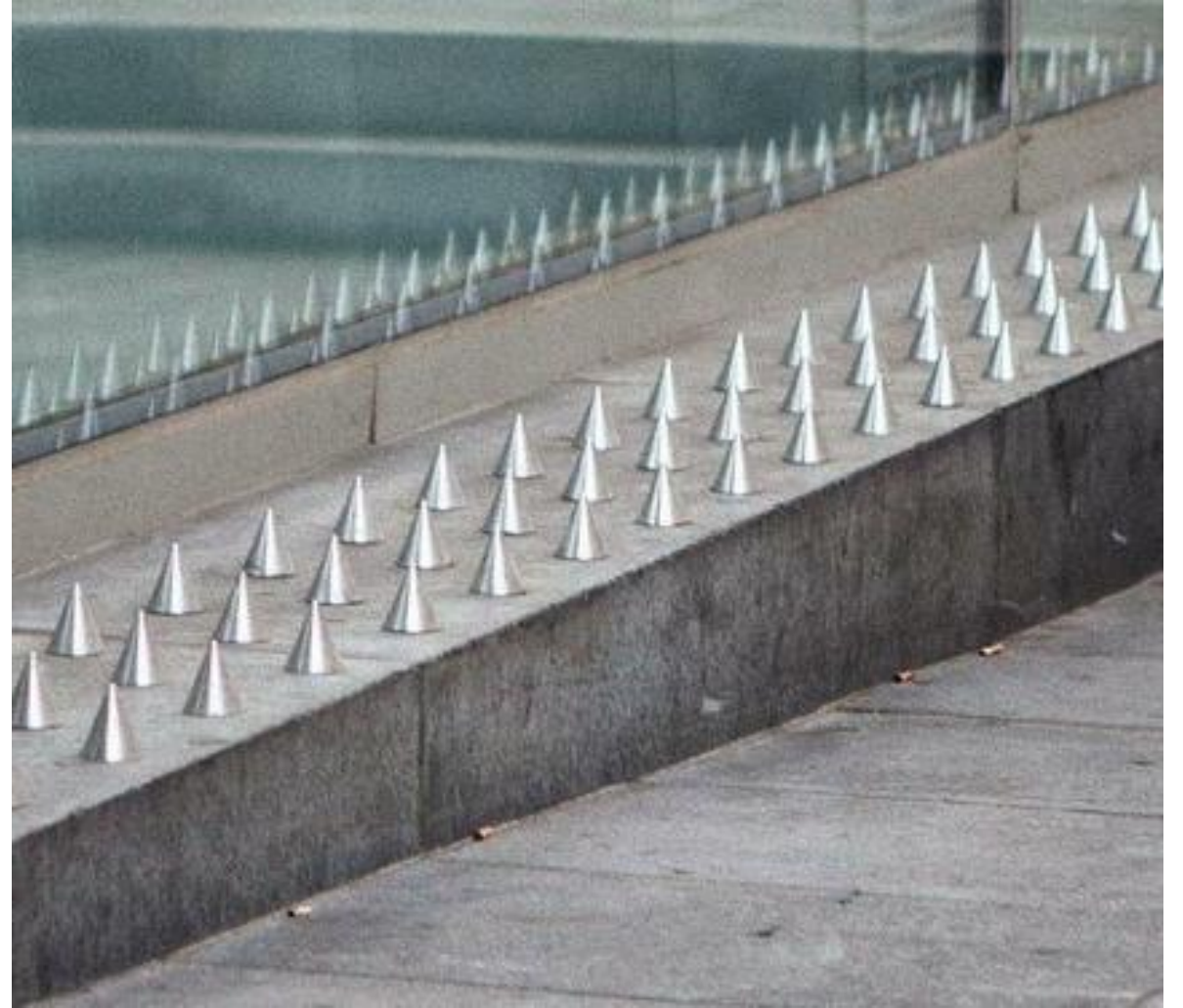
Problem Formulation Statements



- Formulating a problem means treating the **desired solution as good or worthy** of being done.
- **Why** should we care about solving this problem?
- **Who** can agree that this is a problem worth solving?
- Who would **benefit** from its solution?

Problem Formulation

“Homeless people are sleeping here and we want them to stop”



Problem Formulation

“Some people in our community don’t have a place to sleep and we think they should”



Problem Formulation



Who is included in each problem formulation?
Who can agree it's a problem?

Problem Formulation Statements

What is the problem to be solved?

- Is Professor X a good teacher?
- Do students think she is a good teacher?
- Do most students think she is a good teacher?



QUALITY
5.0

CS101 **AWESOME** May 21st, 2015

Attendance: **Not Mandatory** Grade: **A** Textbook: **Yes** Online Class: **Yes**

DIFFICULTY
1.0

This class was awesome. A beginner like me that has never done anything further than facebook on a computer, Professor was very clear and easy to listen to. I very much enjoyed the lectures and how easy it was to learn from such a great teacher. Thank you for all that you do

👍 2 🗑️ 1

Choice of Data

CS 106A Mid-Quarter Evaluation

 kcreel@stanford.edu (not shared) [Switch account](#) 

* Required

Instructor Feedback

These first few questions come from Nick Parlante and Juliette Woodrow. Your anonymous feedback will be reviewed by them, so please be specific and give suggestions for improvement.

What has Nick been doing so far that has worked well for you? What should he continue doing? *

Your answer

- What kind of data should inform our decisions?
- Where will it come from?
- Is it a reflection of what we want to measure?

Data Ethics: Choices and Values



PROBLEM
FORMULATION



LANGUAGE



BIAS AND
REPRESENTATION



FAIRNESS

Why should we care?

Who benefits?

Who's harmed?

What data?

The background of the slide is a dense, chaotic arrangement of white, three-dimensional wooden letters. The letters are scattered across the entire frame, creating a textured, abstract pattern. In the center, a white rectangular box with a thin border contains the word "Language" in a bold, black, sans-serif font. A horizontal line is positioned directly beneath the word.

Language

Descriptive terms? Evaluative terms?

QUALITY

5.0



CS101



AWESOME

May 21st, 2015

Attendance: **Not Mandatory** Grade: **A** Textbook: **Yes** Online Class: **Yes**

DIFFICULTY

1.0

This class was awesome. A beginner like me that has never done anything further than facebook on a computer, Professor Nick was very clear and easy to listen to. I very much enjoyed the lectures and how easy it was to learn from such a great teacher. Thank you for all that you do



Descriptive vs. Normative Language

Descriptive language

- Statements of fact
- What people did
- What happened



- “Lectures are 90-minutes long”
- “Assignments take more than two hours to finish”
- “Sections are mandatory”

QUALITY

5.0

AWESOME

May 21st, 2015

Attendance:
Not
Mandatory

Textbook
Required

atory Grade: A Class: Yes

DIFFICULTY

1.0

me. A beginner like I am can't do anything further than facebook
Professor Nick was very clear and easy to understand. I very much enjoyed the
lectures and how easy it was to learn from such a great teacher. Thank you for all that you do



Descriptive vs. Normative Language

Normative language:

- Evaluative statements
- Express the speaker's opinions/reactions
- How they think things should be



- “right”
- “wrong”
- “good”
- “bad”
- “should”
- “should not”

QUALITY

5.0

DIFFICULTY

1.0



CS101

May 21st, 2015

Attendance: [redacted] Grade: **A** Textbook: **Yes** Online Class: **Yes**

AWESOME

This class was like me that anything further than facebook
on a computer very clear very much enjoyed the
lectures and learn from students thank you for all that you do

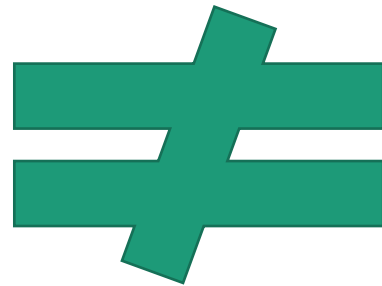
**GREAT
TEACHER**

👍 2 🗨️ 1



Normative language:

- Evaluative statements
- Express the speaker's opinions/reactions
- **How things should be**



Descriptive language

- Statements of fact
- What people did
- What happened
- **How things “are”**

QUALITY

5.0

DIFFICULTY

1.0



CS101

May 21st, 2015

Attendance: [redacted] Grade: **A** Textbook: **Yes** Online Class: **Yes**

CLEAR

This class was very clear and easy to understand. I like me that I don't need anything further than facebook on a computer. I very much enjoyed the lectures and I learned from them. Thank you for all that you do.

EASY TO LISTEN TO

👍 2 🗣️ 1



Thick Normative Terms

Descriptive AND normative:

- Thick normative terms express morally or aesthetically “loaded” descriptions



- Cowardly
- Cautious
- Polite
- Rude
- Chill
- Kind
- Caring
- Smart
- Knowledgeable
- **Professional**

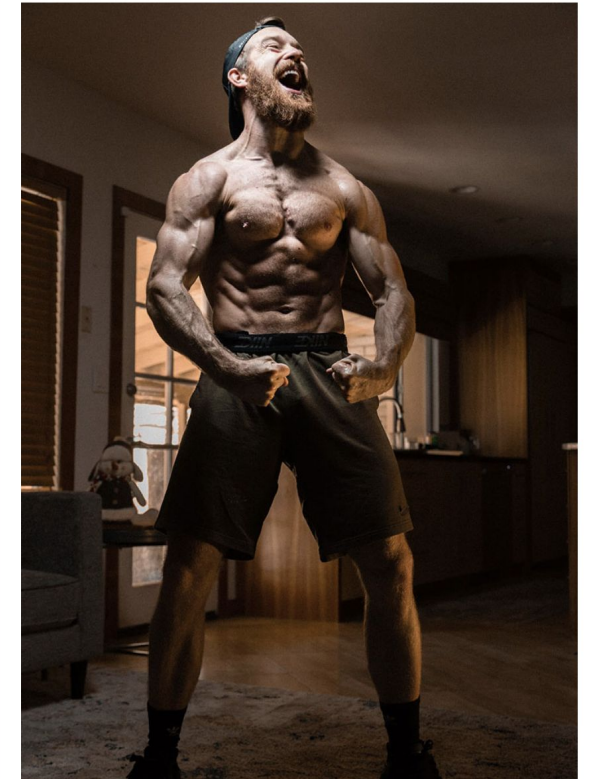
Ex: AI tools' "racy" score & gender bias

- goal: identify "raciness," or how sexually suggestive images are
- images of women more racy than men
- Use of thick normative term. Problematic?



Correct

Google AI	5 of 5
Microsoft AI	98%



Google AI	3 of 5
Microsoft AI	3%

Descriptive or Normative?



Does the program you are writing contain descriptive claims?



Do it contain normative claims or values?



How about thick normative terms?

Data Ethics: Choices and Values



PROBLEM
FORMULATION



LANGUAGE



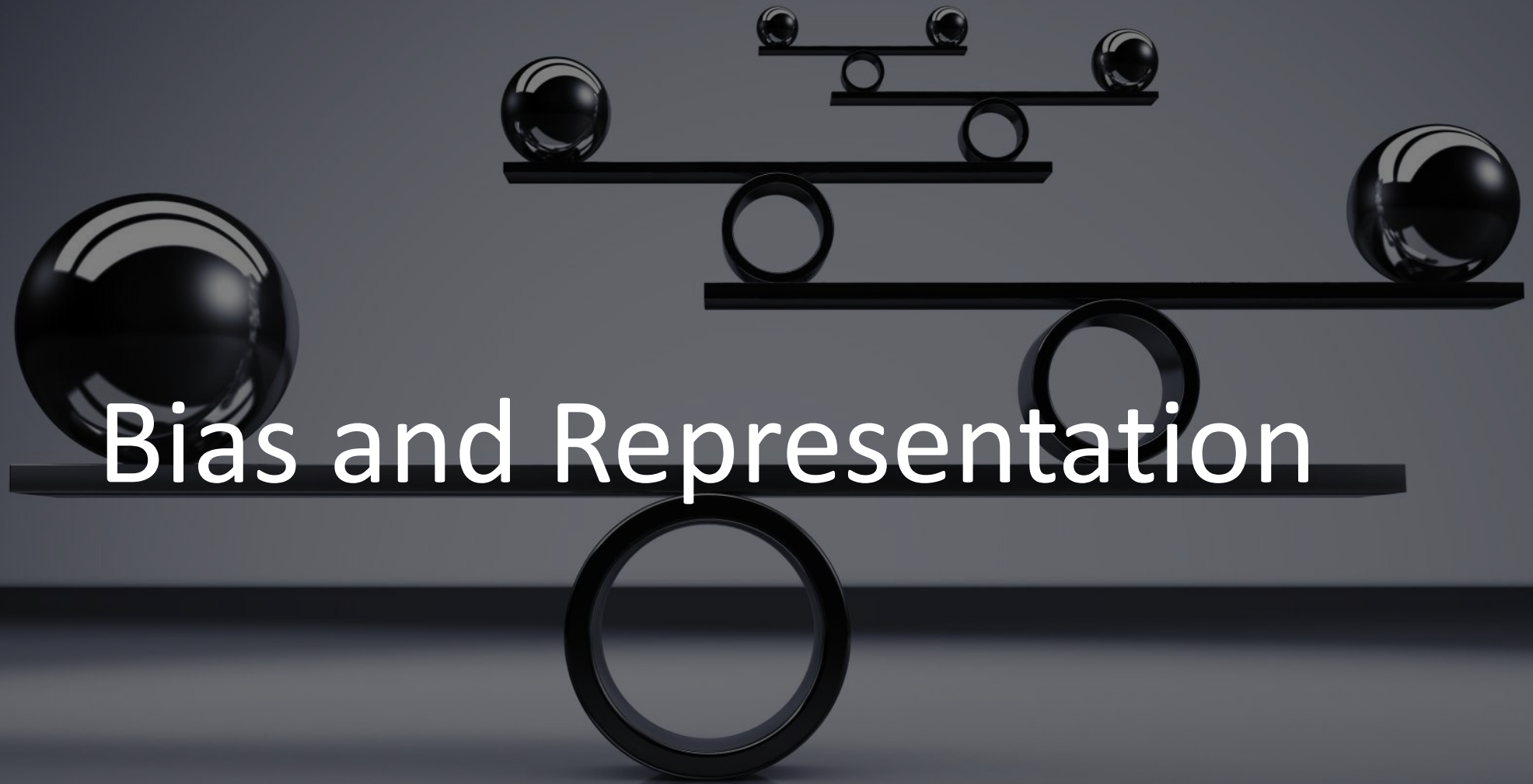
BIAS AND
REPRESENTATION



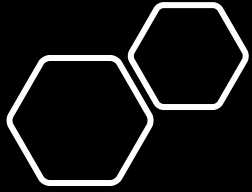
FAIRNESS

Why should we care?
Who benefits?
Who's harmed?
What data?

Descriptive: what is
Normative: what should be
Thick normative: both



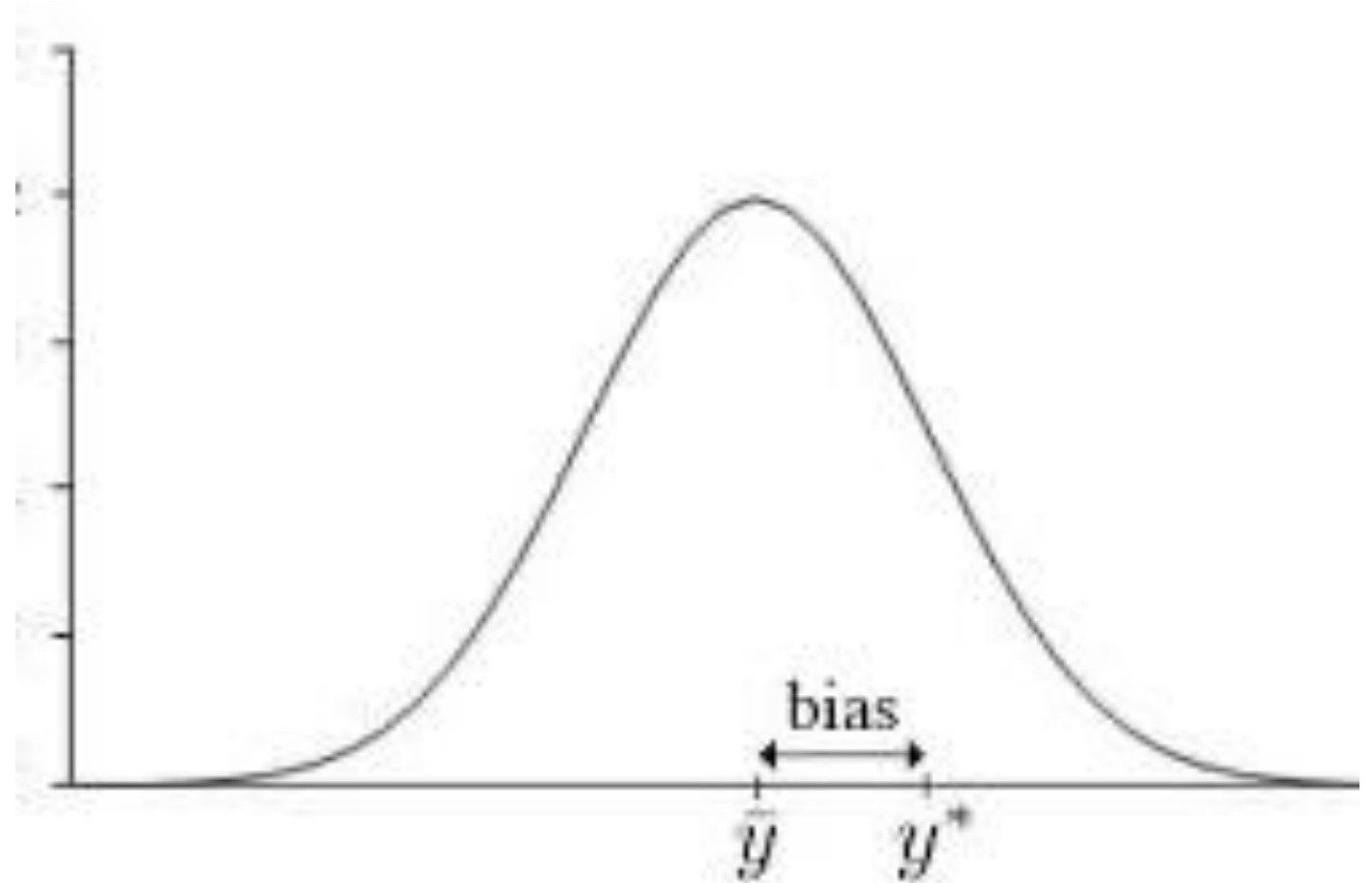
Bias and Representation



What is Bias?

Statistical bias is the difference between measured results and “true” value.

- This is the “neutral” or statistical meaning of the word bias. You will see it often in discussions of patterns in data.





Discrimination

- *Direct discrimination*: discrimination resulting from a negative attitude toward the social group (e.g. animus or indifference)
- *Indirect discrimination*: discrimination that does not result from such an attitude, but from rules and procedures constructed in a way that favors one group over another



Discrimination

”The rules and norms of society **consistently** produce **disproportionately** disadvantageous outcomes for the members of a certain group [and] the outcomes are **unjust** to the members of the disadvantaged group”

(Stanford Encyclopedia of Philosophy)



Discriminatory Bias in Data

Biased measurement or classification

+

Use of that bias that compounds existing
injustice

=

Discriminatory or Unfair Bias

Is this discriminatory bias?

- a. Ratings for Uber drivers were found to be lower for BIPOC drivers. Drivers with too low of ratings would be fired.
- b. Scores on a nursing licensing exam in the United Kingdom were statistically greater for women compared to men. Upon further review, it was found that women tended to perform better on questions about caring for a baby/ infant.

Biased measurement or classification
+
Use of that bias that compounds existing injustice
=
Discriminatory or Unfair Bias



Representation in Survey Data



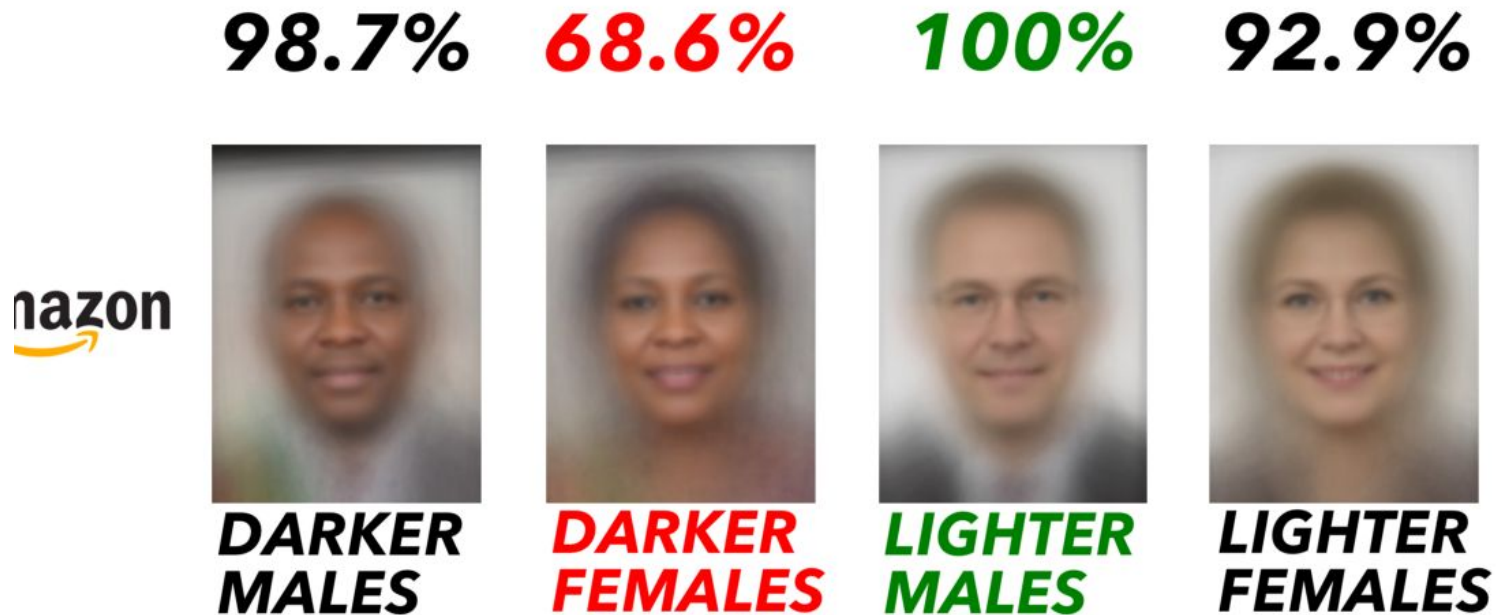
Representational Harms

A person is harmed when her identity is diminished in public representations of her social groups.

**Who is represented in this data?
Who can see themselves in it?**

Ex: Facial Recognition

August 2018 Accuracy on Facial Analysis Pilot Parliaments Benchmark



Amazon Rekognition Performance on Gender Classification

Distributional or Allocative Harm



Allocative Harms

A person is harmed when opportunities, resources, benefits, and protections that would otherwise be allocated to them are **unfairly** withheld.

ex: allocation of social services



Data Ethics: Choices and Values



PROBLEM FORMULATION

Why should we care?
Who benefits?
Who's harmed?
What data?



LANGUAGE

Descriptive: what is
Normative: what should be
Thick normative: both



BIAS AND REPRESENTATION

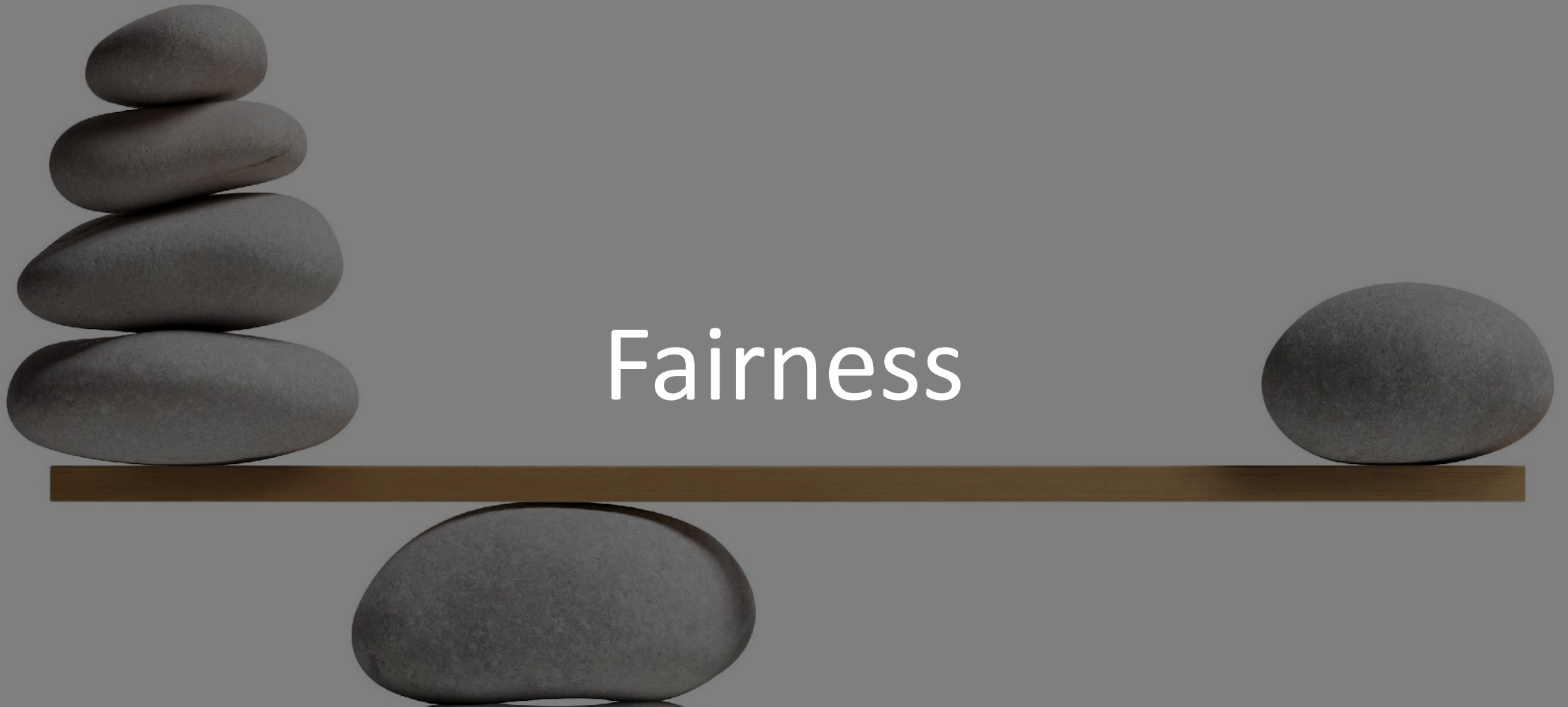
discriminatory bias:
measurement bias + bias
compounds injustice

representation & allocative
harms

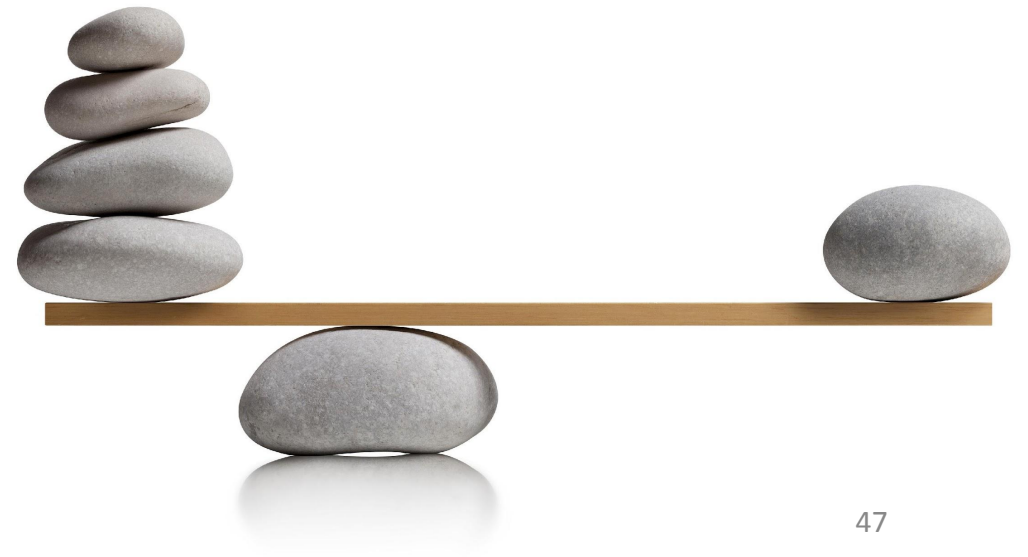


FAIRNESS

Fairness

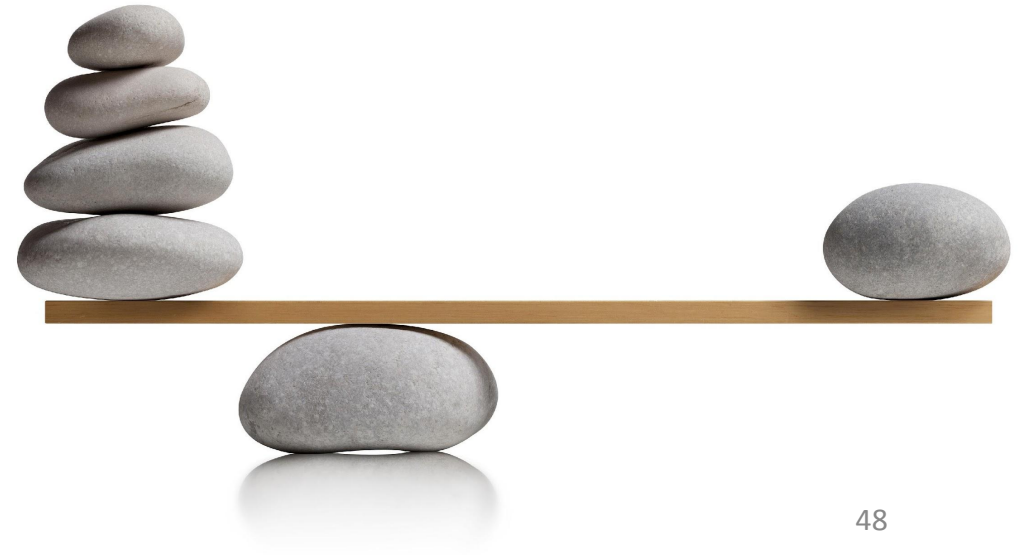


Principle: Distribution of goods should be based on **morally relevant characteristics**, not on morally arbitrary ones.



Formal Equality of Opportunity

- Positions that confer superior advantages should be open to all applicants.
- Applications are assessed on their relevant merits
- applicant deemed most qualified according to appropriate criteria is offered the position
- ex: Everyone has same opportunity to develop skills needed for the job, apply for the job, and get promoted.



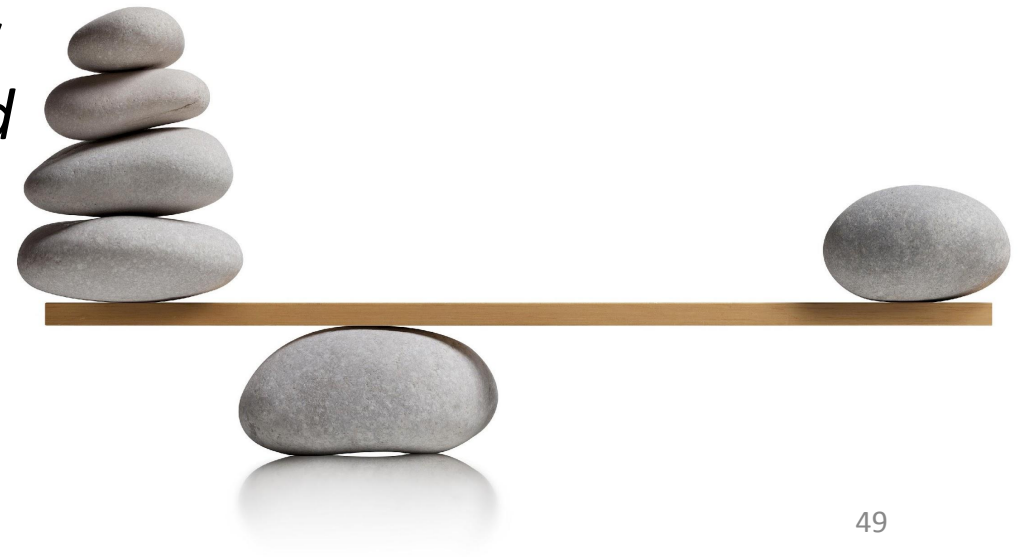
Substantive Equality of Opportunity

Takes into account systemic inequalities to ensure everyone in a community has access to the same opportunities and outcomes. Acknowledging that inequalities exist and works to eliminate them.

ex: Affirmative action

“Race-conscious, holistic selection processes are essential to achieve diversity in STEM programs at selective colleges and universities and to create a pipeline of diverse talent in STEM”

- amicus brief filed by Stanford, 2022

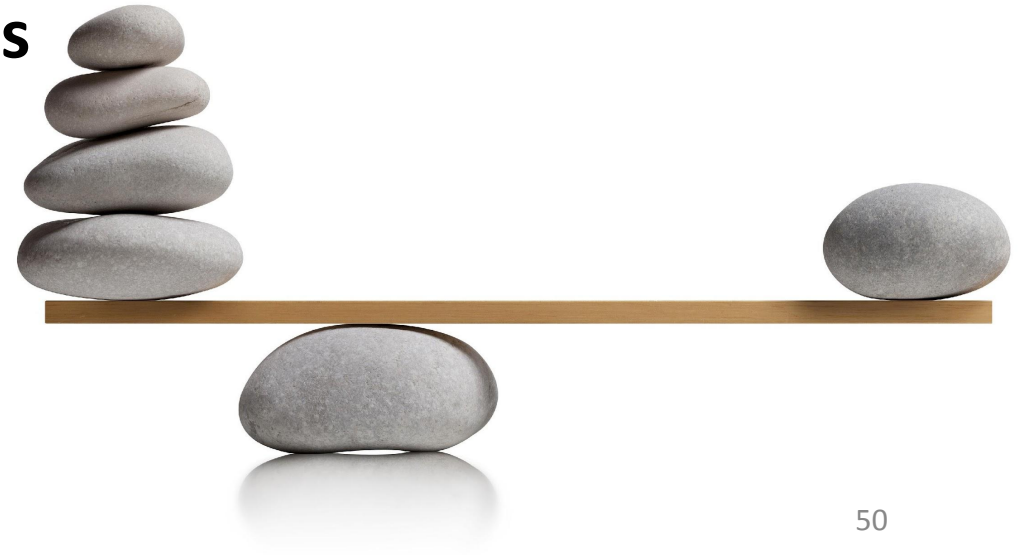


(Demographic) Parity

Because we are equal, we should adjust rules and procedures to ensure that outcomes reflect that.

Ex: People are equally likely to be a good teacher
=> expect numbers of highly rated teachers proportionate to population

VERY common metric of statistical fairness



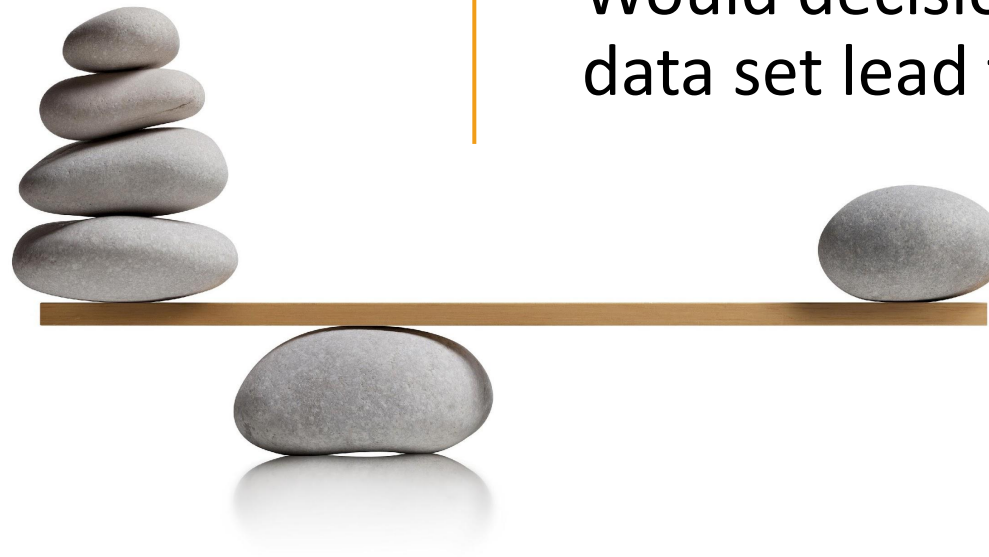
Two Sets of Questions to Ask

Values in Data Set

- What conception of fairness is encoded in the data set, if any?
- Does it lead to discrimination?

Values in data-based decisions

- Given existing biases in the data set, would it be fair to rely on them for our decisions?
- Would decisions based on the data set lead to discrimination?



Data Ethics: Choices and Values



PROBLEM FORMULATION

Why should we care?
Who benefits?
Who's harmed?
What data?



LANGUAGE

Descriptive: what is
Normative: what should be
Thick normative: both



BIAS AND REPRESENTATION

discriminatory bias:
measurement bias + bias
compounds injustice

representation & allocative
harms



FAIRNESS

**fairness: distribution based
on morally relevant
characteristics**

equality of opportunity

parity of outcomes

I have data about people! Now what?

Check for Statistical Bias

What correlations and patterns exist in my dataset?

In what ways do they fail to accurately represent the world?

Decide how to use the data given bias

- For what social purposes would it be appropriate to use this data?
- How should we communicate information about possible biases?

Check for Discriminatory Bias

In what ways do the biases compound existing injustice?

Generative AI: Auto-complete++

ChatGPT 3.5 ▾

You
CS106A is

ChatGPT
CS106A typically refers to an introductory computer science course taught at Stanford University. The course is designed for students with little to no programming experience and aims to teach fundamental concepts in computer science using the programming language Python. Topics covered in CS106A may include variables, loops, conditionals, functions, data structures, and basic algorithms.

It's worth noting that the content and structure of CS106A can vary between universities, and the information provided here is based on a general understanding of the course as of my last knowledge update in January 2022. If there have been any changes to the curriculum or if CS106A refers to a different course or program in your context, I recommend checking the most recent course materials or contacting the relevant educational institution for the latest information.

📄 🌟 🗨️ ↻


DALL-E ▾

Edit the detailed description

CS106A

Generate

RESULTS



The image shows a screenshot of the DALL-E interface. At the top, there's a header with the DALL-E logo and a dropdown menu set to 'DALL-E'. Below that is a text input field containing 'CS106A' and a 'Generate' button. Underneath the input field is a section labeled 'RESULTS' which displays four generated images in a 2x2 grid. The top-left image shows a hand holding a puzzle piece with the letters 'CS' on it. The top-right image shows a complex industrial piping system with blue and red pipes. The bottom-left image shows a wooden post with the text 'CS106A' in a snowy field. The bottom-right image shows a red and white striped tower structure.

Alignment problem



PROBLEM FORMULATION

Alignment problem:
AI doesn't align with
intended goals,
preferences, values

Write a Review *

Discuss the professor's professional abilities including teaching style and ability to convey the material clearly

! Guidelines ^

- Your rating could be removed if you use profanity or derogatory terms.
- Don't claim that the professor shows bias or favoritism for or against students.
- Don't forget to proof read!

[View all guidelines](#)

What do you want other students to know about this professor?

0/350

Homogenization of Language



LANGUAGE

homogenization:

biases writing towards
similar tone, language

Trained on form, not social
context

BE You

Tell me what to say to sound cool to college students



ChatGPT

To sound cool to college students, it's essential to strike a balance between being genuine, relatable, and incorporating some elements of current trends and culture. Here are some conversation starters and phrases that might help:

1. Start with a Casual Greeting:

- "Hey, what's up?"
- "What's good?"
- "How's it going?"

2. Use Slang Appropriately:

- "That's lit!"
- "Totally vibing with that."
- "No cap, it was an awesome weekend."

Reproducing social biases



BIAS AND REPRESENTATION

reproduces social biases:
(e.g. stereotypes)

often aligns most with
dominant US culture

Choose a first adjective (or leave this blank!)

intellectual

Choose a first group

teacher

intellectual teacher



Choose a second adjective (or leave this blank!)

pleasant

Choose a second group

teacher

pleasant teacher



Fairness and AI



FAIRNESS

(in)equality of opportunity:
poor performance w/
dialects, disabilities

(non-)parity of outcomes:
disparate risk of harmful
content



challenging cisgender bias in AI (J Rosenbaum)



PROBLEM FORMULATION

Write a Review *

Discuss the professor's professional abilities including teaching style and ability to convey the material clearly

Guidelines

- Your rating could be removed if you use profanity or derogatory terms.
- Don't claim that the professor shows bias or favoritism for or against students.
- Don't forget to proof read!

[View all guidelines](#)

What do you want other students to know about this professor?

0/350



LANGUAGE

You

Tell me what to say to sound cool to college students

ChatGPT

To sound cool to college students, it's essential to strike a balance between being genuine, relatable, and incorporating some elements of current trends and culture. Here are some conversation starters and phrases that might help:

- Start with a Casual Greeting:**
 - "Hey, what's up?"
 - "What's good?"
 - "How's it going?"
- Use Slang Appropriately:**
 - "That's lit!"
 - "Totally vibing with that."
 - "No cap, it was an awesome weekend."



BIAS AND REPRESENTATION



intellectual vs pleasant teachers (DALLE 2)



FAIRNESS



challenging cisgender bias in AI (J Rosenbaum)

What about AI?



PROBLEM FORMULATION

Alignment problem:
AI doesn't align with intended goals, preferences, values



LANGUAGE

homogenization:
biases writing towards similar tone, language

Trained on form, not social context



BIAS AND REPRESENTATION

reproduces social biases:
(e.g. stereotypes)

often aligns most with dominant US culture



FAIRNESS

(in)equality of opportunity:
poor performance w/ dialects, disabilities

(non-)parity of outcomes:
disparate risk of harmful content

Ethics goals for CS106a

- a. Image manipulation should not compromise people's **autonomy**
- b. End to end encryption addresses some **privacy** considerations
- c. When using data, our choices reflect direct and collateral **values**

CS182: Ethics, Public Policy, and Technological Change

Central Themes

- The impacts of technology are not fixed. They reflect a set of “design” choices. Those design choices encode a set of values.
- The impacts also reflect choices about what policies and regulations society chooses to put in place.
- When competing values are at stake, they must be weighed against one another. Who weighs these values and how? This is a critical question of governance, politics, and power.
- You are a central participant in this drama. Understanding your role(s) and exploring/debating the values you want to see encoded are a modern form of civic duty.



Taught by Prof. Mehran Sahami

Tech Ethics Center

- talks & events
- minor in ethics & technology
- grants for UGs

Ethics, Society, and Technology (EST) Hub

- paid summer fellowships
- research opportunities

[https://ethicsinsociety.stanford.edu/
undergraduate-offerings-overview](https://ethicsinsociety.stanford.edu/undergraduate-offerings-overview)

Human-Centered AI Inst.

- talks & events
- student affinity groups
- grants for UGs

<https://hai.stanford.edu/>

university-wide effort to make community service an essential feature of Stanford education

- courses
- small grants to students, students orgs
- service opportunities (quarter or longer)
- (paid) summer opportunities, fellowships
- career opportunities

<https://haas.stanford.edu/>

Data Ethics: Choices and Values

Benjamin Xie, Ph.D. (benjixie@stanford.edu)



PROBLEM FORMULATION

Why should we care?
Who benefits?
Who's harmed?
What data?



LANGUAGE

Descriptive: what is
Normative: what should be
Thick normative: both



BIAS AND REPRESENTATION

discriminatory bias:
measurement bias + bias
compounds injustice

representation & allocative
harms



FAIRNESS

fairness: distribution based
on morally relevant
characteristics

equality of opportunity

parity of outcomes

Thank you!

