Data Ethics: Choices and Values

Benjamin Xie, Ph.D.

McCoy Family Center for Ethics in Society - HAI

original slides & content by Kathleen Creel, Diana Acosta-Navas

We use data to inform our decisions

- Evidence-based
- Impartial
- Reliable

QUALITY

5.0

DIFFICULTY

1.0

CS101 😎 **AWESOME**

May 21st, 2015

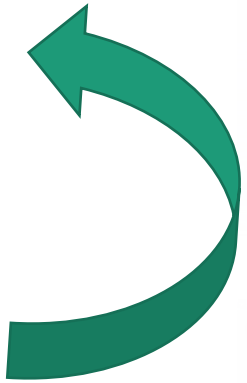
Attendance: **Not Mandatory** Grade: **A** Textbook: **Yes** Online Class: **Yes**

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

2 1

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 find ethically relevant elements?

VALUES IN DESIGN

PROBLEM FORMULATION

LANGUAGE

BIAS AND REPRESENTATION

FAIRNESS

Values in Design

Values in Design

- Design decisions encode values.
- They are expressive of what we care about
- They reveal our assumptions about the world, the people who will be interacting with our design, and benefiting from it

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

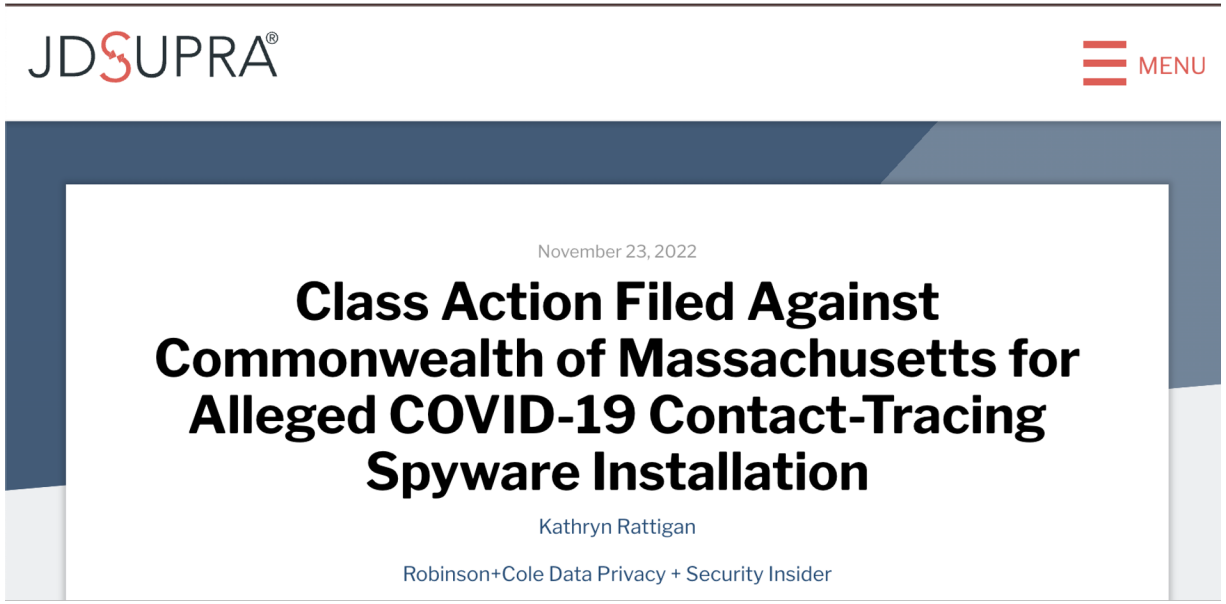
Explicit Values

Contact-tracing

Health
Safety
Efficiency
Public interest

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?



Explicit Values

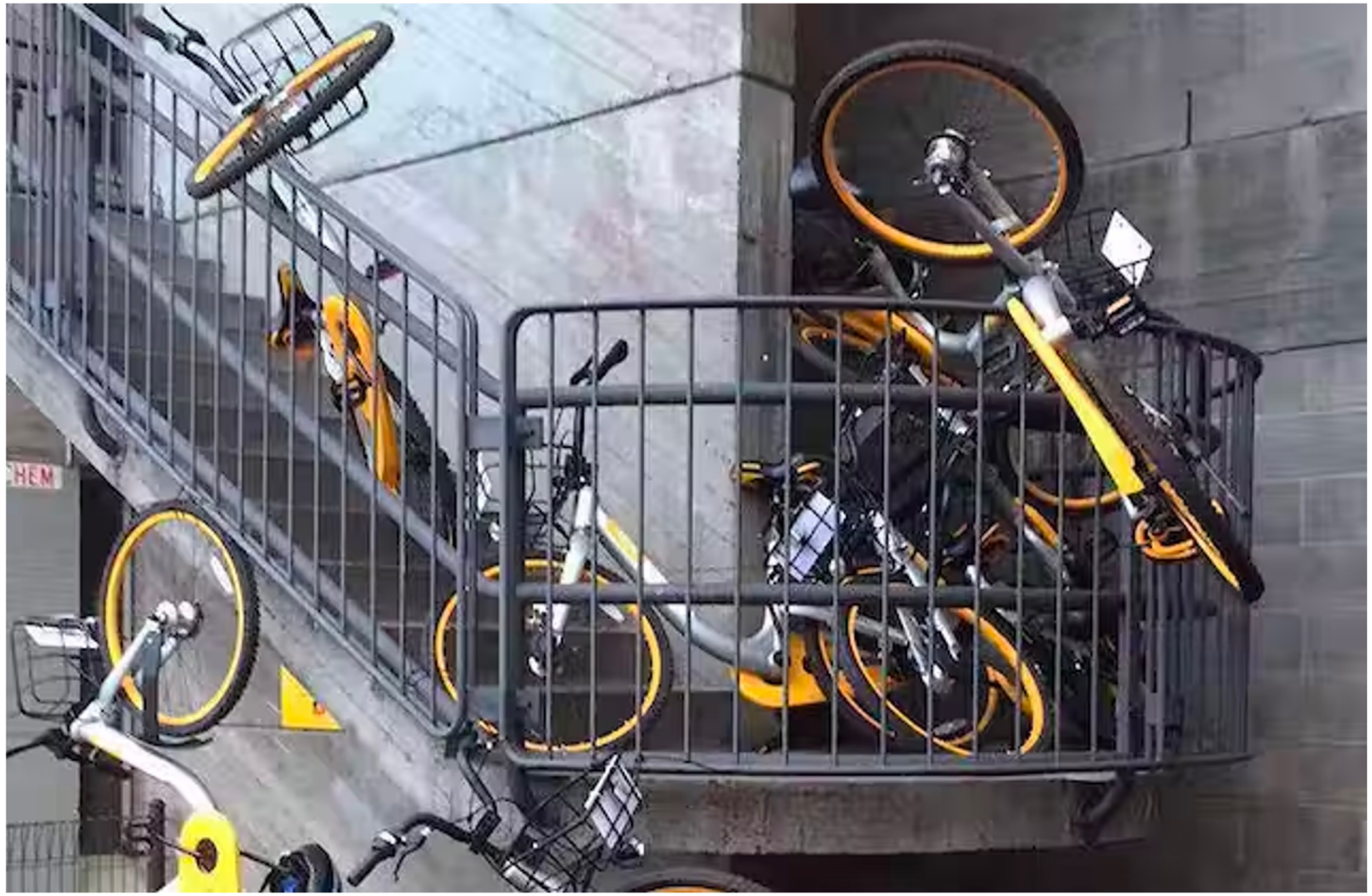
Bike-sharing app

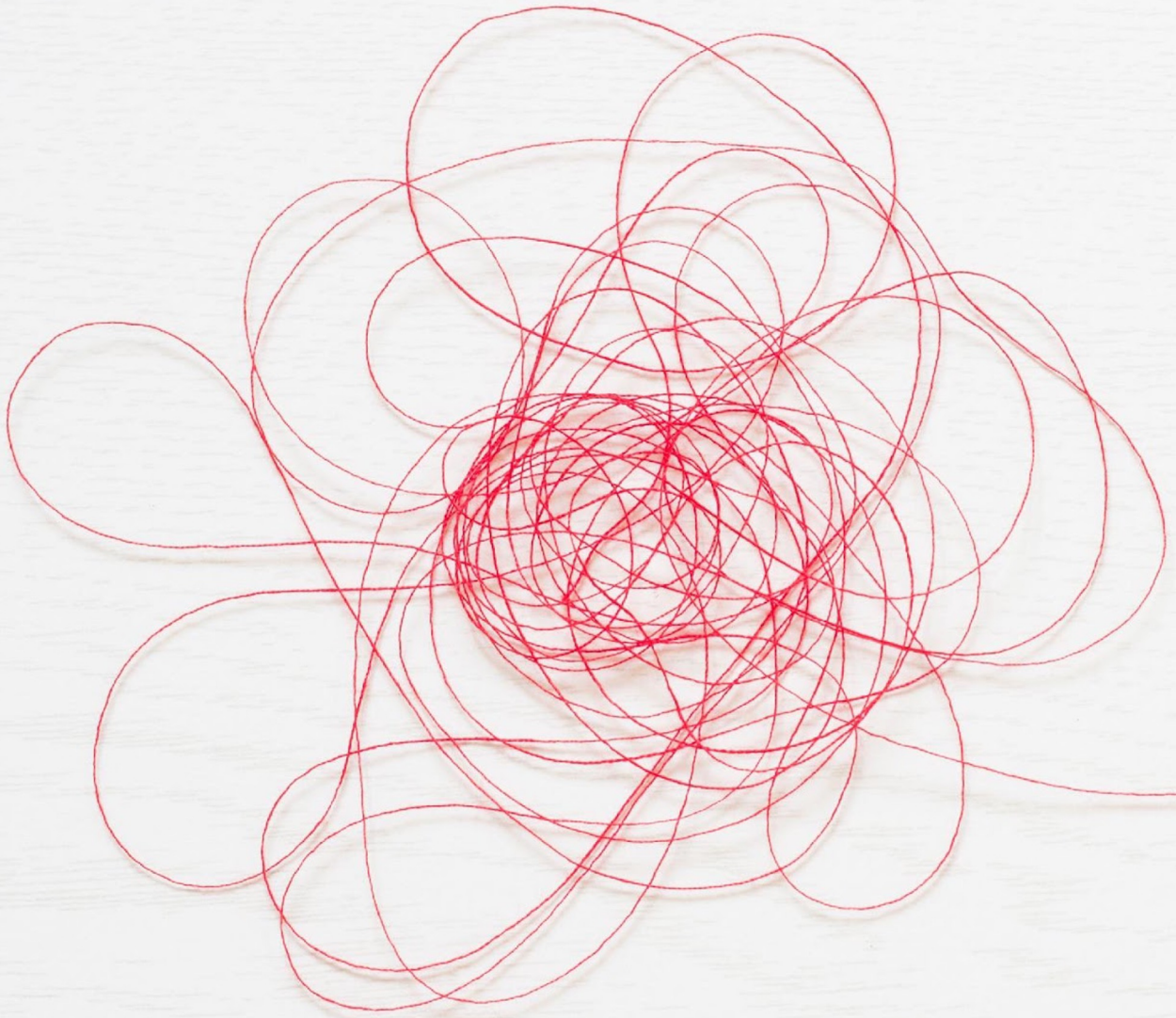
Mobility
Health
Sustainability
Inclusion

Collateral Values

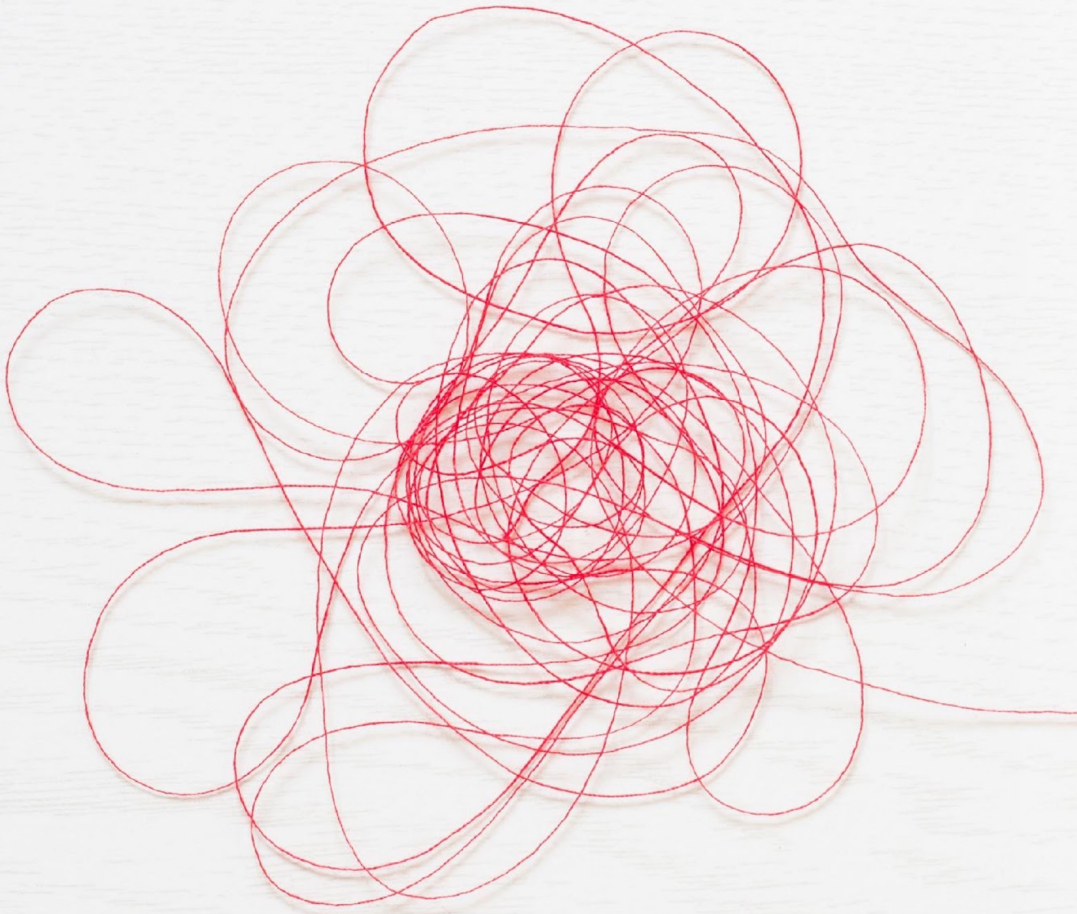
Bike-sharing

Who is the default user?
Impact on other mobility?





Problem Formulation

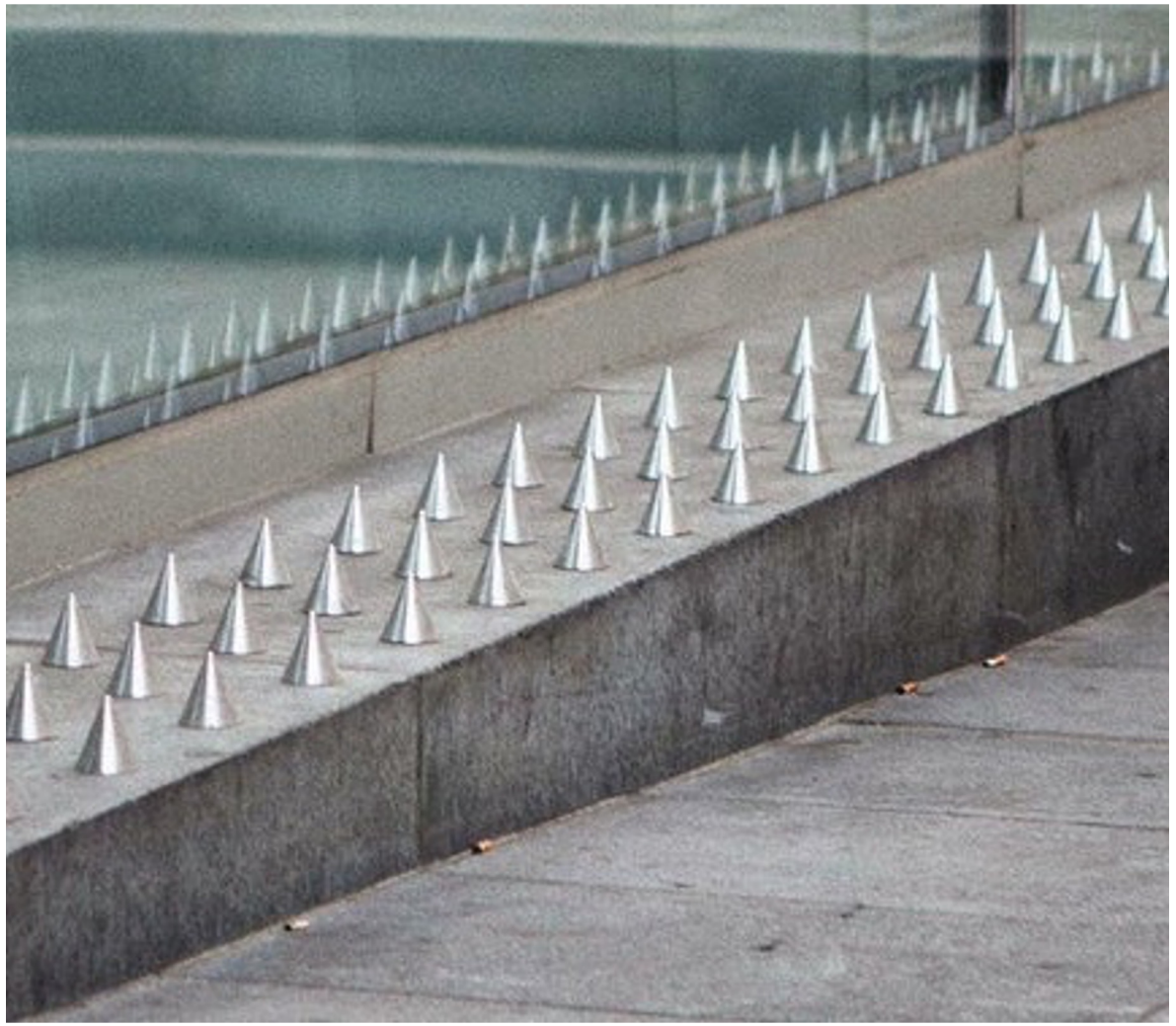


Problem Formulation Statements

- 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

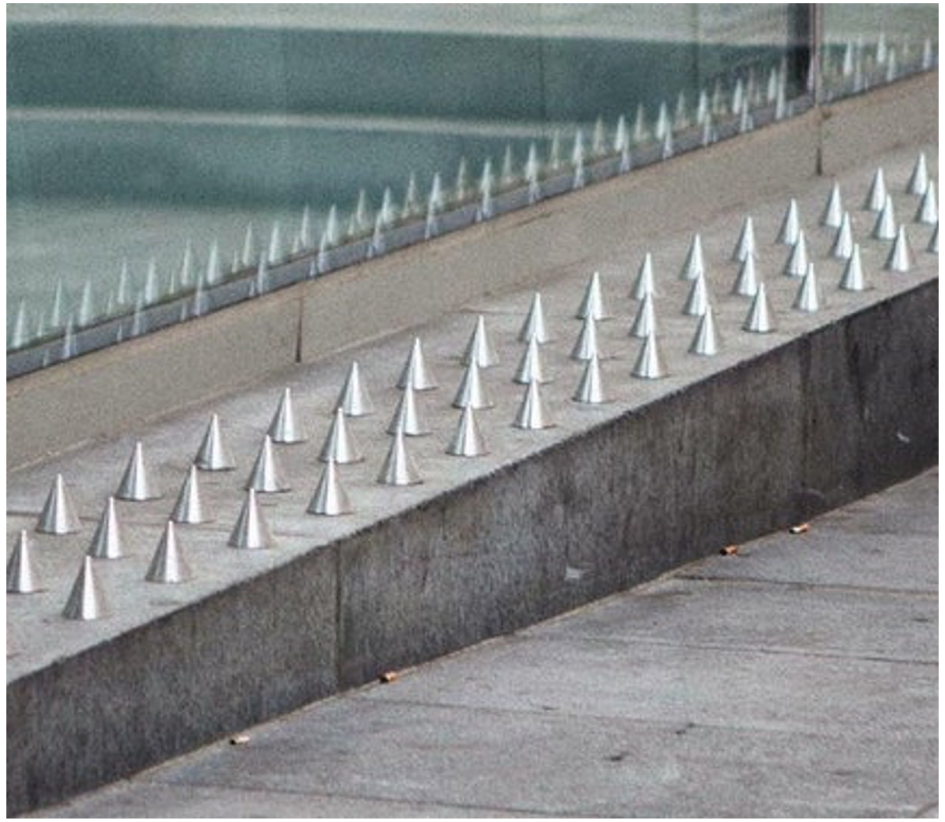
“Homeless people are sleeping here and *we* want them to stop”



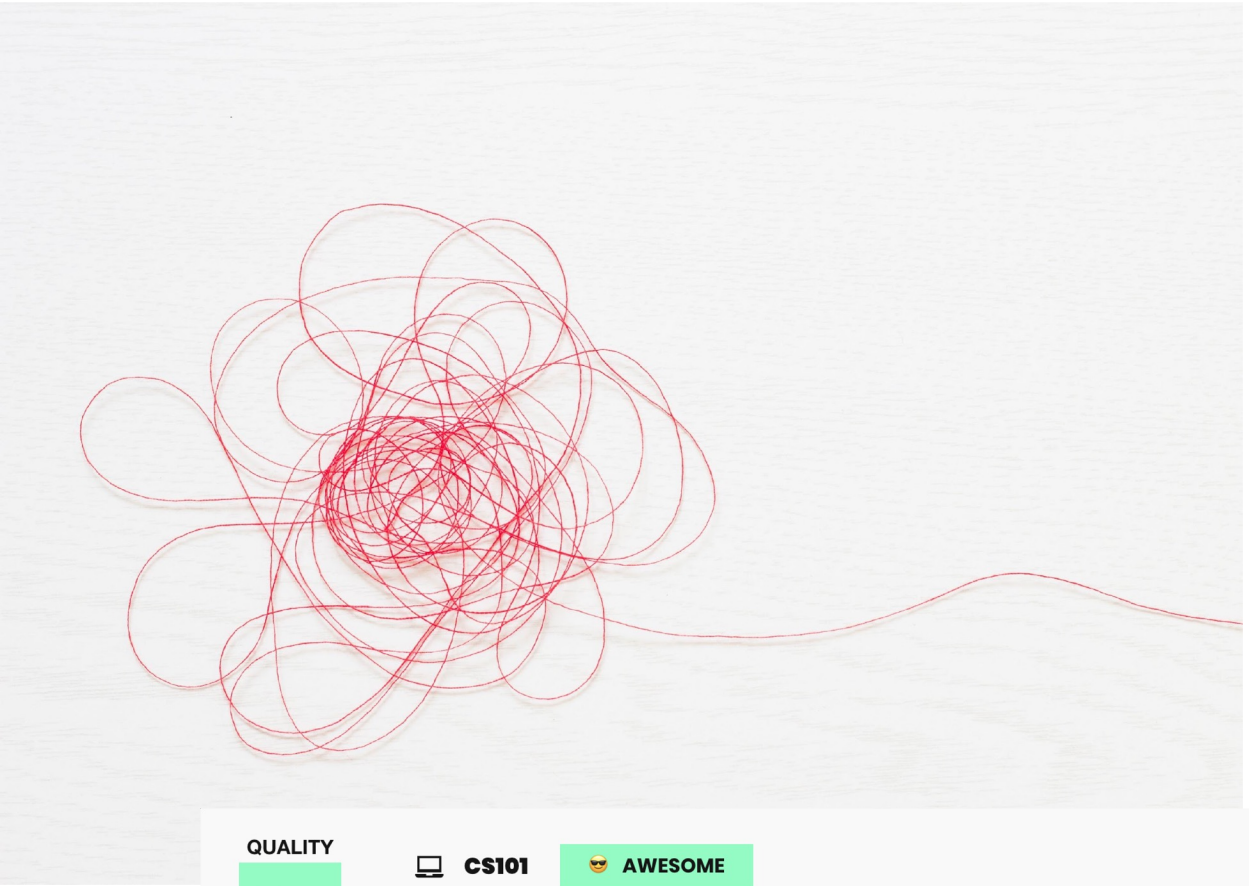
Problem Formulation

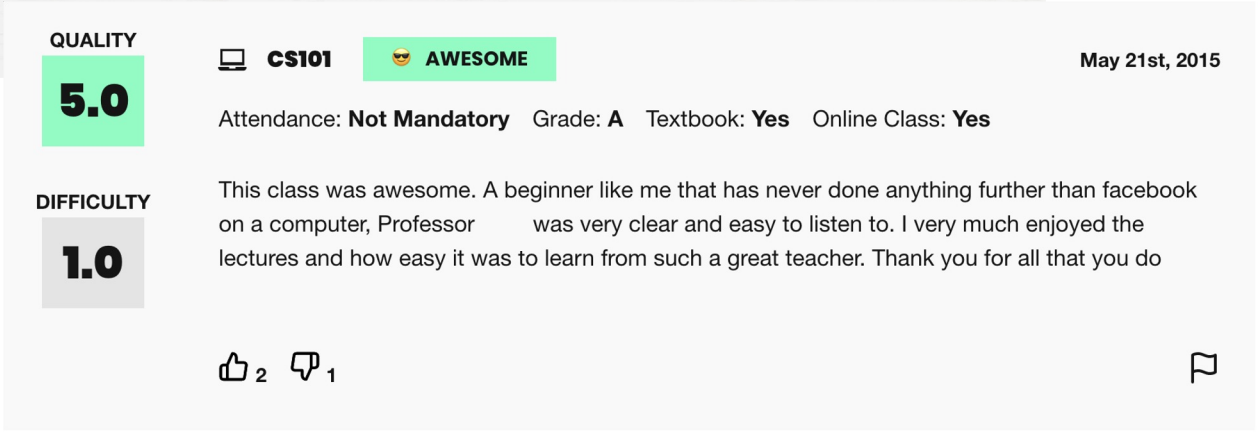
"Some people in our community don't have a place to sleep and *we* think they should"

Problem Formulation



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



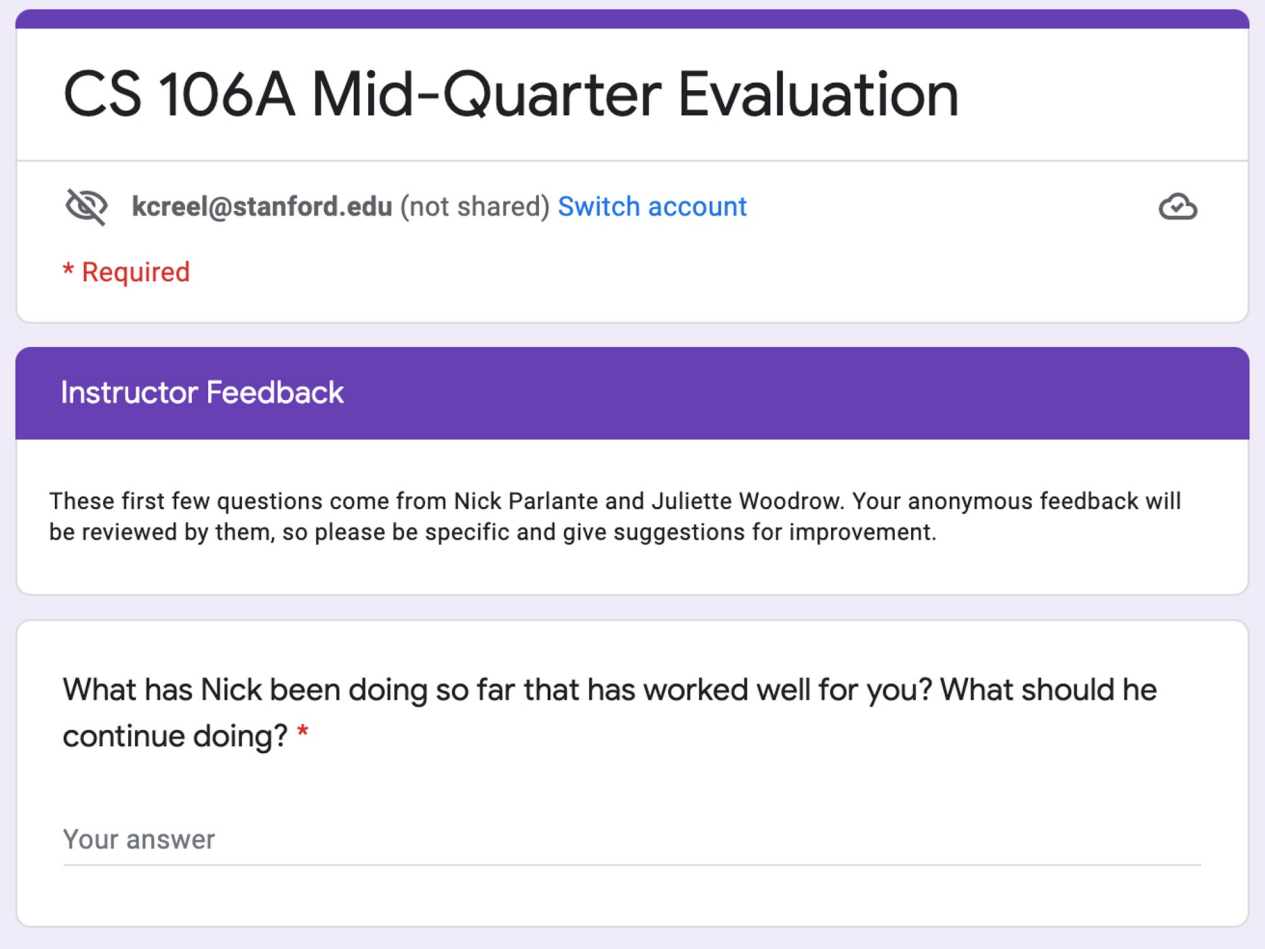


Problem Formulation Statements

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?

Choice of Data



- 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?

Language

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

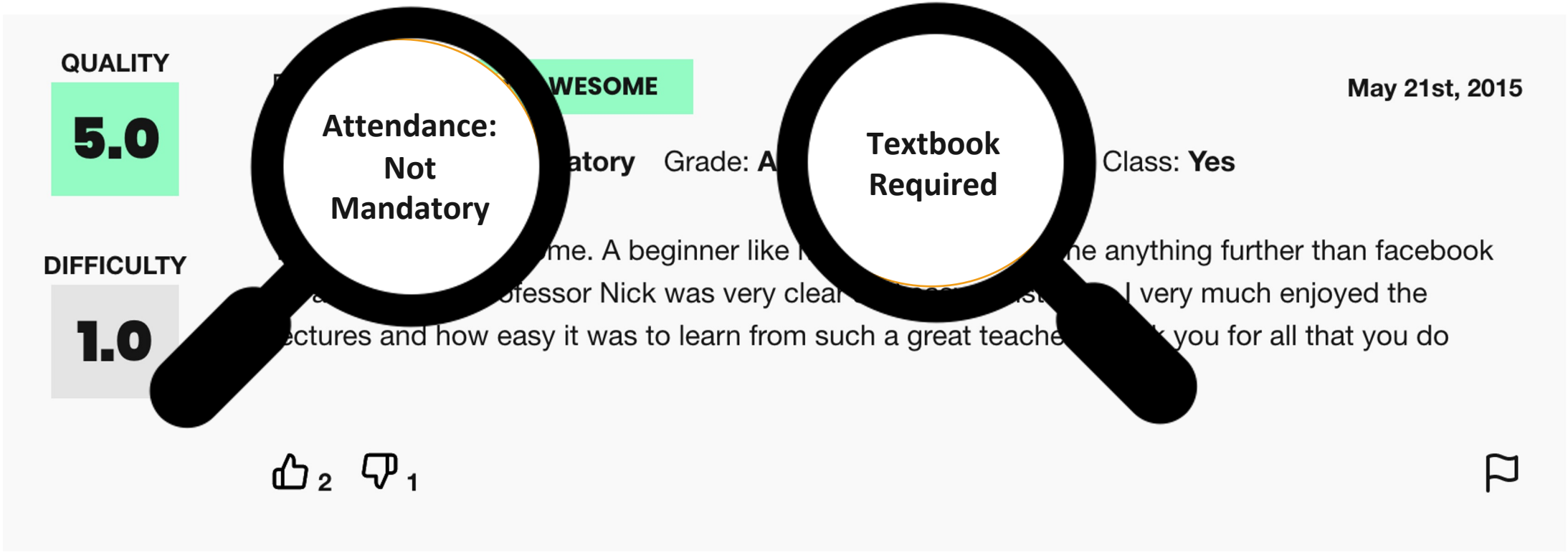
2 1

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
DIFFICULTY
1.0
WESOME
May 21st, 2015
atory Grade: A
Class: Yes
me. A beginner like
ne anything further than facebook
ofessor Nick was very clear
very much enjoyed the
ectures and how easy it was to learn from such a great teache
you for all that you do
2
1
Attendance: Not Mandatory
Textbook Required

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: e: **A** Textbook: **Yes** Online Class: **Yes**

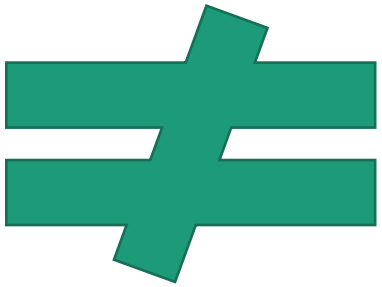
This class w like me tha anything further than facebook on a compute very clear very much enjoyed the lectures and l earn from su k you for all that you do

2 1

Descriptive vs. Normative Language

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: e: **A** Textbook: **Yes** Online Class: **Yes**

This class w like me tha anything further than facebook on a compute very clear very much enjoyed the lectures and l earn from s nk you for all that you do

2 1

CLEAR

EASY TO LISTEN TO

Thick Normative Terms

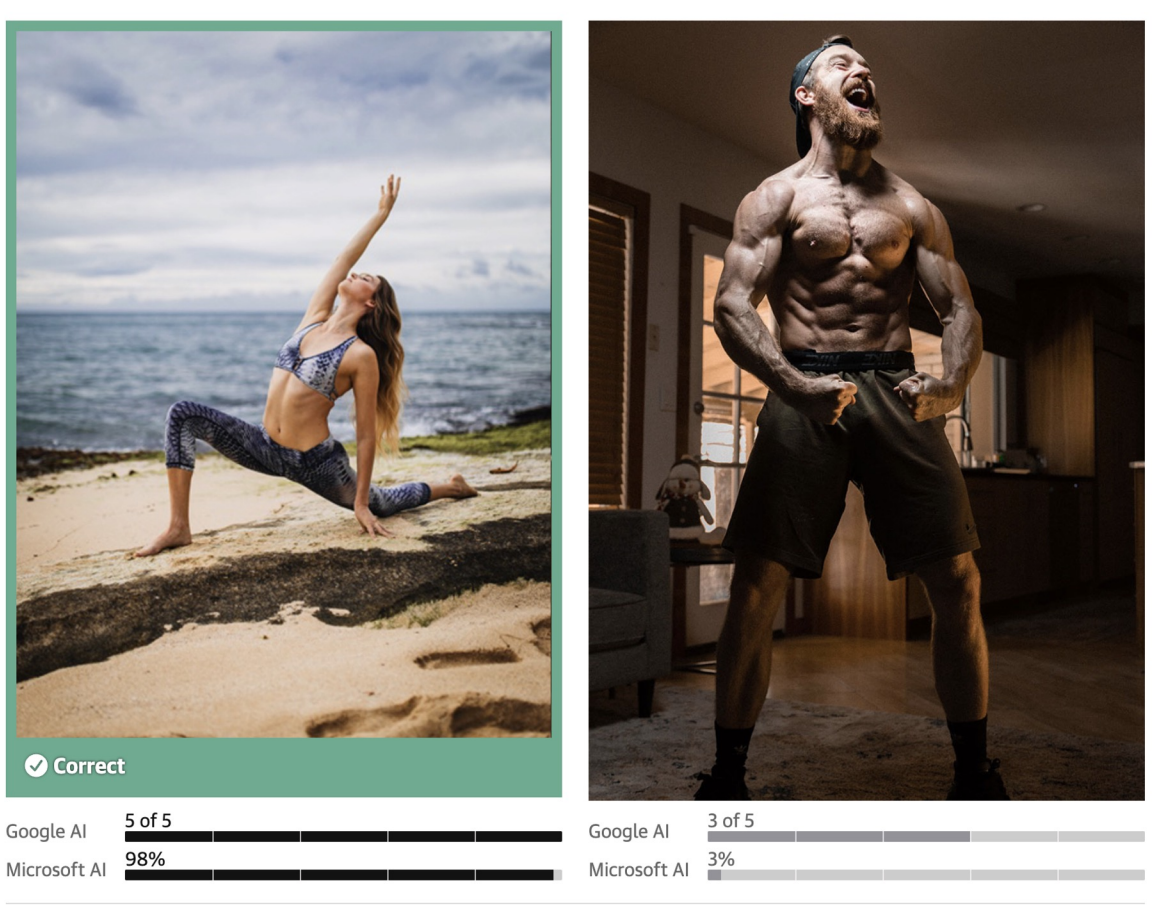
Descriptive AND normative:

- Thick normative terms express morally or aesthetically "loaded" descriptions

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

Ex: AI tools' "racy" score

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



Guardian. Feb 2023

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

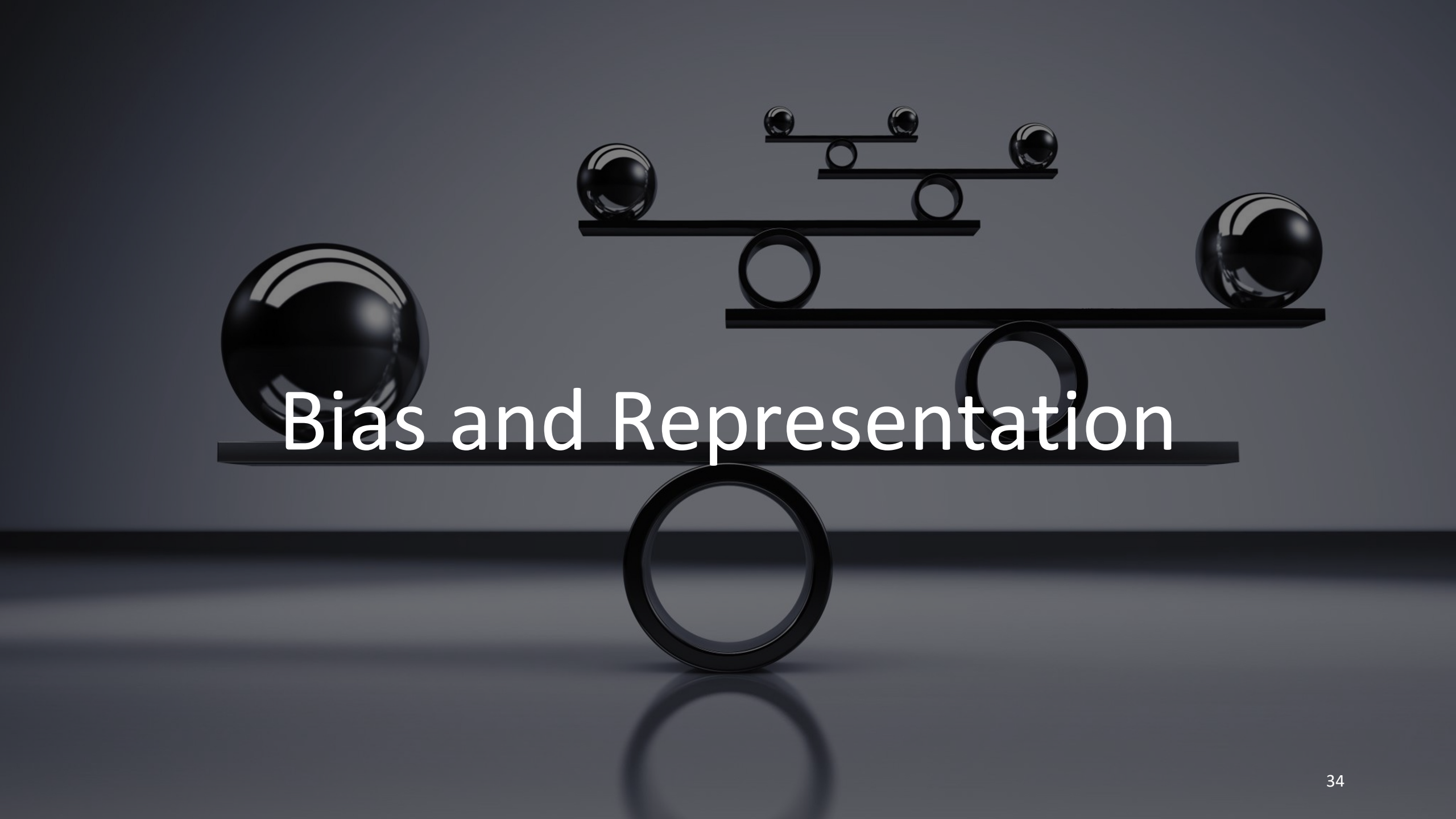
LANGUAGE

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

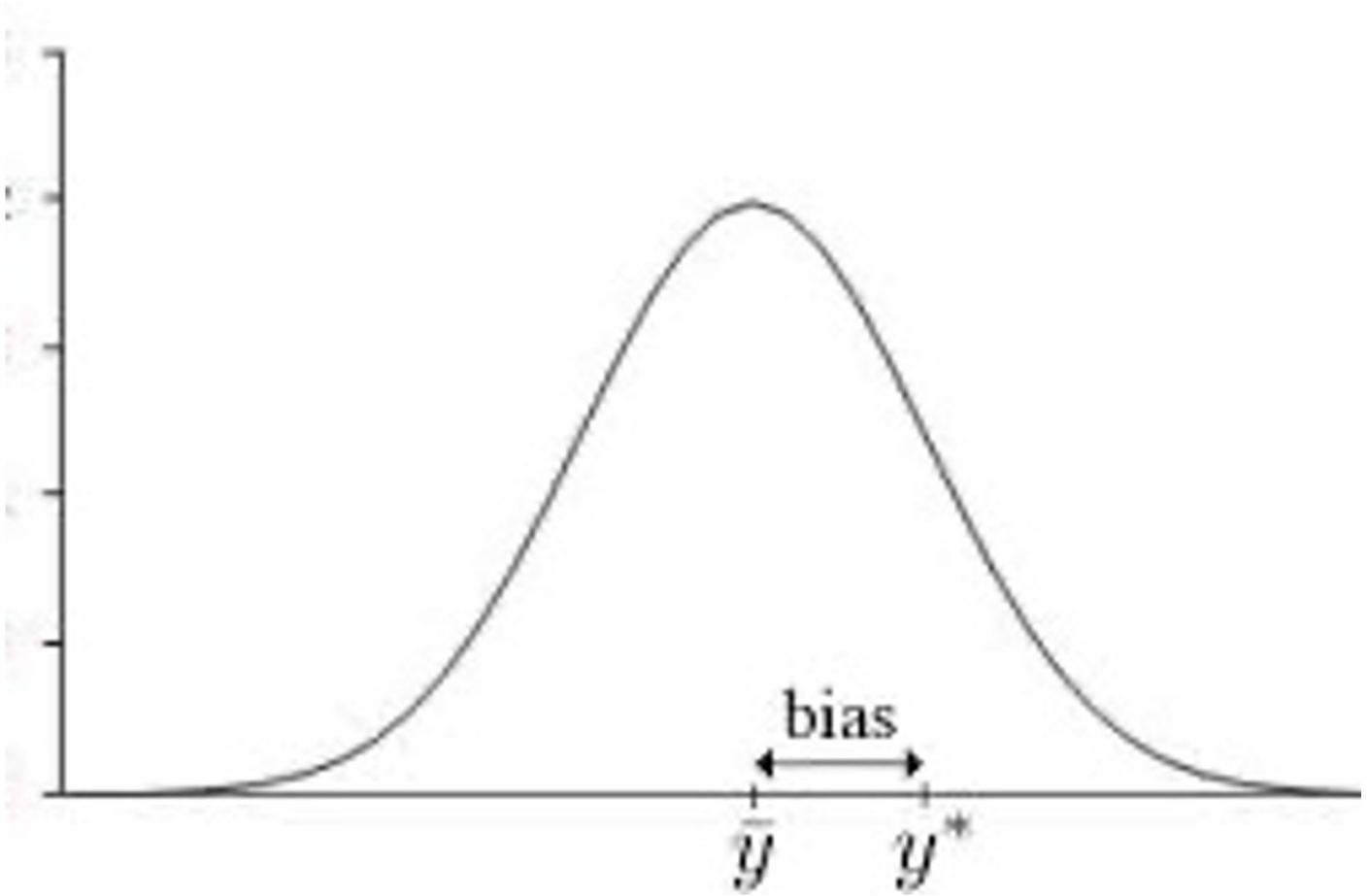


Bias and Representation

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

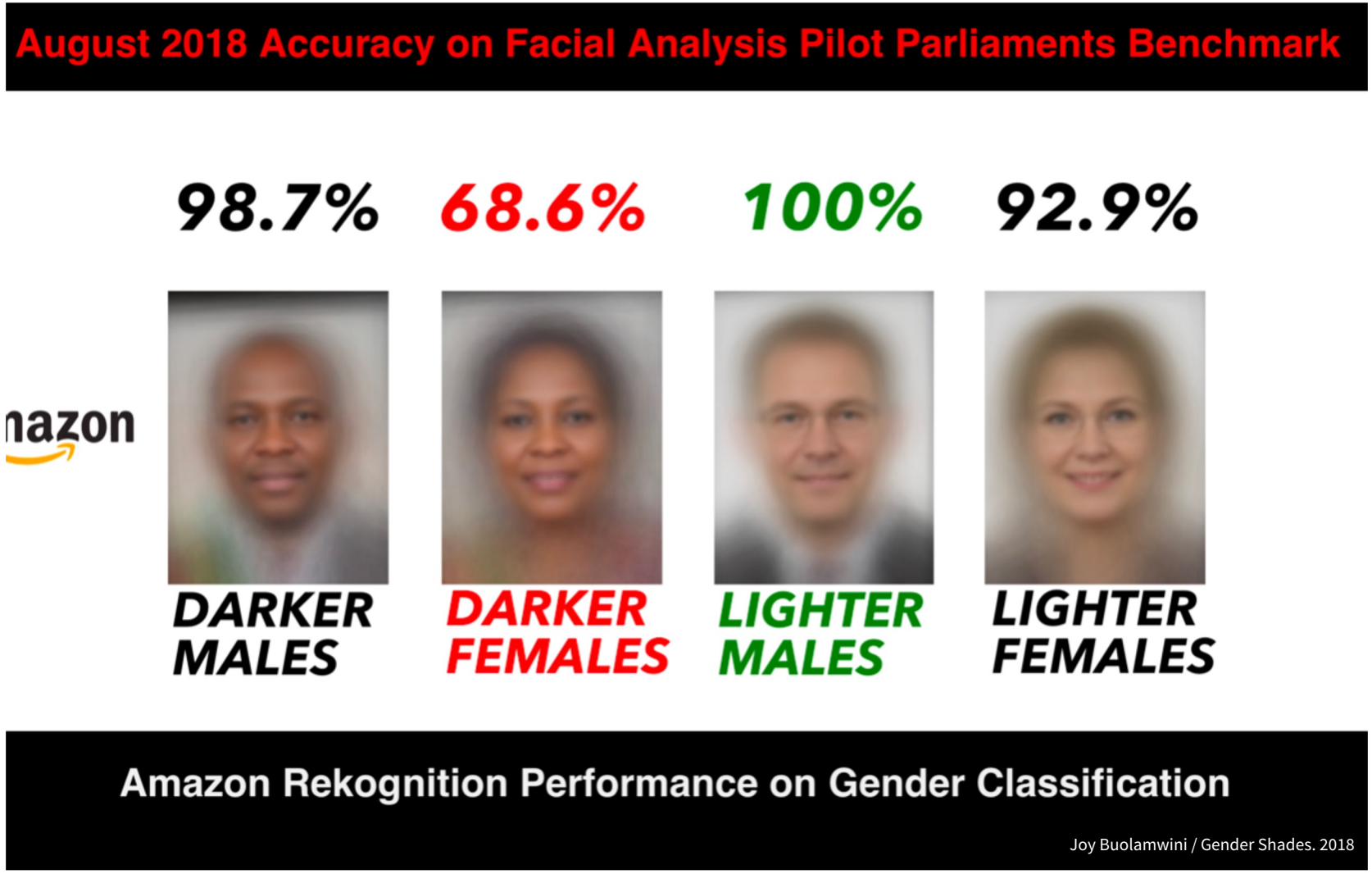
Representational Harms

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

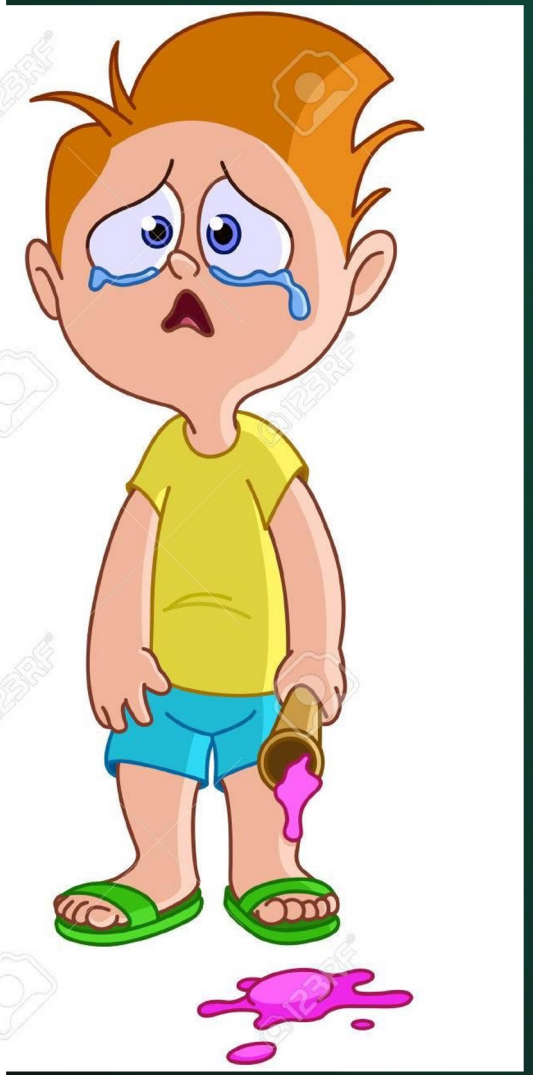


Distributional or Allocative Harm

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

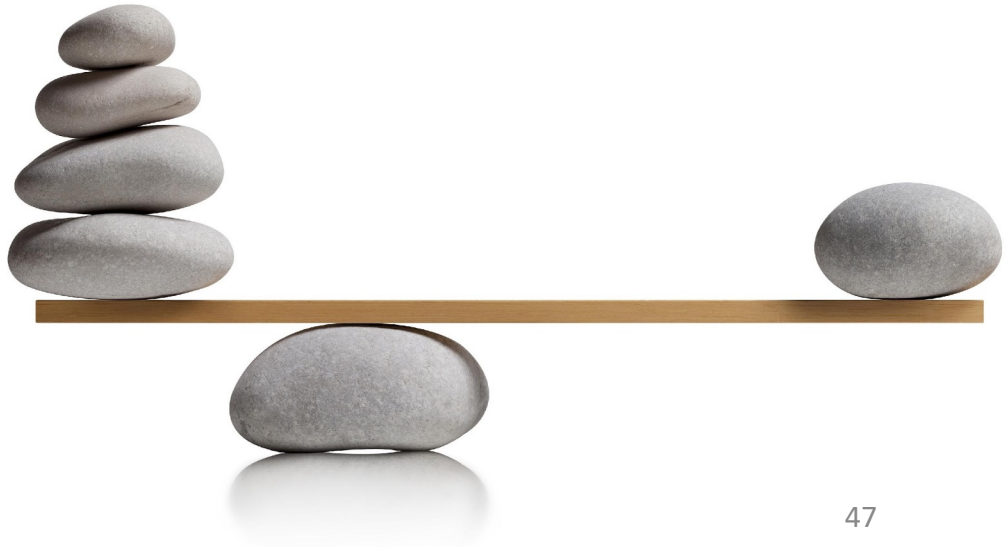
discriminatory bias: measurement bias + bias compounds injustice

representation & allocative harms

FAIRNESS

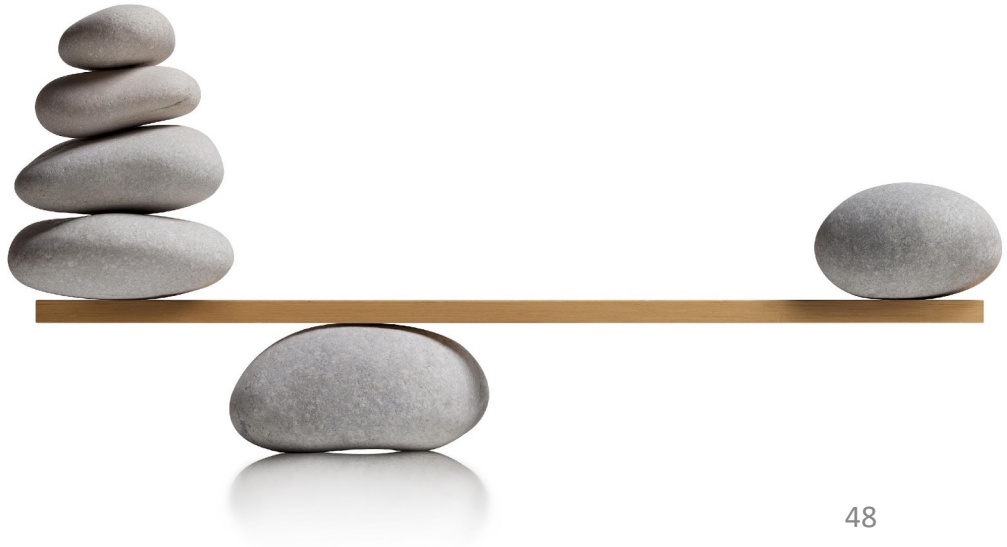
Fairness

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



Formal Equality of Opportunity

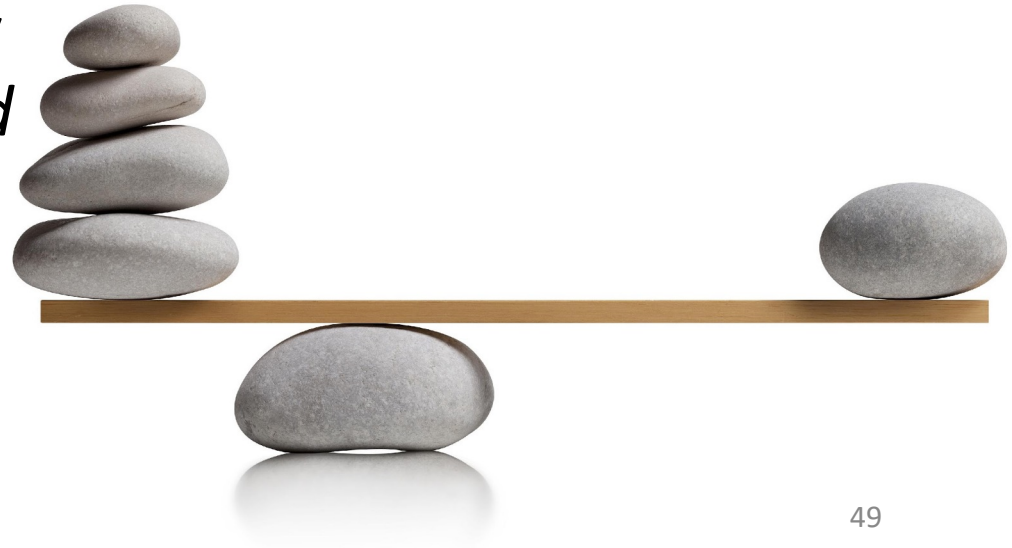
- Positions that confer superior advantages should be open to all applicants.
- Applications are assessed on their relevant merits
- applicant deemed most qualified according to appropriate criteria is offered the position
- ex: Everyone has same opportunity to develop skills needed for the job, apply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

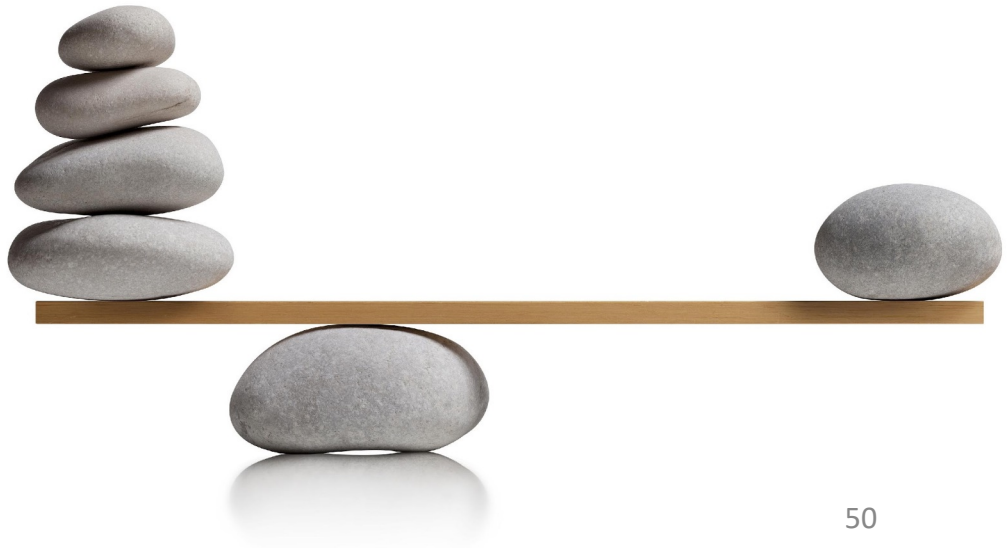


Parity

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?

Check for Discriminatory Bias

In what ways do the biases compound existing injustice?

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?**

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

https://web.stanford.edu/class/cs182/

Stanford | McCoy Family Center for Ethics in Society
SCHOOL OF HUMANITIES & SCIENCES

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

Human-Centered AI Inst.

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

https://hai.stanford.edu/

Stanford | Haas Center for Public Service

Stanford | Cardinal Service

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!