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
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
• 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
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?
VALUES IN DESIGN

- PROBLEM FORMULATION
- LANGUAGE
- BIAS AND REPRESENTATION
- FAIRNESS
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

Why should we care?
Who benefits?
Who’s harmed?
What data?
Language
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

**Normative language**

- “Lectures are 90-minutes long”
- “Assignments take more than two hours to finish”
- “Sections are mandatory”
Attendance: Not Mandatory

Textbook Required

May 21st, 2015

QUALITY
5.0

DIFFICULTY
1.0

Review: AWESOME

Professor Nick was very clear in his 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”
May 21st, 2015

Course: CS101

Attendance: A

Textbook: Yes

Online Class: Yes

Rating: 5.0

Difficulty: 1.0

This class was AWESOME for those of you like me that are not fans of doing anything further than facebook on a computer. The lectures were very clear and informative. I have very much enjoyed the lectures and look forward to what I can learn from subsequent assignments. Thank you for all that you do.

NET 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”
This class was great. I really like me that I can learn anything further than facebook on a computer. The lectures were very clear and I really much enjoyed the lectures and class. I will learn from some of the topics. Thank you for all that you do.
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?

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
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

FAIRNESS
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

98.7% 68.6% 100% 92.9%

Amazon Rekognition Performance on Gender Classification

Joy Buolamwini / Gender Shades. 2011
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 **moral**y 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?

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?
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.
Alignment problem

Alignment problem:
AI doesn’t align with intended goals, preferences, values
Homogenization of Language

**homogenization:** biases writing towards similar tone, language

Trained on form, not social context

---

**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

https://huggingface.co/spaces/society-ethics/DiffusionBiasExplorer
Fairness and AI

FAIRNESS

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

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

calling cisgender bias in AI (J Rosenbaum)
**PROBLEM FORMULATION**
Discuss the professor’s professional abilities including teaching style and ability to convey the material clearly.

**LANGUAGE**
- **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 proofread!

**BIAS AND REPRESENTATION**
- **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."

**FAIRNESS**
- **intellectual vs pleasant teachers (DALLE 2)**

- **challenging cisgender bias in AI (J Rosenbaum)**
What about AI?

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

**PROBLEM FORMULATION**

**homogenization:** biases writing towards similar tone, language

Trained on form, not social context

**LANGUAGE**

**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 \textbf{autonomy}.
b. End to end encryption addresses some \textbf{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.

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

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/](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!