Causal Inference Combined Lecture

Kenneth (Kenny) Joseph

University at Buffalo
Department of Computer Science and Engineering
School of Engineering and Applied Sciences

Sign up to meet with the librarians!

Here is a link to the form to sign up to talk with the librarians:

https://forms.gle/8Pn4dDzydv5wbuDc8

Few followup comments/reminders:

- Only **one person per group** needs to fill in the form above
    - As a group decide one the time slot that *most* of you can make and sign up for that one.
    - **Please email us (Dr. Muller, Kenny and I) once your group has filled in the form**
    - Please sign up by **Sunday, Feb 25**
- Remember that y'all need to go to the librarians with your questions. See the Unit 1 group submission instructions for more on sort of questions y'all should come prepared with.
    - See @11 for some resources provided by the librarians
    - In the spirit of trust but verify in Unit 2 group submission we'll ask y'all to report back on your meetings.

Today's plan

- **AMA / AYA** (5-10 mins)
- A whirlwind intro to ML (20 mins)
- Break
- Causal inference (thinking quantitatively about the social world) (rest of class)

What is Machine Learning? / aAa(ask Atri anything

ML as a recipe creator

> *An algorithm is like a recipe. It takes "inputs" (the ingredients), performs a set of simple and (hopefully) well-defined steps, and then terminates after producing an "output" (the meal)*

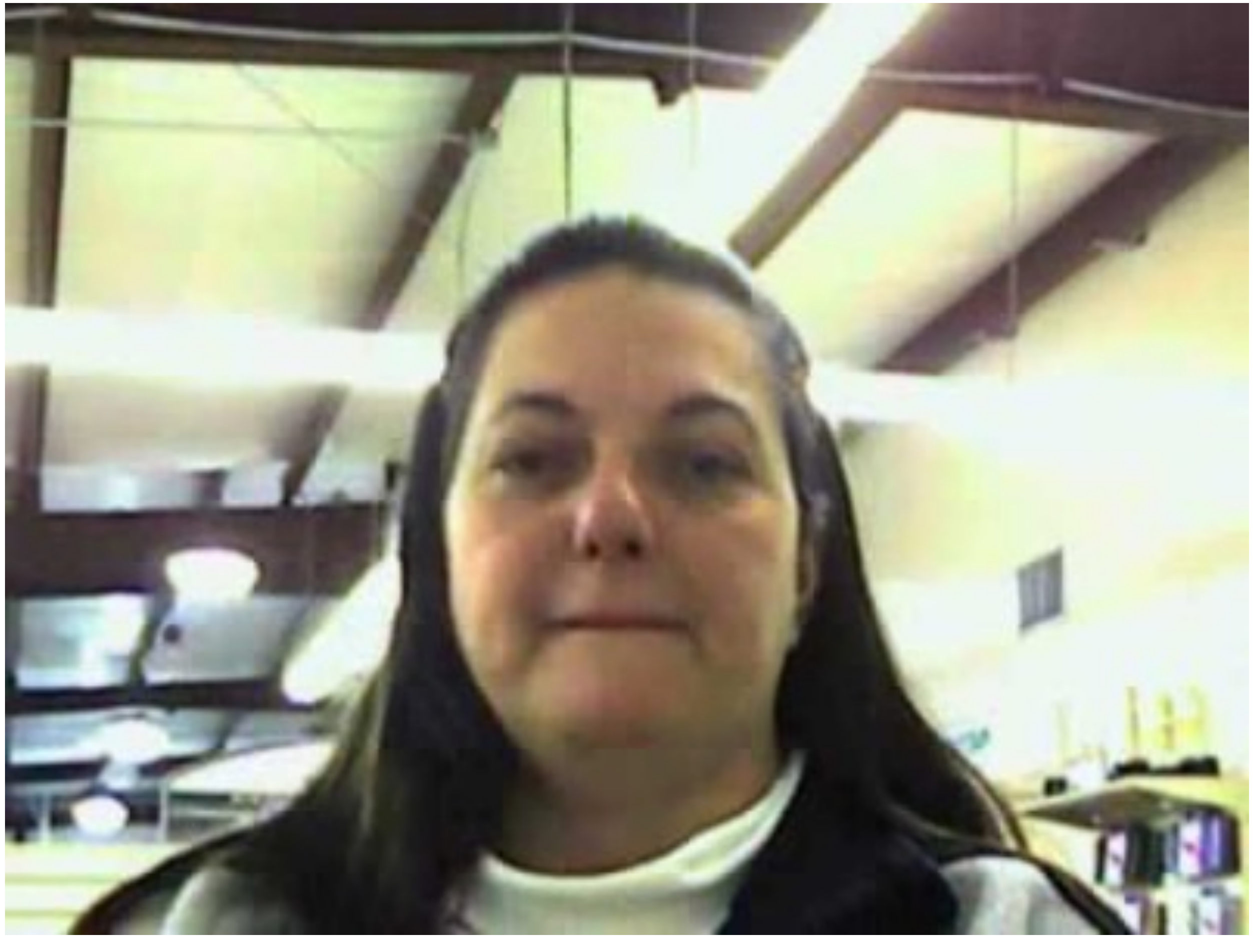
A learning algorithm is a game of roulette on a 50 dimensional wheel that lands on a particular spot (a recipe) based completely on how it was trained, what examples it saw, and how long it took to search.

https://medium.com/@geomblog/when-an-algorithm-isn-t-2b9fe01b9bb5

A boy saw 17 doctors over 3 years for chronic pain. ChatGPT found the diagnosis

Alex experienced pain that stopped him from playing with other children but doctors had no answers to why. His frustrated mom asked ChatGPT for help.

Why this course?



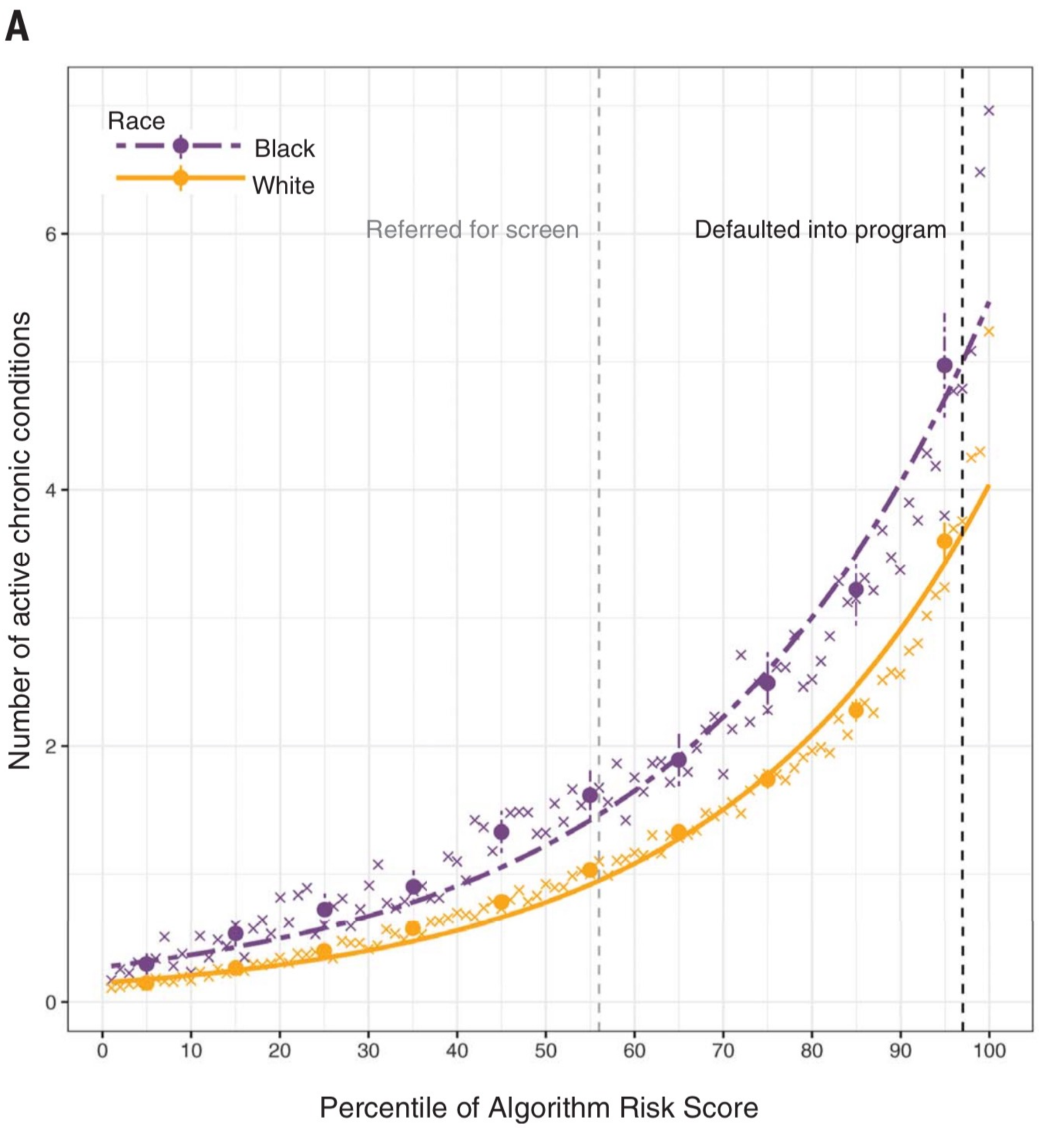


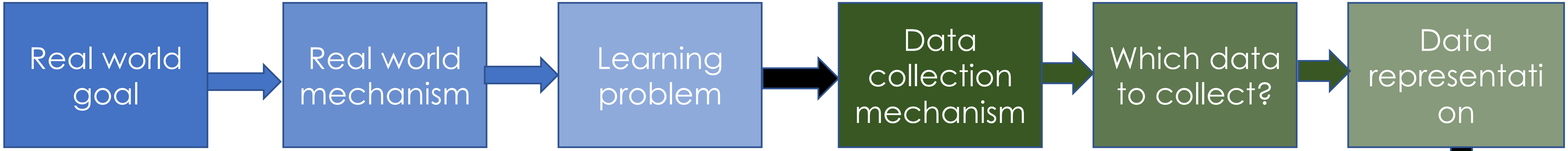
Fig. 1. Number of chronic illnesses versus algorithm-predicted risk, by race. (**A**) Mean number of chronic conditions by race, plotted against

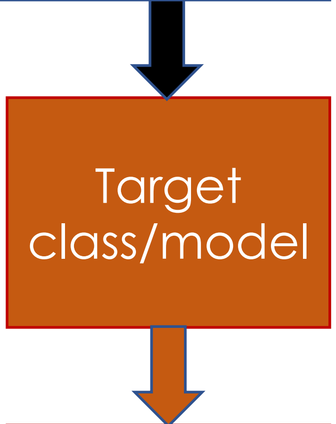
Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, *366*(6464), 447–453. https://doi.org/10.1126/science.aax2342

Fixing structural issues

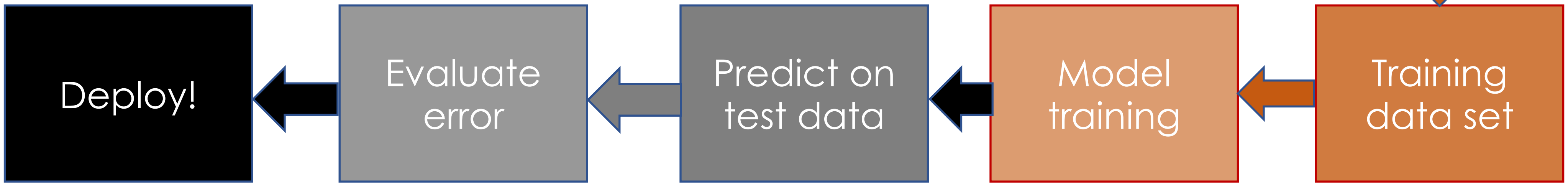
Talking about "bad data"

Putting biased ML into context: Step 1

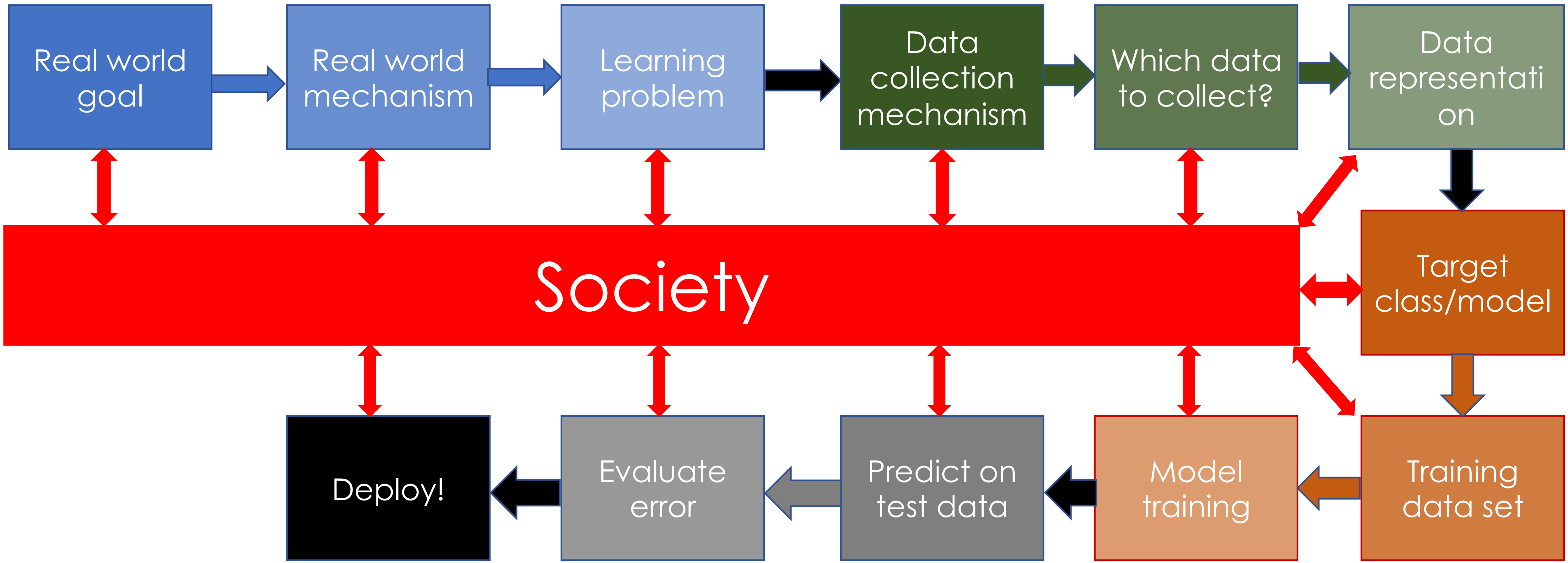




Machine learning, as you all know, does not happen in a vacuum of computers beeping and bopping.



Putting biased ML into context



A very brief run-through

For the details:

https://www-student.cse.buffalo.edu/~atri/ml-and-soc/support/notes/half-pipeline/index.html

Real world goal

Real world goal

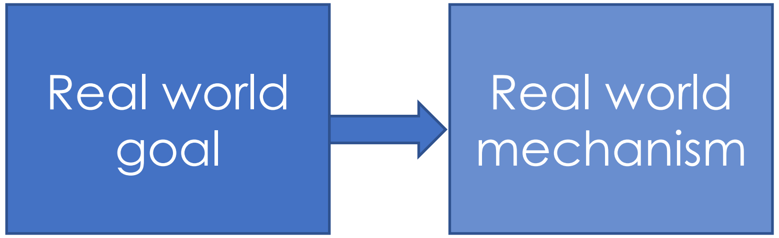
Real world goal: Example 1

Your company wants to increase revenue. A majority of revenue for your company comes from facilitating online ads. Your group has to attain this high level goal.

Real world goal: Example 2

Your hospital learns of a new government program that provides hospitals with additional resources to help manage health of patients with significant needs. The hospital management wants your hospital to utilize these funds since the hospital has been losing money in the last few quarters. However, the funds can only help a (relatively) small fraction of the patients in your hospital.

Real world mechanism



Real world mechanism: Example 1

Since online ads make up a majority of the company's revenue your group decides to improve upon the ad display (with the hope that this can generate more revenue.

Real world mechanism: Example 2

Here you get conflicting demands: the management wants to use the extra funds to cut spending (i.e. keep the current service at their current level) while doctors want to use the extra funds to supplement the existing services (i.e. add on to the existing services).

Learning problem

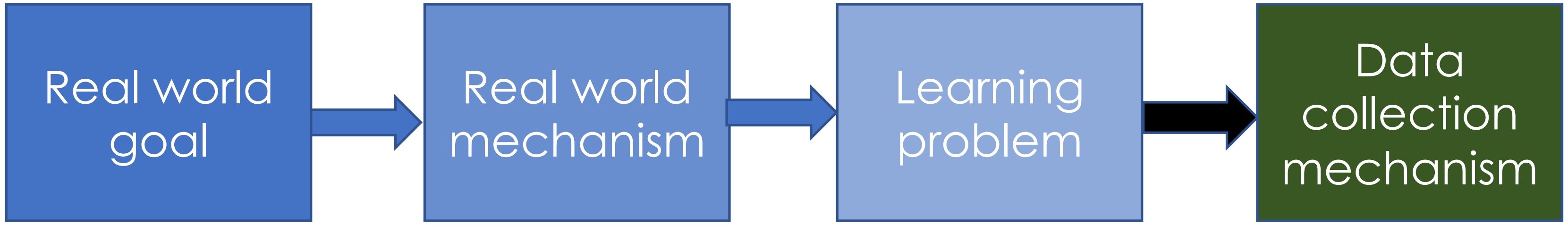
Learning problem: Example 1

Your group decides to predict the click through rate, which a measure of the likelihood that a user will click on your ad. Based on these prediction, you will better place ads.

Learning problem: Example 2

The doctors had their way so your group decides to predict the patients with most need so that they can targeted with the supplementary practice.

Data collection mechanism



Data collection mechanism: Example 1

Your group decides to log interactions with ads in the current system.

Data collection mechanism: Example 2

Your group decides to use the existing patient electronic health records (which includes details of the current care the patients receive in your hospital but possibly other details).

Data collection mechanism: Data doesn't exist

Real world goal → Real world mechanism → Learning problem → Data collection mechanism

Potential issues?

Use 3rd party data brokers

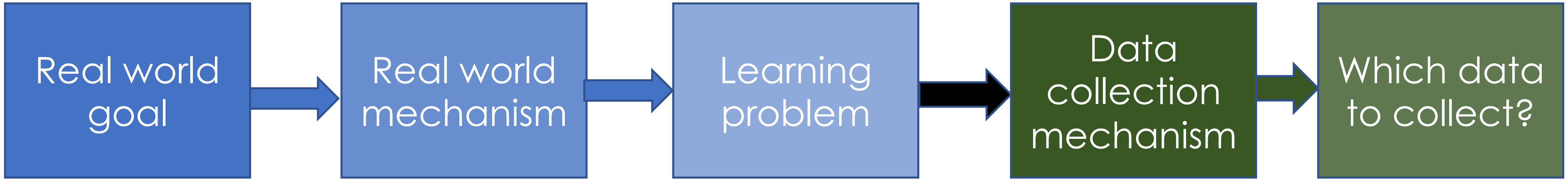

The Data Brokers So Powerful Even Facebook Bought Their Data – But They Got Me Wildly Wrong

Kalev Leetaru Contributor

AI & Big Data

I write about the broad intersection of data and society.

Which data to collect?



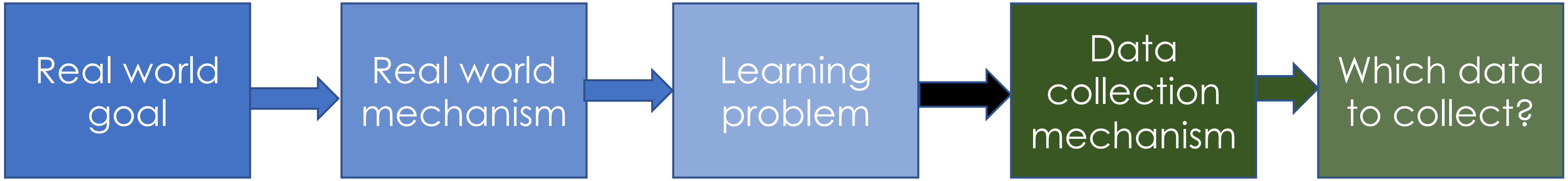
Which data to collect?: Example 1

Even though you have access to the current system, you cannot log everything. This could be because e.g. sorting everything would need a lot of storage or perhaps if the system were to log every action it observes then just the act of logging everything can slow down the system (which is not desirable). For example, your group (as Hal suggests) decides to log queries (for which ads are generated), ads and clicks.

Which data to collect?: Example 2

In this example,by restricting yourself to electronic health records, you are limiting yourself to what is logged into the electronic health records. One could e.g. try and use doctor's notes to glean more information but these are not necessarily standardized and its not clear how to extract information from doctor's notes. Further, there have been complaints from doctors on the usability of electronic health records, which raises issues about accuracy of data being collected. Finally, for the study that your group is planning will most probably need IRB approval from your hospital, which could in turn restrict which data can be collected/used for your system.

Which data to collect?: General thoughts



Expense might determine what gets collected

Time to finish a survey also has implications

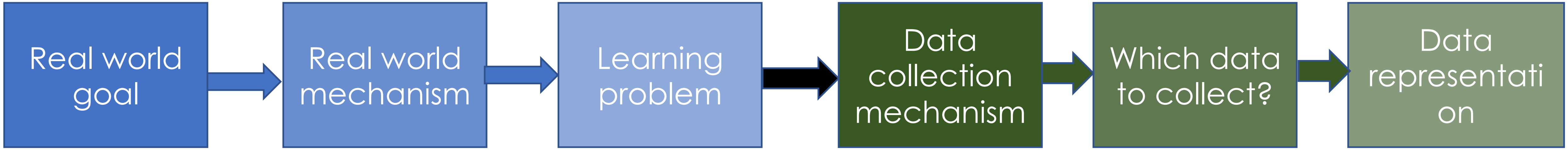
Other restrictions, e.g. from an IRB

Data representation

Real world goal → Real world mechanism → Learning problem → Data collection mechanism → Which data to collect? → Data representati on

https://www.history101.com/april-14-2003-the-human-genome-project-completed/

Data representation



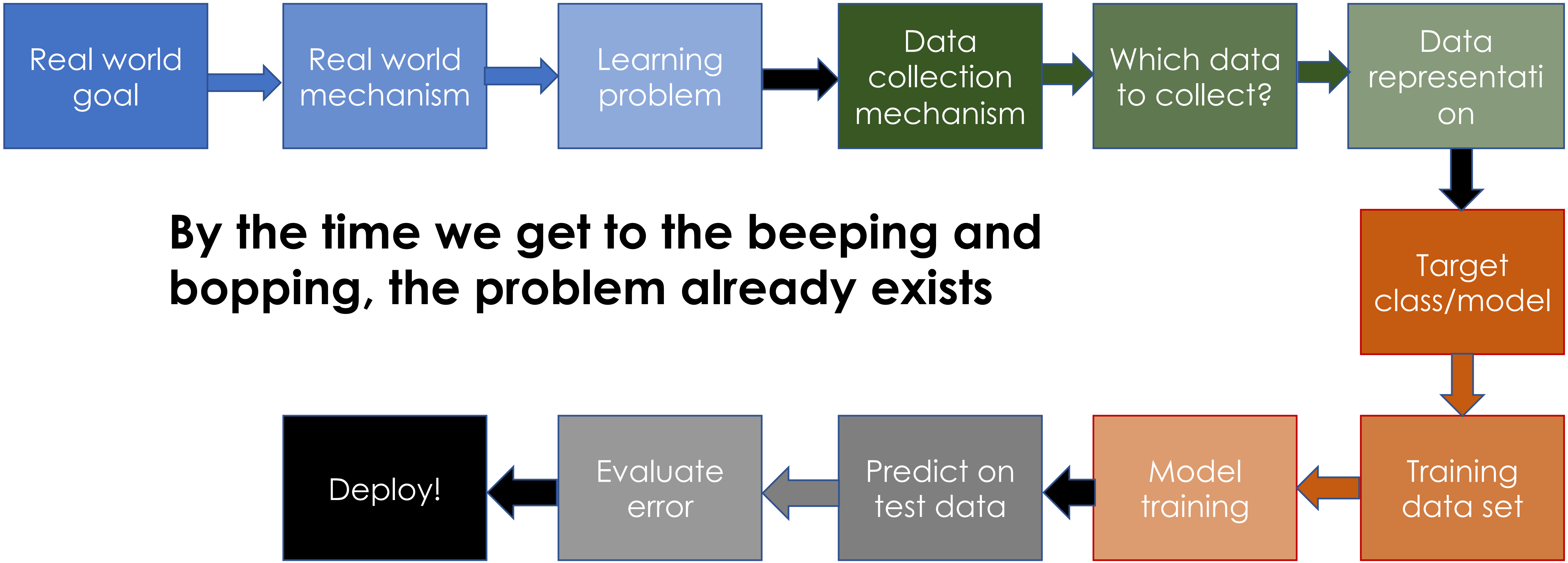
Data representation: Example 1

Your group has zeroed in on query, ad and clicks. For the latter perhaps the most natural way to represent this to encode whether a user clicked on ad or not (so either $+$ for clicked and $-$ for not clicked or 1 for clicked and 0 for not clicked. The representation for query and the ad is not as straightforward. We could store the exact text for the query and the ad but that seems to indicate issues (e.g. what is you ad text are distinct strings but are essentially the "same" for human consumption or what if someone runs a query that has the same keywords as another query but in different order). To get around this issues by using the text as is, your group decides to use a representation that is more standard in natural language processing: bag of words model.

Data representation: Example 2

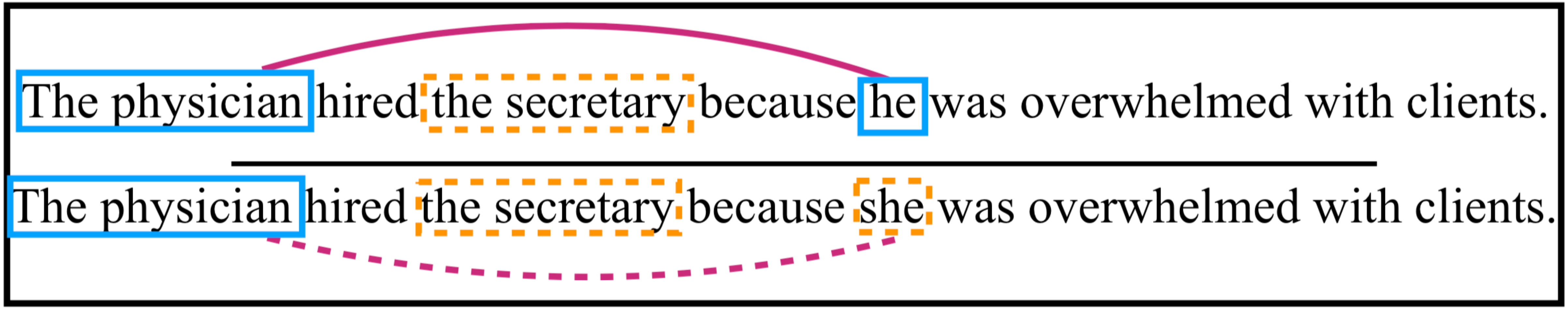
In this case since your group is using the electronic health records, then the data representation is pretty much already fixed for your group. Perhaps one exception could be to represent the doctor's notes in the bag of words model as above.

Putting biased ML into context: Step 1



By the time we get to the beeping and bopping, the problem already exists

A brief example

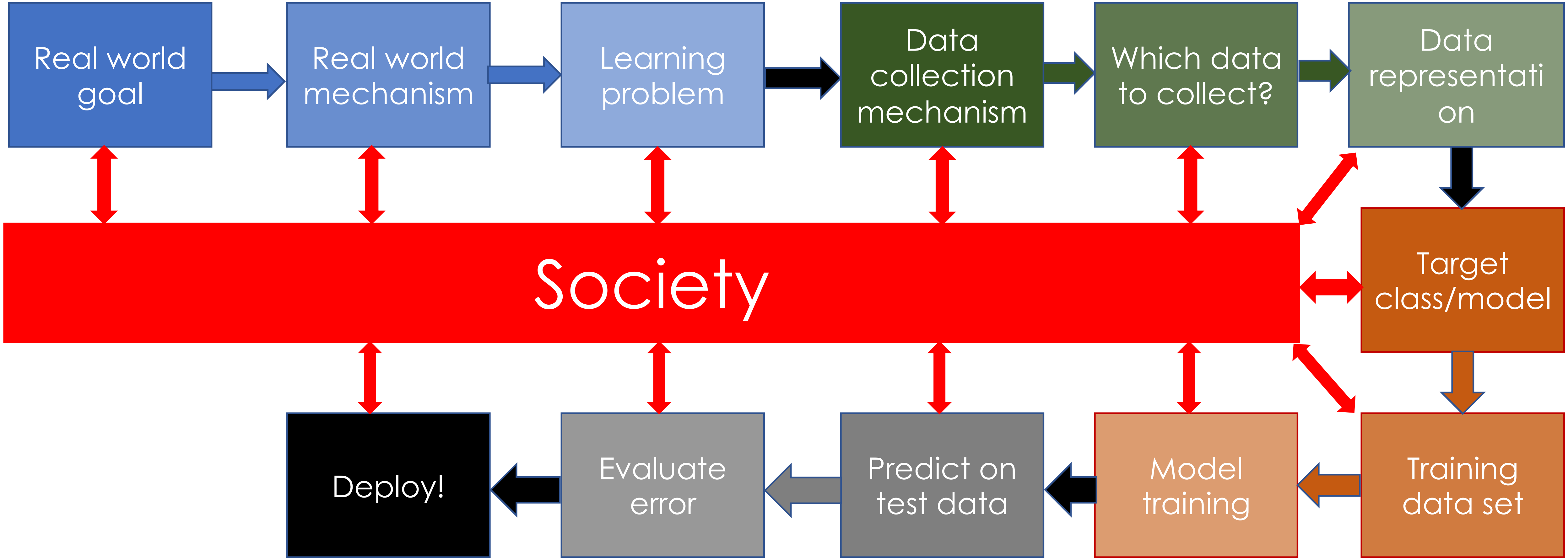


Zhao, J., Wang, T., Yatskar, M., Ordonez, V., & Chang, K.-W. (2018). Gender Bias in Coreference Resolution: Evaluation and Debiasing Methods. *ArXiv:1804.06876 [Cs]*.

Summary

- There are clear (at least to us) issues with the ML pipeline
- These are driven by various social processes
- I am a **computational social scientist**, in the I prefer to **use computation to study these processes**
- This is what Unit 1 was about for your ML&Soc teammates

Putting biased ML into context



Break!

Looking at society through a CSS lens

Three things you need for your project/to do good [computational social] science (IMO)

1. The ability to **understand/reason about** the social world
2. The ability to use that understanding to **design interventions** on the world
3. The ability to **conduct measurements** that assess the [potential] **effect** of that intervention

Looking at society through a CSS lens

Three things you need for your project/to do good [computational social] science (IMO)

1. The ability to **understand/reason about** the social world
2. The ability to use that understanding to **design interventions** on the world
3. The ability to **conduct measurements** that assess the [potential] **effect** of that intervention

To help us with this, I'm going to introduce three things, and then tie them together:

~~1. Probability/stats~~

2. Causal inference
3. Directed Acyclic Graphs

First – the math

- Goal: give us the bare-bones quantitative language
  - [Also, good to remind ourselves of these concepts]
- Concepts
  - Univariate Stats
    - Probability, Random Variables (RVs), Probability of RVs
    - Expectation & Variance
    - Probability distributions
  - Multivariate Stats
    - Conditional Probability
    - Covariance and Correlation

Seeing Theory

https://seeing-theory.brown.edu/basic-probability/

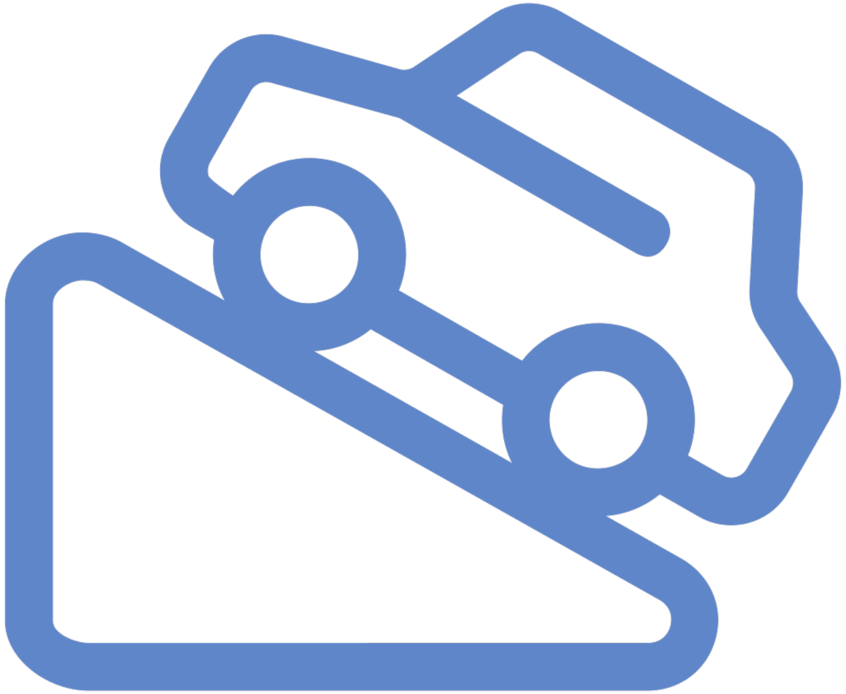
Correlation vs. Causation

- Correlation: A **measure of the relationship between two variables**
- Causation
  - Informally: "if I change A, then B will also change"
  - Formally, two ways to think about this:
    - "Do calculus"
    - Potential outcomes
    - People argue about which of these is better. Most of that is semantics. We'll make use of both.
- Causal inference: the process of inferring the causal relationships between variables.

Correlation vs causation

https://tylervigen.com/spurious-scholar

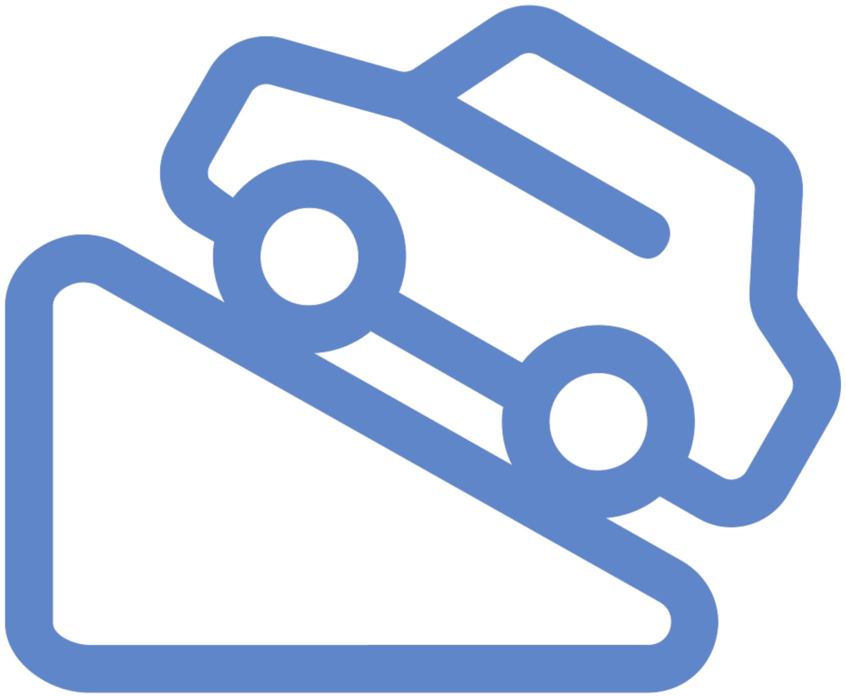
- Can you summarize the difference between correlation and causation?
- Does correlation always mean causation?
- Does *no* correlation always mean *no* causation?



https://goodauthority.org/news/milton-friedmans-thermostat/

No correlation != No causation

- What does pressing down the gas pedal do to speed?
- What does going up/down a hill do to speed?
- If we didn’t know that, what conclusions might we draw from a skilled driver?
- **Can you think of another example?**



https://goodauthority.org/news/milton-friedmans-thermostat/

Some relevant questions

- Why do we care about causation?
- How do we know something causes something else, and how do we prove it?

- Sidebar: What is the goal of machine learning?

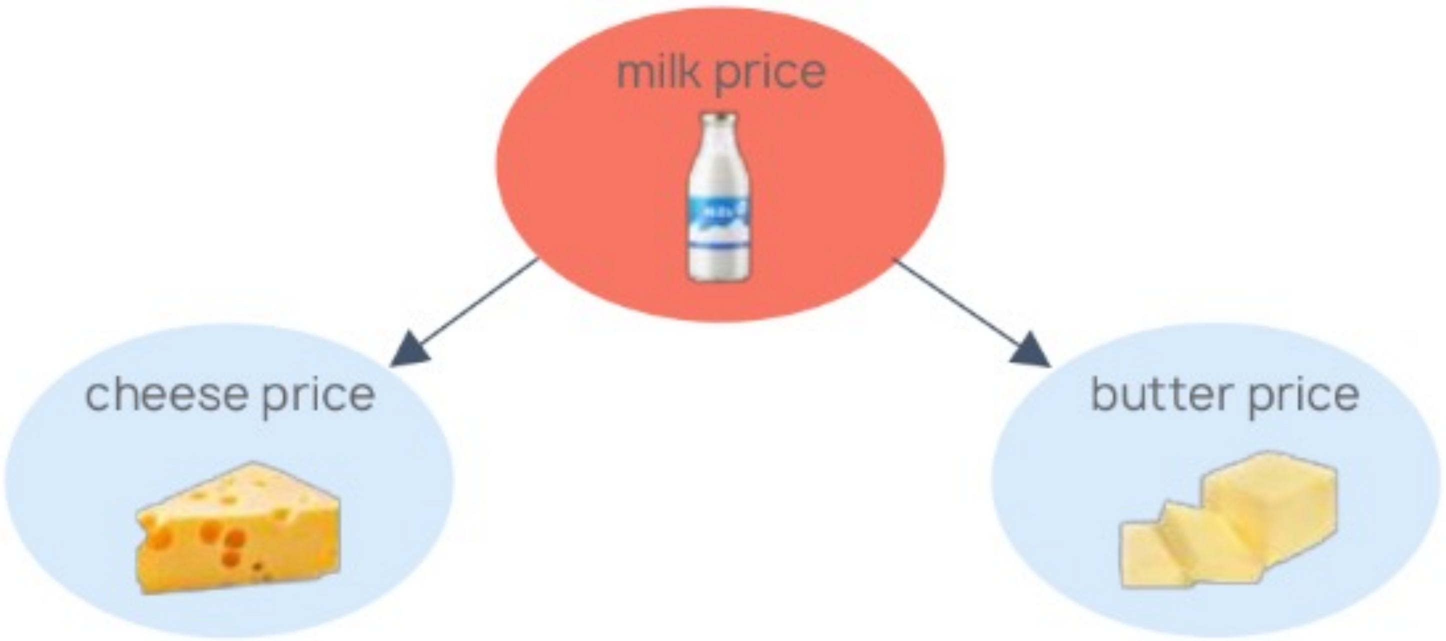
Looking at society through a CSS lens

Three things you need for your project/to do good [computational social] science (IMO)

1. The ability to **understand/reason about** the social world
2. The ability to use that understanding to **design interventions** on the world
3. The ability to **conduct measurements** that assess the [potential] **effect** of that intervention

Causal DAGs– a tool to explain causal reasoning

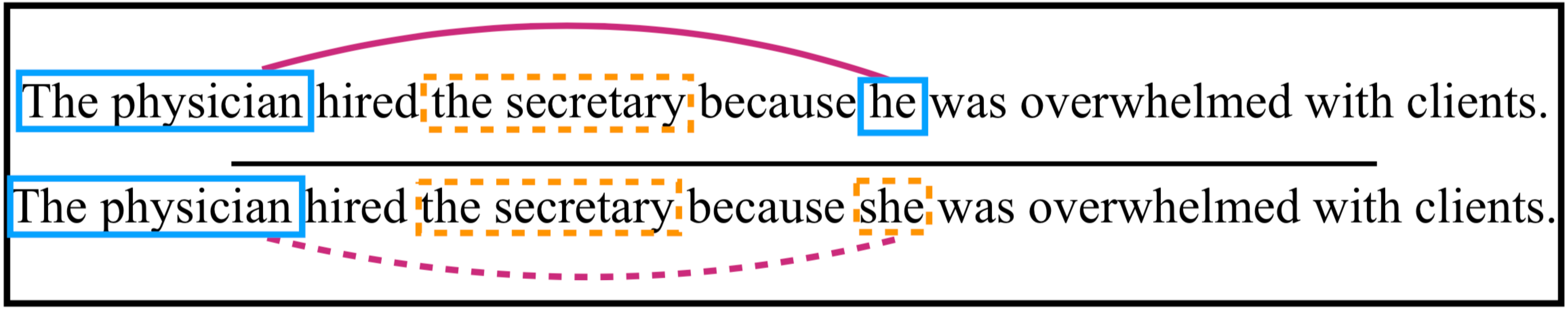
https://causalens.com/resources/white-papers/why-correlation-based-machine-learning-leads-to-bad-predictions/



What is a causal graphical model?

- A **diagram** that helps us **explain** our **assumed causal** relationships between two things
  - Nodes are **random variables**
  - Edges are **causal relationships between RVs**
- Let's practice!

A brief example



Zhao, J., Wang, T., Yatskar, M., Ordonez, V., & Chang, K.-W. (2018). Gender Bias in Coreference Resolution: Evaluation and Debiasing Methods. *ArXiv:1804.06876 [Cs]*.

Some questions about DAGs

- DAGs have limitations!
  - Cycles are real
  - Hard to differentiate certain kinds of probabilistic relationships
- We disagree on both semantics and beliefs about the issues
  - How do we resolve differences between your DAGs?
- Some things are hard to put a number on
  - We have to get suuuper specific to actually run experiments, and **we have to be right**
  - "No causation without intervention"
- How do you know there's a node? An edge?
  - What counts as evidence?
  - How do we reference evidence?

How do we know?

- **Prior work**

- **Activity:**
  - Try to find an academic article that provides support for the following statement:

The more severe your illness, the more likely you are to die, regardless of treatment

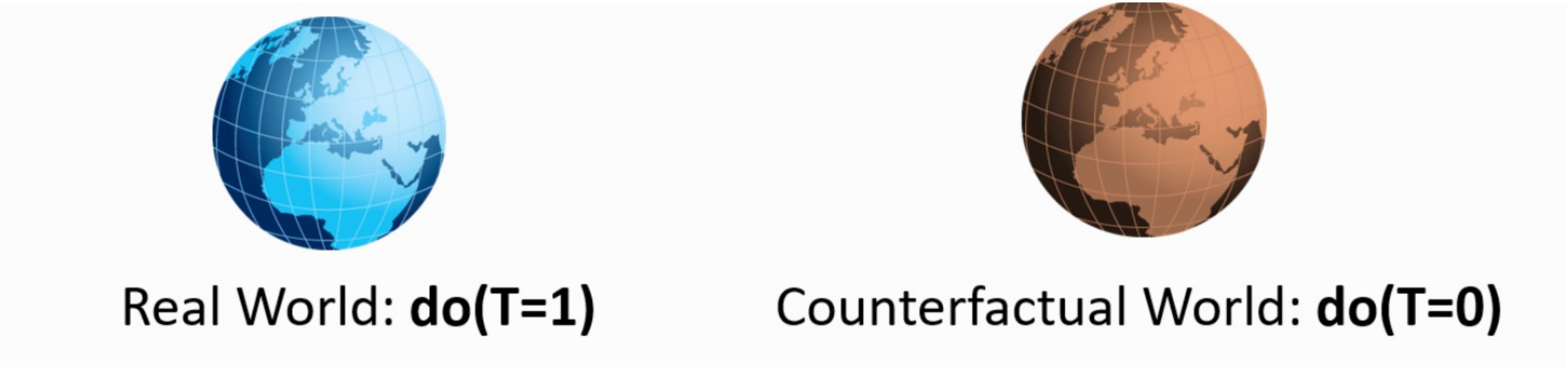
How do we know?

- **Prior work**
- Conditional dependencies + Theory
  - ***What is theory?***
  - Sidebar: How do we identify conditional dependencies that are not likely to be random?
  - Sidebar to the sidebar: The "standard" way here is to assume linearity, do linear regression, and look at statistical significance of coefficients. This is *fiiiineeee*.
- Causal Evidence

Introducing causal inference

What if we had a method that could **explicitly account for these challenges and help us measure *real causation*?**

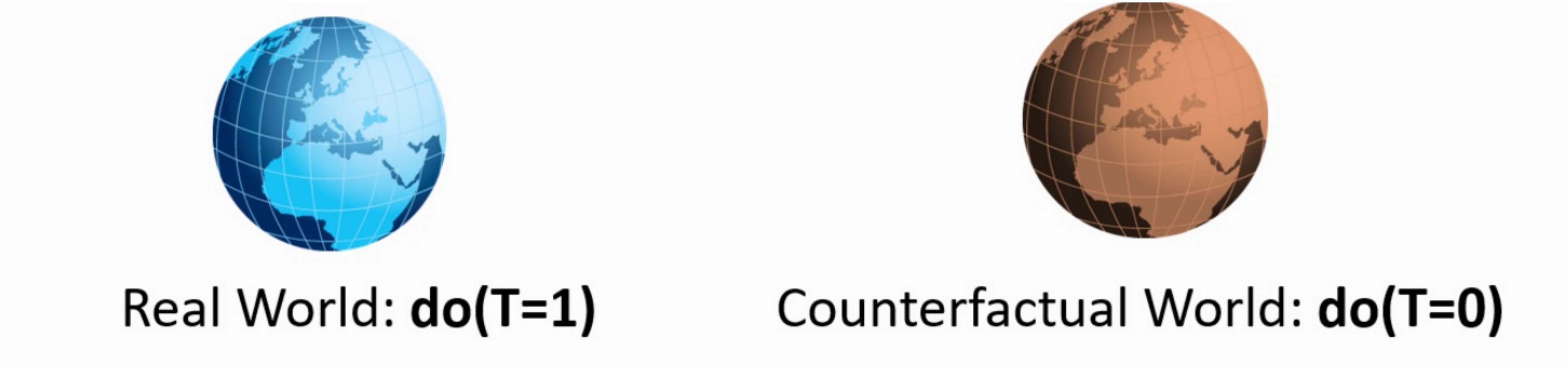
Enter causal inference!



https://microsoft.github.io/dowhy/example_notebooks/tutorial-causalinference-machinelearning-using-dowhy-econml.html

Sounds great! But there's a catch...

I want to know if telling Bernie he is going to get an A no matter what will make his participation better or worse



I can't both give Bernie an A AND not give it to him!
This is the **Fundamental Problem of Causal Inference**

https://microsoft.github.io/dowhy/example_notebooks/tutorial-causalinference-machinelearning-using-dowhy-econml.html

Problems with the simple story

- Sometimes, experimentation is unethical
  - I suspect half of you would be very angry if I gave you a placebo study guide ☺
- Other times, we might have wanted to experiment but simply couldn't, and are left with a bunch of observational data

Uh-oh

- Now:
  - We have an idea how to evaluate an intervention, but when we can't experiment, we have to **control for factors associated with both the intervention and the outcome**
  - Related: how do we decide on a treatment in the first place?
- Enter DAGs [and then, causal graphical models]. Informally, there are two "kinds" of probabilistic DAGs
  - Bayes Nets encode the factorizations of any joint probability distribution
  - Causal Graphical Models put explicit assumptions about causation into DAGs.
  - We'll focus on the latter

As always – easier said than done

DIONE CORWIN

8555 PROSACCO SQUARES, CHICAGO, IL

PHONE
+1 (555) 870 1476

EXPERIENCE

BAUMBACH, KOZEY AND NIENOW 04/2018 – present
Philadelphia, PA // *Director of Internal Audit*

- Resides in the Greater Houston Marketplace
- Knowledge of accounting principles, practices, and financial reporting
- Experience working with automated ledger applications
- Strong knowledge of the PCAOB standards, SOX Compliance requirements and COSO's Internal Control Integrated Framework
- Requires strong written and verbal communication skills, analytical and project management skills
- Excellent communication, collaboration and presentation skills
- Requires strong collaboration skills, working closely with management and external audit teams to develop efficient audit plans that address key risks

YUNDT, JACOBI AND SCHINNER 01/2012 – 11/2017
Philadelphia, PA // *Head of Audit*

- Strong experience in the financial services industry, recent experience
- Internal or external audit experience
- To drive the audit framework continuous review and improvements
- Lead the audits in the Digital Bank, focusing on all risks arising from the activities of the Bank
- Plan and perform risk-based assessments, understand and evaluate the business environment, related controls and processes
- Review draft/ final report and participate in the presentation of audit findings

EDUCATION

OTIS COLLEGE OF ART AND DESIGN
Bachelor's in Accounting

SKILLS

- Business acumen including knowledge of the Company's strategy, industry, key risks, operations, and culture
- Outstanding communications skills, including presenting to, advising, and educating executive leadership and Board of Directors
- A quick learner that displays excellent judgment and problem-solving skills which enable timely and appropriately risk balanced advice and guidance
- A team player who displays self-confidence, encourages collaboration, and establishes credibility that earns organizational trust from superiors and peers who can find common ground in solving problems
- Extensive experience with risk-based internal auditing and annual planning

Rachel Corwin

8555 PROSACCO SQUARES, CHICAGO, IL

PHONE
+1 (555) 870 1476

EXPERIENCE

BAUMBACH, KOZEY AND NIENOW 04/2018 – present
Philadelphia, PA // *Director of Internal Audit*

- Resides in the Greater Houston Marketplace
- Knowledge of accounting principles, practices, and financial reporting
- Experience working with automated ledger applications
- Strong knowledge of the PCAOB standards, SOX Compliance requirements and COSO's Internal Control Integrated Framework
- Requires strong written and verbal communication skills, analytical and project management skills
- Excellent communication, collaboration and presentation skills
- Requires strong collaboration skills, working closely with management and external audit teams to develop efficient audit plans that address key risks

YUNDT, JACOBI AND SCHINNER 01/2012 – 11/2017
Philadelphia, PA // *Head of Audit*

- Strong experience in the financial services industry, recent experience
- Internal or external audit experience
- To drive the audit framework continuous review and improvements
- Lead the audits in the Digital Bank, focusing on all risks arising from the activities of the Bank
- Plan and perform risk-based assessments, understand and evaluate the business environment, related controls and processes
- Review draft/ final report and participate in the presentation of audit findings

EDUCATION

OTIS COLLEGE OF ART AND DESIGN
Bachelor's in Accounting

SKILLS

- Business acumen including knowledge of the Company's strategy, industry, key risks, operations, and culture
- Outstanding communications skills, including presenting to, advising, and educating executive leadership and Board of Directors
- A quick learner that displays excellent judgment and problem-solving skills which enable timely and appropriately risk balanced advice and guidance
- A team player who displays self-confidence, encourages collaboration, and establishes credibility that earns organizational trust from superiors and peers who can find common ground in solving problems
- Extensive experience with risk-based internal auditing and annual planning

Threats to validity

How Black Are Lakisha and Jamal? Racial Perceptions from Names Used in Correspondence Audit Studies

S. Michael Gaddis

University of California, Los Angeles

Abstract: Online correspondence audit studies have emerged as the primary method to examine racial discrimination. Although audits use distinctive names to signal race, few studies scientifically examine data regarding the perception of race from names. Different names treated as black or white may be perceived in heterogeneous ways. I conduct a survey experiment that asks respondents to identify the race they associate with a series of names. I alter the first names given to each respondent and inclusion of last names. Names more commonly given by highly educated black mothers (e.g., Jalen and Nia) are less likely to be perceived as black than names given by less educated black mothers (e.g., DaShawn and Tanisha). The results suggest that a large body of social science evidence on racial discrimination operates under a misguided assumption that all black names are alike, and the findings from correspondence audits are likely sensitive to name selection.

Keywords: racial discrimination; inequality; names; audit studies; experiments

Some questions about DAGs

- DAGs have limitations!
  - Cycles are real
  - Hard to differentiate certain kinds of probabilistic relationships
- We disagree on both semantics and beliefs about the issues
  - How do we resolve differences between your DAGs?
- Some things are hard to put a number on
  - We have to get suuuper specific to actually run experiments, and **we have to be right**
  - “No causation without intervention”
- How do you know there's a node? An edge?
  - What counts as evidence?
  - How do we reference evidence?

Looking at society through a CSS lens

Three things you need for your project/to do good [computational social] science (IMO)

1. The ability to **understand/reason about** the social world
2. The ability to use that understanding to **design interventions** on the world
3. The ability to **conduct measurements** that assess the [potential] **effect** of that intervention

Putting biased ML into context