

ML and Society

Apr 12, 2022

Discussion Summary for Thursday

note @55 🗨️ ★ 🔒 ⌵ stop following 3 views

Discussion Summary (Fairness and Abstraction in Sociotechnical Systems) now open on Autolab

Sorry for the delay but Autolab is now accepting submissions for the 4th discussion summary, which are due by **5pm on Wednesday**.

logistics autolab discussion_summary

edit · good note | 0 Updated 19 hours ago by Atri Rudra

Today's OH

 note @56    

[stop following](#)

1 views

Today's office hours ends at 2:35pm

I have to leave campus at 2:35pm today so will cur short my office hours then. If you were planning to stop by between 2:35 and 2:50pm, please let me know and we'll figure out an alternative.

office_hours

[edit](#)

· [good note](#) | 0

Updated 34 seconds ago by Atri Rudra

What is bias?

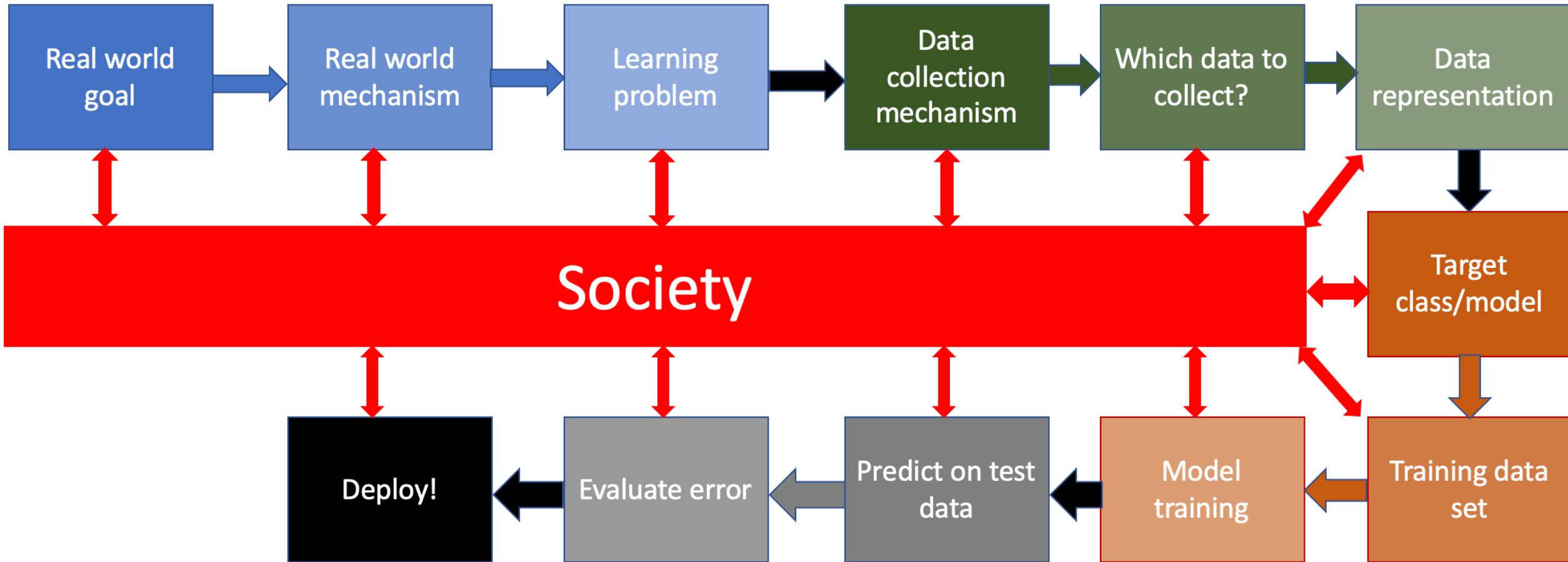
What is bias?

Another loaded term that we will use is the term **bias**. In particular, there are roughly three kinds of notions of bias that is relevant to these notes:

1. The first notion (which might be the least known) occurs in a dataset where there are certain specific collection of input variable values occur more than others. This essentially measure how [far away from a truly random](#) dataset the given dataset is. Note that this notion is bias is **necessary** for ML to work. If all the datapoints are completely random (i.e. both their input and target variable values are completely random), then there is no bias for a classifier to "exploit"-- in other words, one might as well just output a random label for prediction.
2. The second notion of bias is that of [statistical bias](#), where in our setting this would mean that the binary classifier outcome does not reflect the distribution of the underlying target variable. Such a classifier would be [well calibrated](#), if this does not happen. One could consider a well-calibrated binary classifier to be fair in some sense. This will be one notion of fairness that will come up in the COMPAS story. (This is the notion of fairness used in the rejoinder to the ProPublica article).
3. The finally notion of bias is the [colloquial use of the term](#) that is mean to denote an outcome that is **not fair**. Most of the definitions of fairness in the literature deal with this notion of bias. And a couple of definition of this kind of fairness will also play a part in the COMPAS story (this is the notion of fairness used in the ProPublica article).



Bringing society back into the picture



Six categories of bias of the 3rd kind

A Framework for Understanding Unintended Consequences of Machine Learning

Harini Suresh
MIT
hsuresh@mit.edu

John V. Guttag
MIT
guttag@mit.edu

Abstract

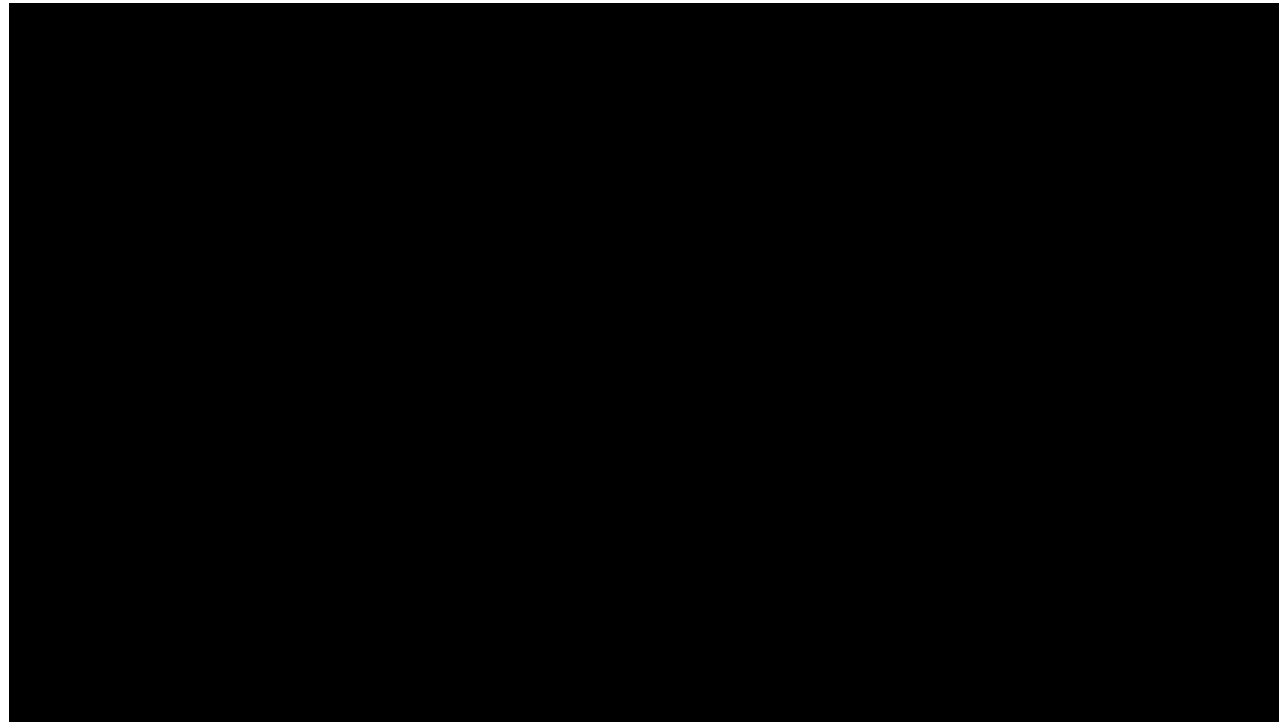
As machine learning increasingly affects people and society, it is important that we strive for a comprehensive and unified understanding of potential sources of unwanted consequences. For instance, downstream harms to particular groups are often blamed on “biased data,” but this concept encompasses too many issues to be useful in developing solutions. In this paper, we provide a framework that partitions sources of downstream harm in machine learning into six distinct categories spanning the data generation and machine learning pipeline. We describe how these issues arise, how they are relevant to particular applications, and how they motivate different solutions. In doing so, we aim to facilitate the development of solutions that stem from an understanding of application-specific populations and data generation processes, rather than relying on general statements about what may or may not be “fair.”

Consider the following toy scenario: an engineer building a smile-detection system observes that the system has a higher false negative rate for women. Over the next week, she collects many more images of women, so that the proportions of men and women are now equal, and is happy to see the performance on the female subset improve. Meanwhile, her co-worker has a dataset of job candidates and human-assigned ratings, and wants to build an algorithm for predicting the suitability of a candidate. He notices that women are much less likely to be predicted as suitable candidates than men. Inspired by his colleague’s success, he collects many more samples of women, but is dismayed to see that his model’s behavior does not change. Why did this happen? The *sources* of the disparate performance were different. In the first case, it arose because of a lack of data on women, and introducing more data solved the issue. In the second case, the use of a proxy label (human assessment of suitability) meant that the model’s (statistical) definition of

Historical Bias

Bias ingrained in society

Cannot be avoided even with **perfect sampling** of the population



Earlier example: Google Image search “ceo”

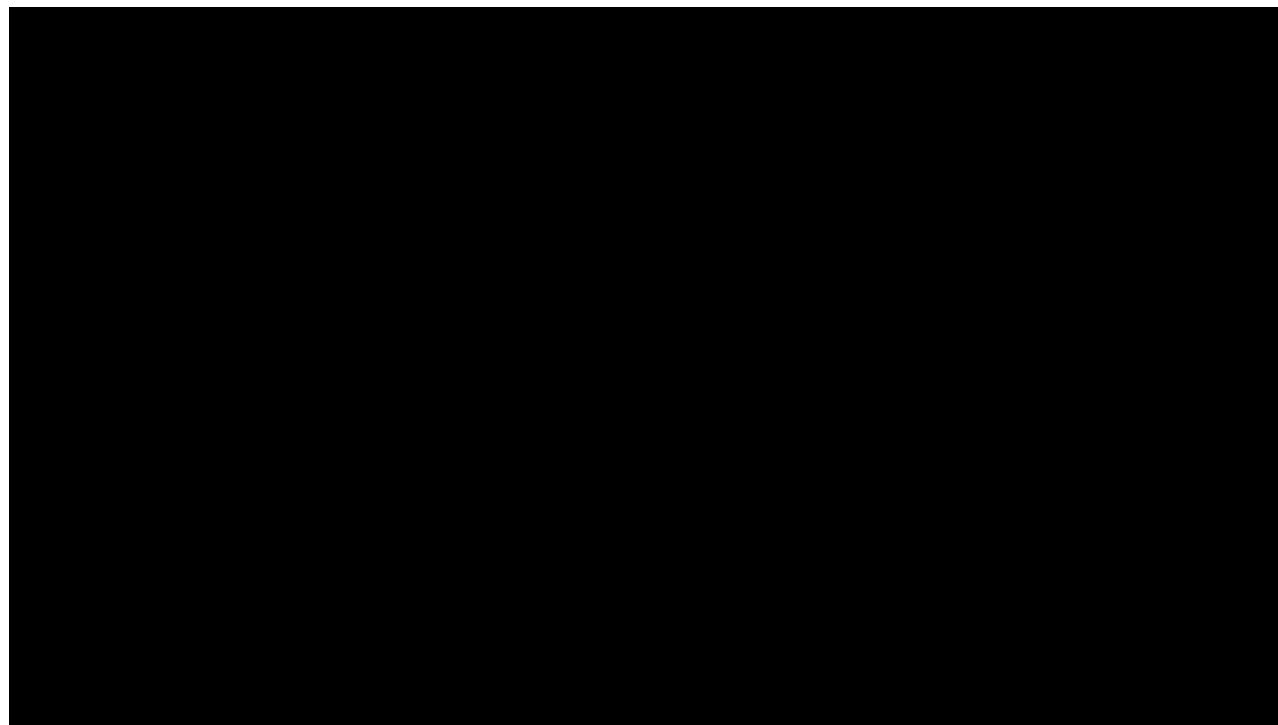
The image shows a Google search interface for the term "ceo". The search bar at the top contains the text "ceo" and a search icon. Below the search bar, there are navigation tabs for "All", "News", "Images", "Books", "Videos", "More", "Settings", and "Tools". The "Images" tab is selected. To the right of the search bar, there is a "Sign in" button and a "SafeSearch" option.

Below the navigation tabs, there is a row of circular icons representing various brands and categories, including "business", "google", "cartoon", "snapchat", "microsoft", "apple", "woman", "desk", "amazon", "uber", "pepsi", "youtube", "black", "facebook", "starbucks", "successful", "twitter", and "salary".

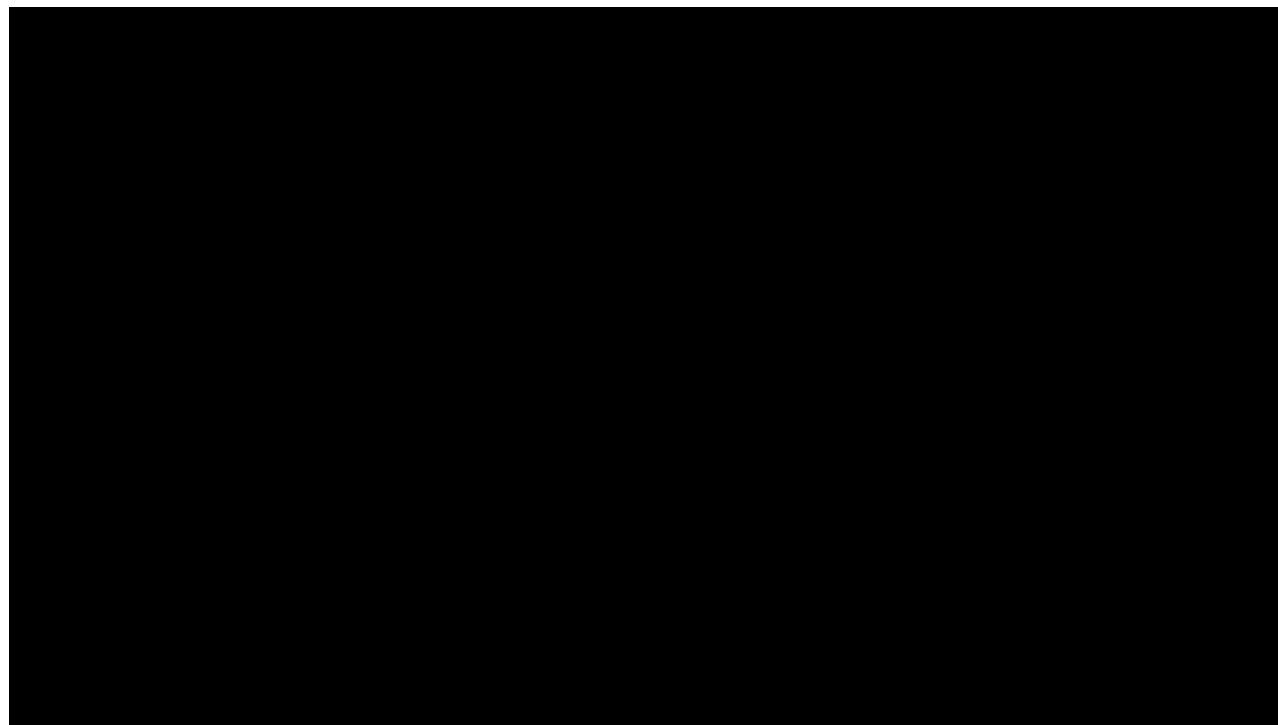
The main content area displays a grid of image search results. Each result consists of a small thumbnail image and a caption with a source URL. The results include:

- Chief executive officer - Wikipedia
- Casey's Anno... businesswire...
- What do CEOs do? A CEO ... steverrobbins.com
- Marriott CEO Arne Sor... chiefexecutive.net
- Harvard study: What CEO... cnbc.com
- McDonald's Fires CEO Over... forbes.com
- How to use 'CEO magic' ... europeanceo.com
- Rise of the next-gen bank CEO americanbanker.com
- HSBC's Caretaker CEO ... bloomberg.com
- Siemens USA CEO say... businessinsider.com
- Selective CEO begins the ... insurancebusinessmag...
- You are the CEO of Your Life... personalexcellence.co
- CEO Confidence Ticks U... chiefexecutive.net
- Meet Our CEO -... stellarairportstor...
- Proximus approves CEO app... mobileworldlive.com
- Byron Sanders as New Presi... bigthought.org
- John Furner President & CEO... corporate.walmart.com
- CEO MESSAGE | JCB Glo... global.job
- CEO Message | Company | AGC agc.com
- visits Egypt to promote U.S. inv... egypttoday.com
- CEO vs. Owner: The Key Diff... onlinemasters.ohio.edu
- Bank of America CEO Bria... bloomberg.com
- Lockheed Martin's Marilyn... chiefexecutive.net
- Former Bain, eBay CEO John ... consulting.us
- brilliant answer in a job int... cnbc.com
- Four CEOs Were Dethron... forbes.com
- Mike Roman | 3M... 3m.com
- Carnival Corp. CEO Arnold... blackenterprise.com
- NBCUniversal CEO Steve Bur... adweek.com
- Rackspace appoi... techcrunch.com
- Ford CEO Jim Hackett Earne... motor1.com
- Boeing CEO pushed out a... theverge.com
- Message from the CEO | Canon... global.canon
- Abbott Labs' CEO Is Step... barrons.com
- Boeing CEO Dennis Mullenbur... cnbc.com
- Not the ideal messenger': Faceb... aljazeera.com
- A Message From Your Favorite C... youtube.com
- Giannico Farrugia, M.D., ... newsnetwork.mayoclinic.org
- Wellforce health... bostonglobe.com
- Roche - Meet our CEO roche.com
- China has become a CEO-level... mckinsey.com
- Boeing CEO is Parkland ... mcall.com
- New Northrop Grumman CEO... fortune.com
- The Future Of Real Estate? L... forbes.com
- Markus Duesmann to ... audi-mediacentr.com
- New CEO at Porsche Korea newsroom.porsche.com
- Related searches: cartoon ceo, ceo text, desk ceo
- Related searches: ceo word, ceo logo, ceo female
- Why Novartis CEO wants to ... cnbc.com
- BMW promotes Oliver Zi... reuters.com
- Appointed CE... airbus.com
- F5 Networks names new ... geekwire.com
- Boeing CEO Dennis Mullenbur... businessinsider.com

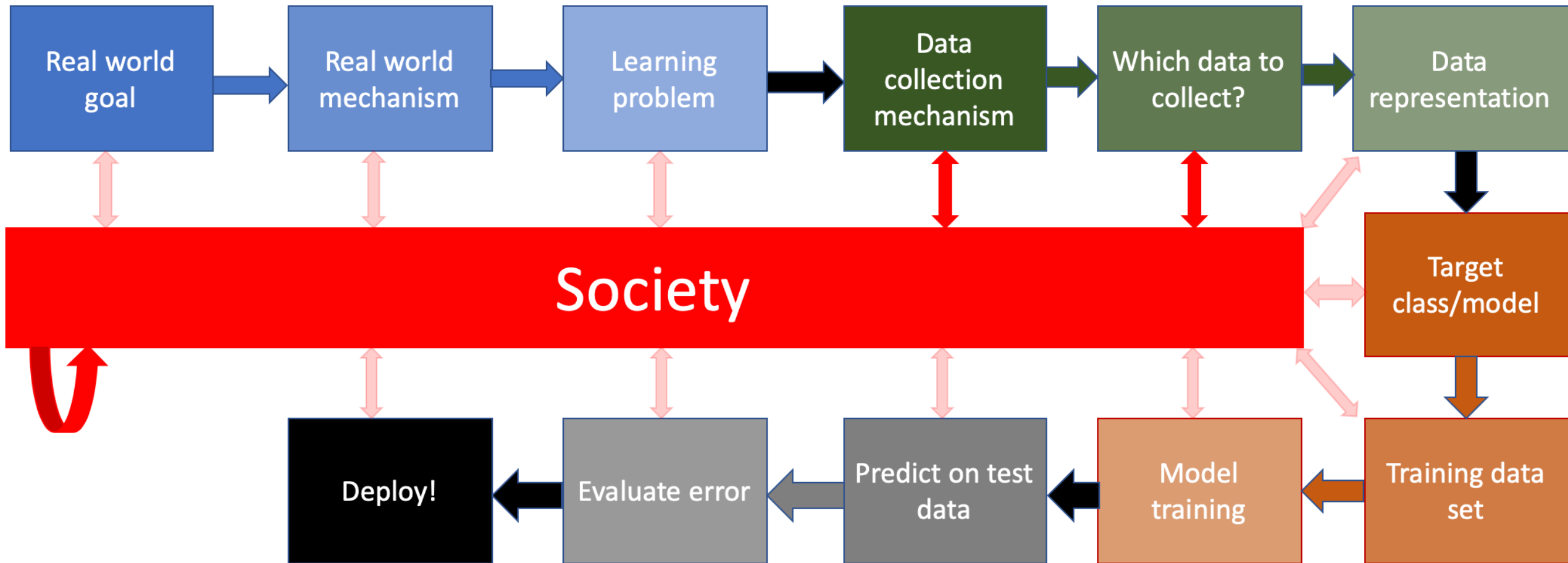
There *are* few women CEOs



Another interesting video



What are the relevant interactions?



How do we handle historical bias?

How we handle historical bias in the ML pipeline?

Among the five classes of bias we are going to study in these notes, historical bias is something that the ML pipeline has least control over. However, this **does not mean that you as a developer of the ML pipeline should just raise your hands and "give up"**. At the **very least you should be aware** of historical bias so that you can make sure that the latter stages of the ML pipeline do not exacerbate the historical bias when you deploy your ML pipeline.

In certain situations, your real-life goal means you will **have** to account for historical bias. E.g. if the real life goal is to improve diversity in your company workforce, then if you are designing an ML pipeline to hire more diverse folks who will be successful in your company, then historical bias in your workforce (e.g. technology companies having much fewer women and even fewer people of color) is something that your ML pipeline will have to **explicitly** handle.

Automated hiring

ML pipeline are being deployed for hiring in many places. This is a pipeline where you should pay attention to historical biases in the data that you are using to figure out who would make for a "good" employee.

To get a sense of what can happen/go wrong: **play this online game:** [Survival of the best fit](#) .



Representation bias

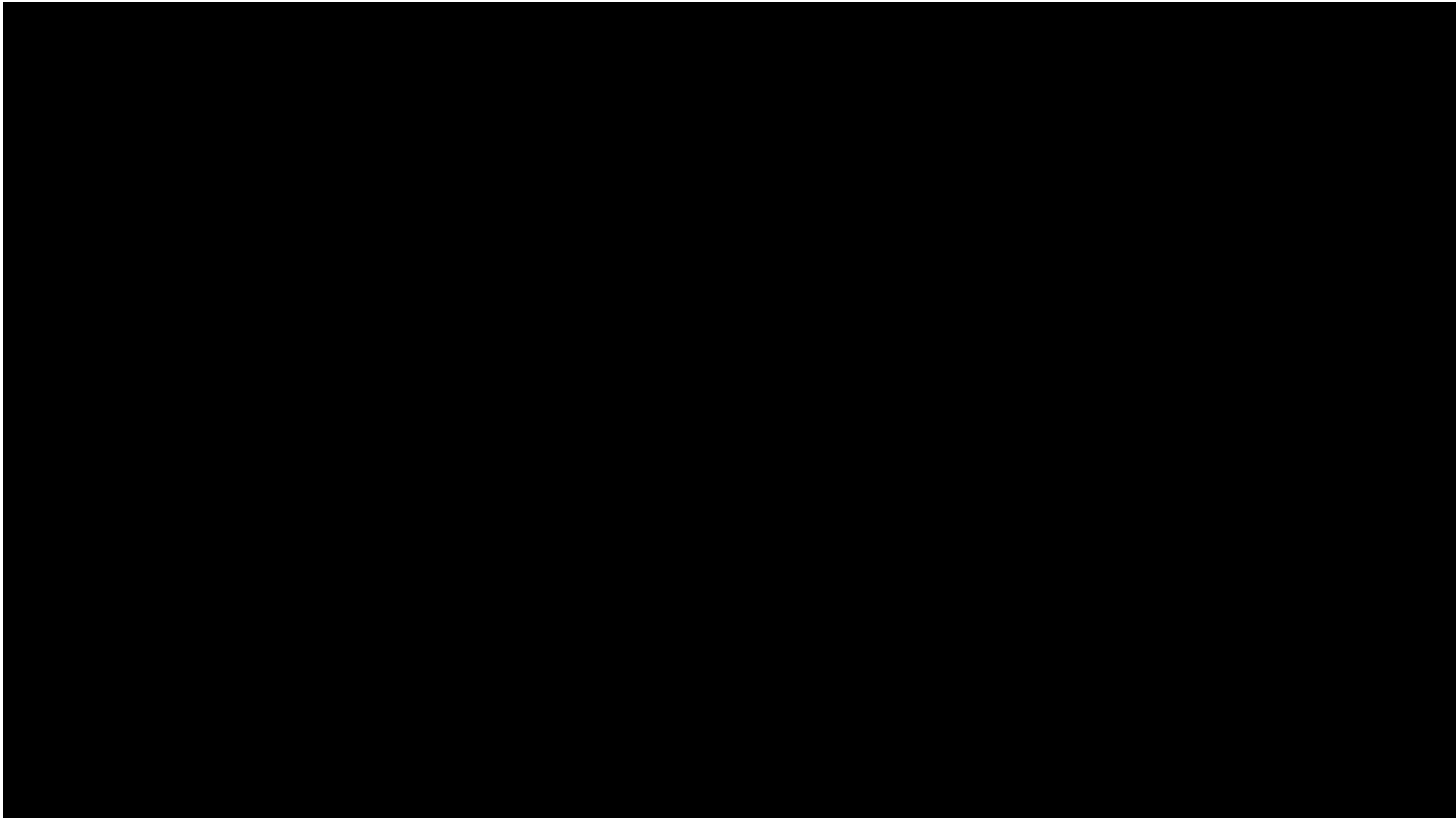
Certain section(s) of population excluded in your dataset

Relation to statistical bias

Statistical bias refers to the issue that the your (training) dataset is not a perfectly random sample from the ground truth. I.e., the dataset point are not representative of the underlying population distribution. Thus, selection bias **by definition** leads to representative bias.

However, it is possible that there is **representative bias even in the absence of selection bias**. We will see a particular reason later on but here is another scenario. Consider the [demographics of Finland](#), where non-whites form a tiny fraction of the Finnish population. Now if even if we had a truly random sample of the population of Finland, unless the sample size is very large, there will be very few non-whites in your sample. In other words, even though technically there is no selection bias, there will be presentation bias in your system.

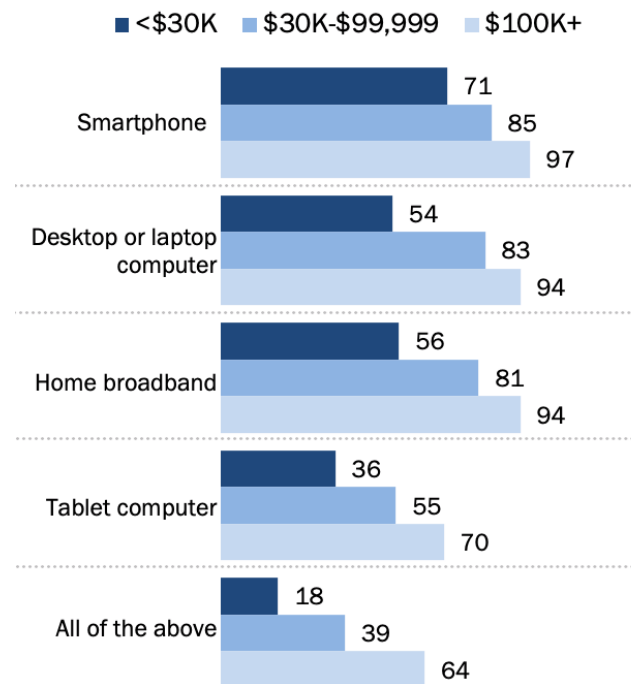
Reason #1: Data collection method excludes



Who is being excluded here?

Lower-income Americans have lower levels of technology adoption

% of U.S. adults who say they have the following ...




Note: Respondents who did not give an answer are not shown.
Source: Survey conducted Jan. 8-Feb. 7, 2019.

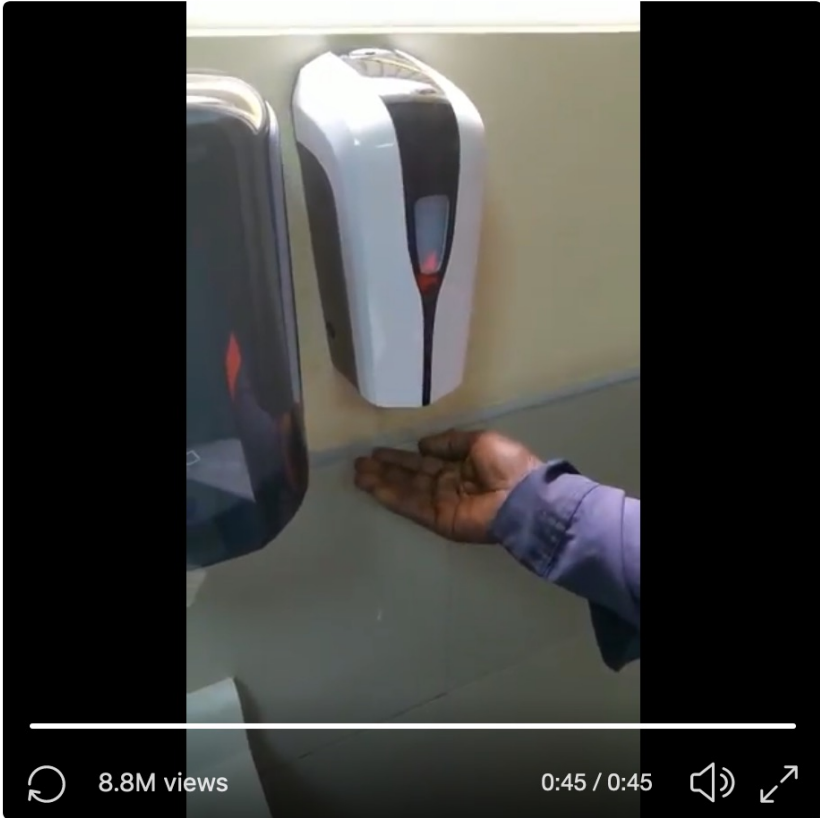
Acknowledgement: Kate Crawford



Reason #2: training data is not representative

 **Chukwuemeka Afigbo**
@nke_ise

If you have ever had a problem grasping the importance of diversity in tech and its impact on society, watch this video

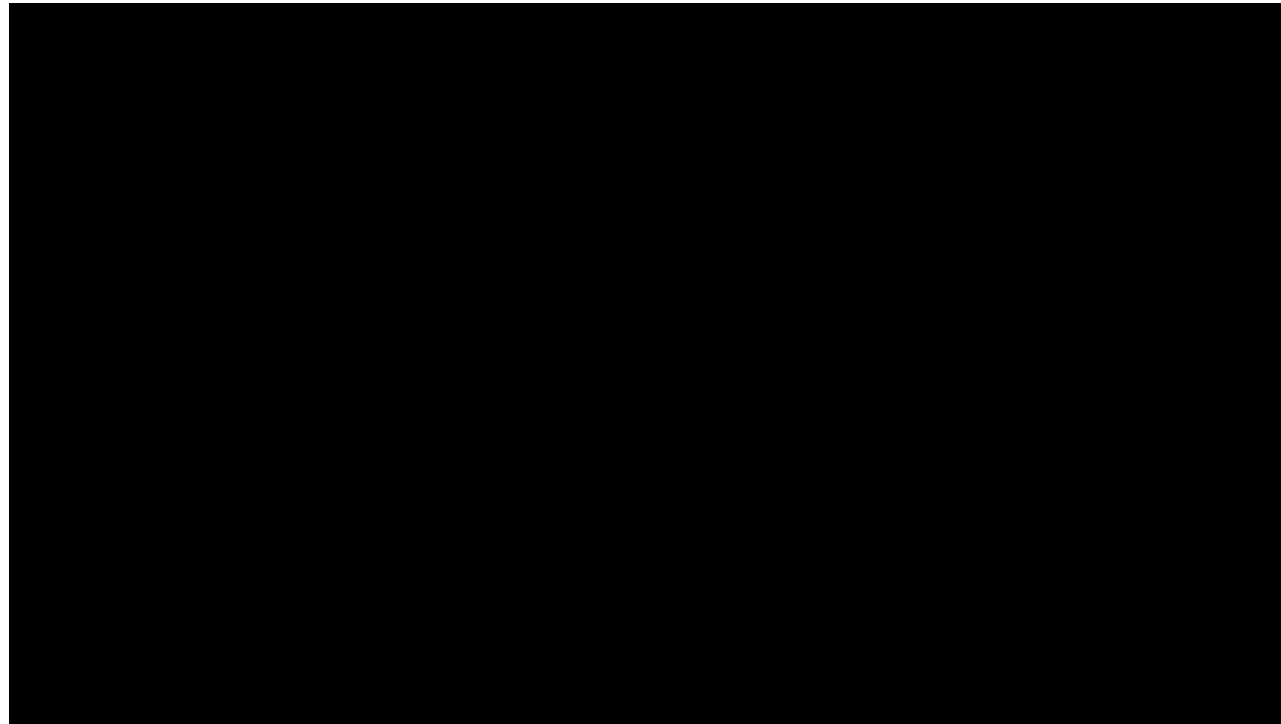


8.8M views 0:45 / 0:45

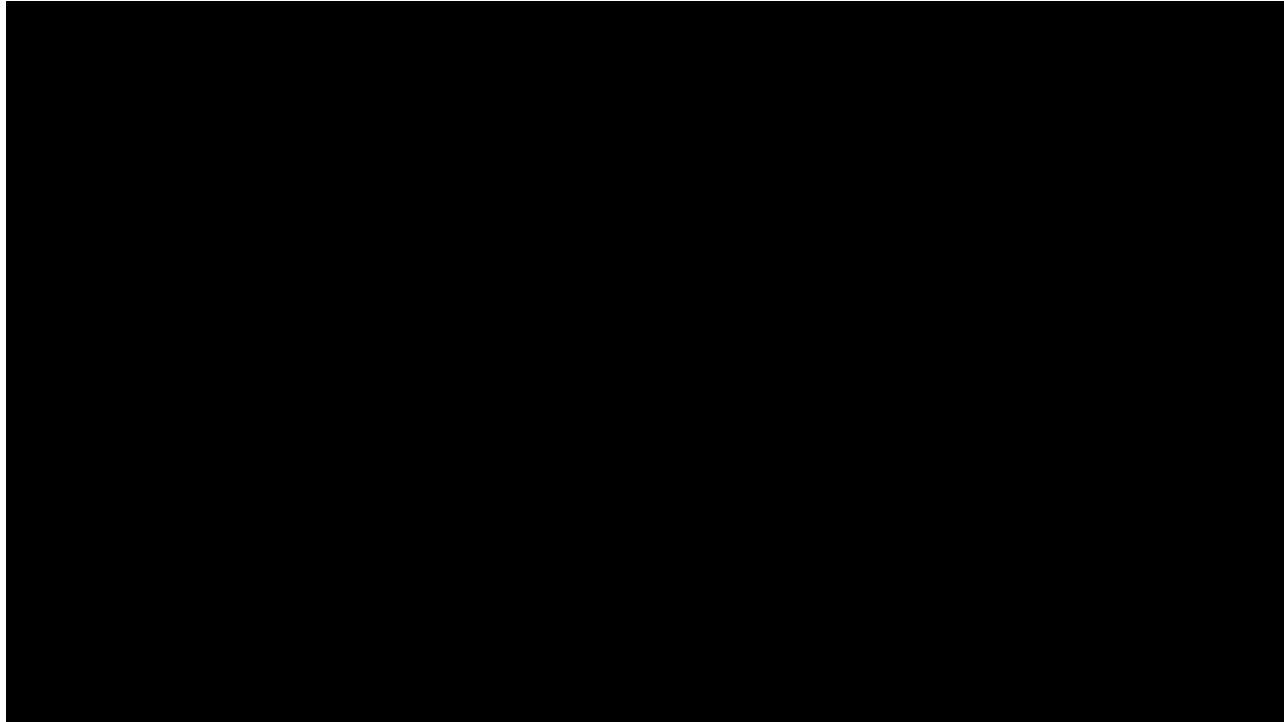
220K 5:48 AM - Aug 16, 2017

Not specific to Machine Learning per se...

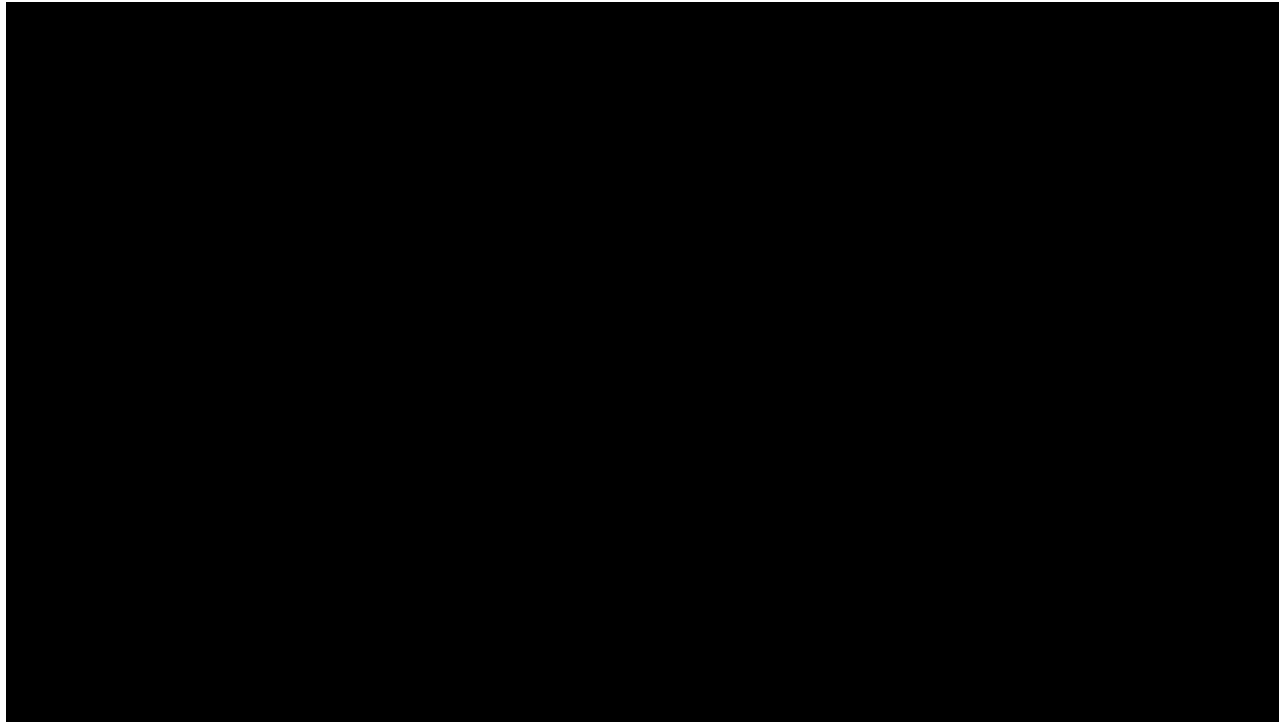
Do you know what a camera/film roll is?



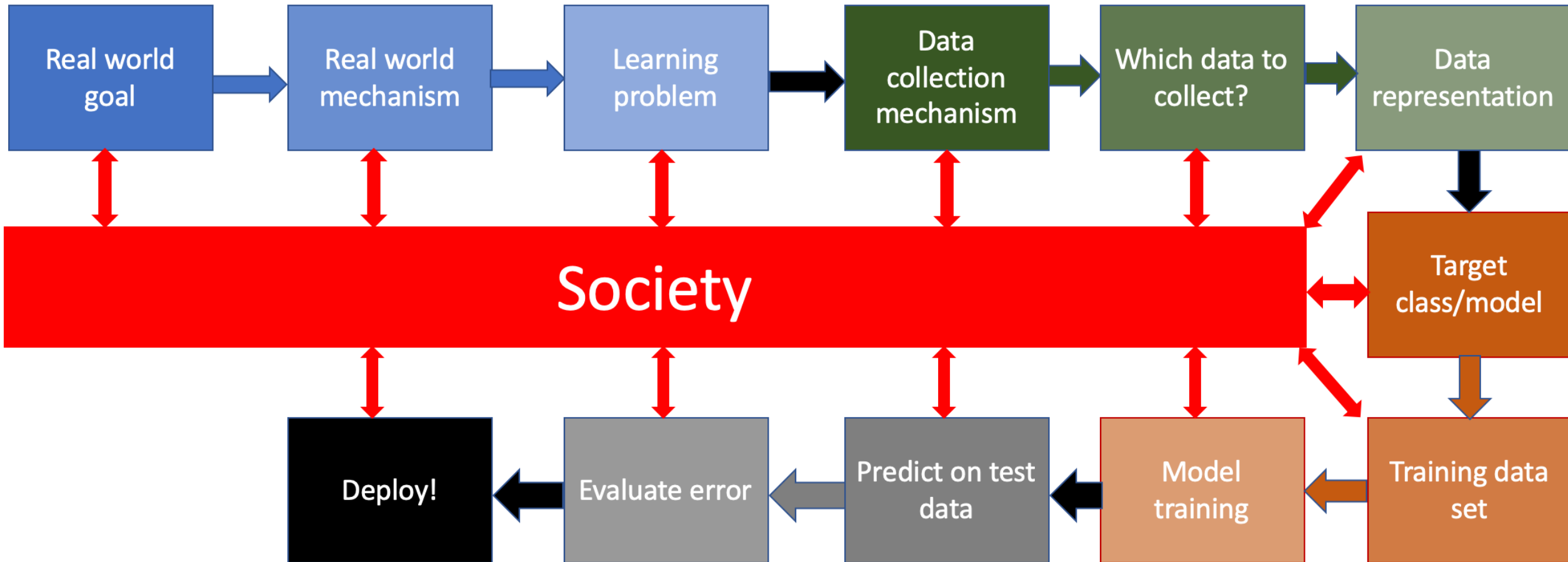
Camera rolls had an issue...



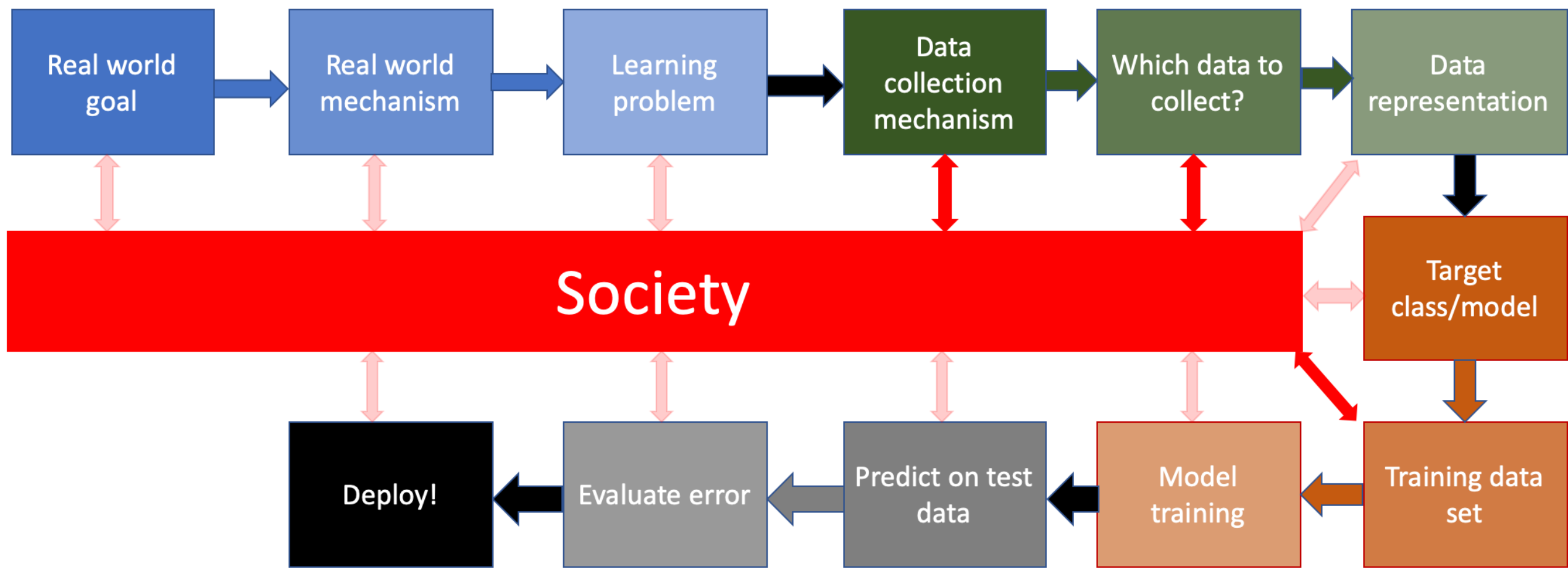
Representative ex. for representation bias



Back to the ML pipeline



What are the relevant interactions?





How do we handle representation bias?

How we handle representation bias in the ML pipeline?

Unlike the case of [handling historical bias](#), handling representation bias is something that should be handled in the ML pipeline because one potentially could have more control over this kind of bias than historical bias.

If when creating the ML pipeline, you have control over the data collection mechanism, then you should make an effort to not create representation bias. However, we should note that this could be expensive. See e.g. this video that talk about this issue with the census (along with other issues):



If your ML pipeline uses an existing dataset then you should understand the history of the dataset. It would be helpful if the dataset has a [datasheet](#) associated with it.

Measurement bias

Using a proxy variable instead of the “real” variable

Digression: How do you measure recidivism

This is a good time to clarify/remind you that the recidivism rates being higher for blacks than whites does **not** imply that blacks necessarily reoffend at a higher rate than whites. Think about why this could be the case.

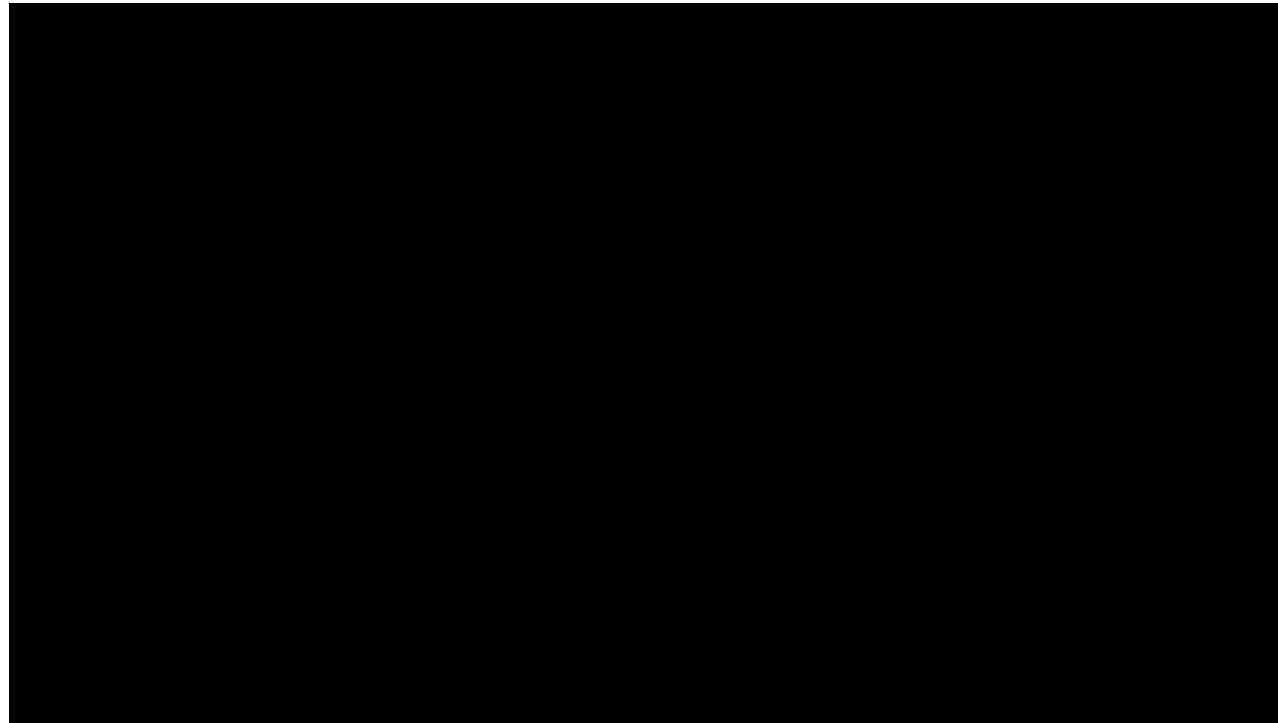
Hint: How would you measure whether someone reoffended or not?

[Click here to see the answer](#)

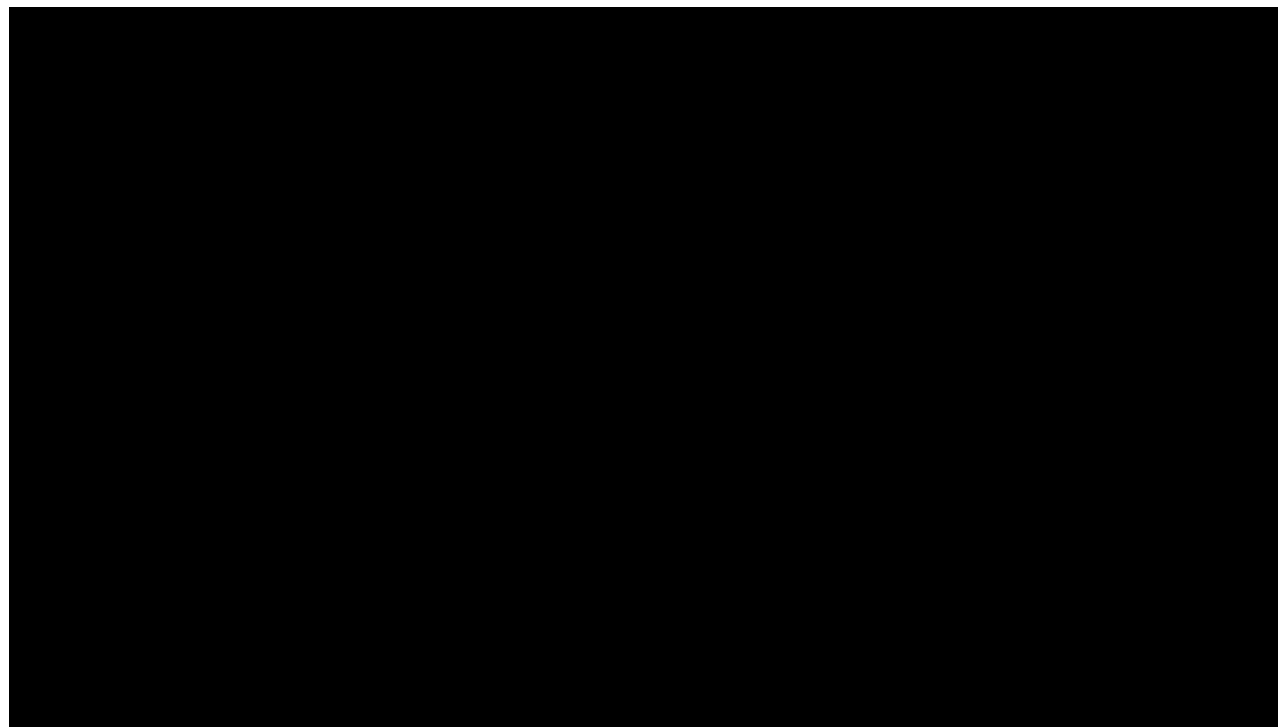
The issue is the question mentioned in the hint. There is no way for sure to know whether a person reoffended in a certain time frame or not: e.g. what about the case when someone commits a crime but never gets caught for it? On the other hand, if someone is caught/arrested for committing a crime that can be recorded.

The notion of recidivism in the COMPAS dataset was whether someone was **arrested** for another crime in a two year period. Thus, while it **is** true that more blacks than whites were arrested for a reoffense, this does not mean that the same holds for actually committing a repeat reoffense.

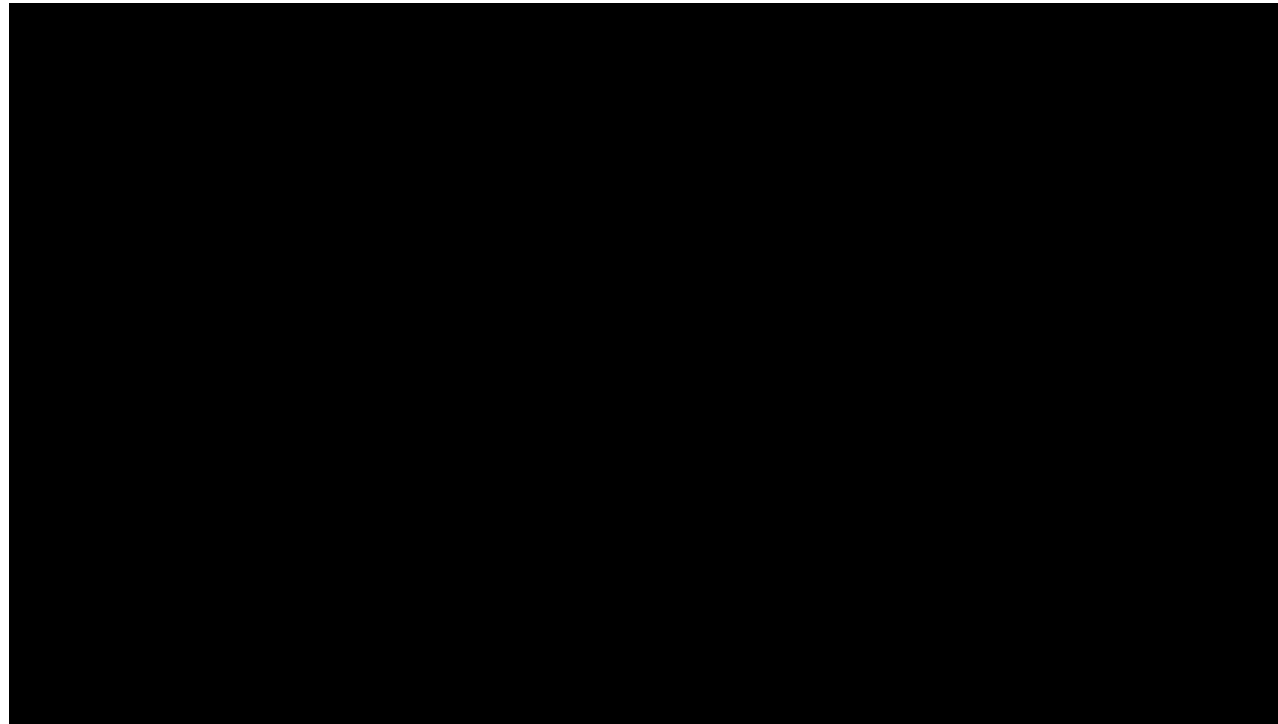
Reason #1: Prevalence of proxy differs in grps



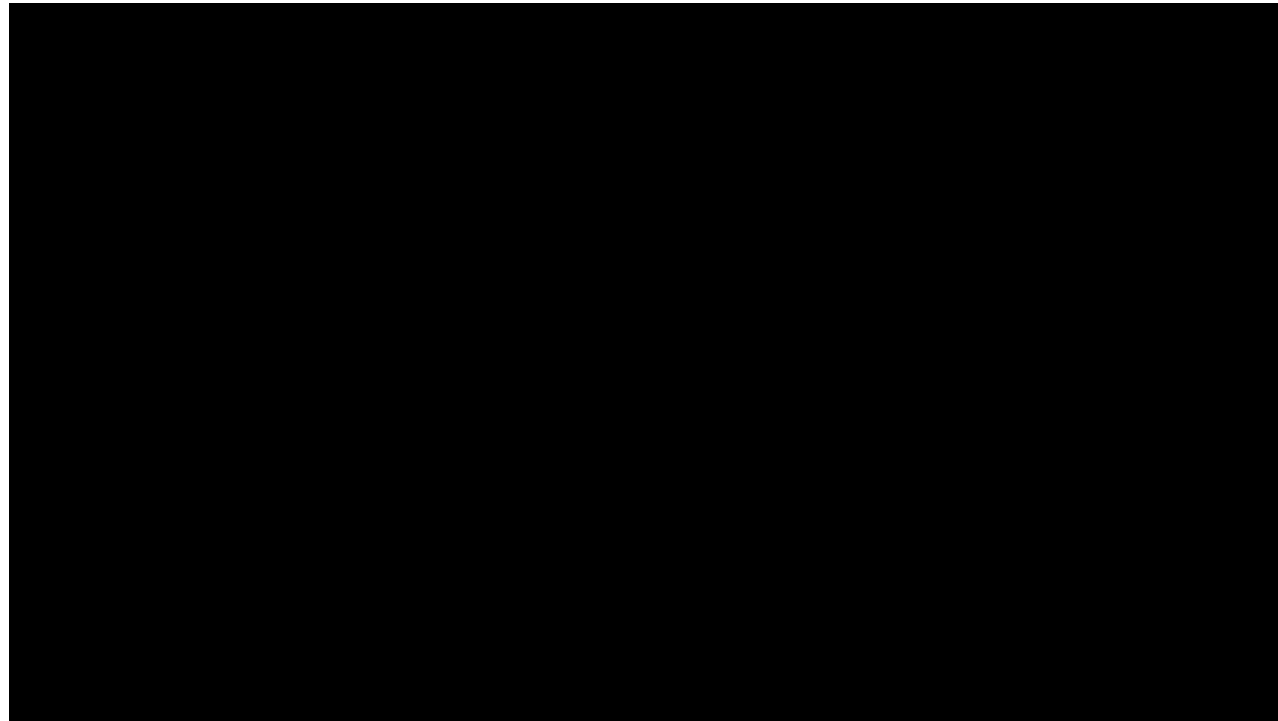
Another example



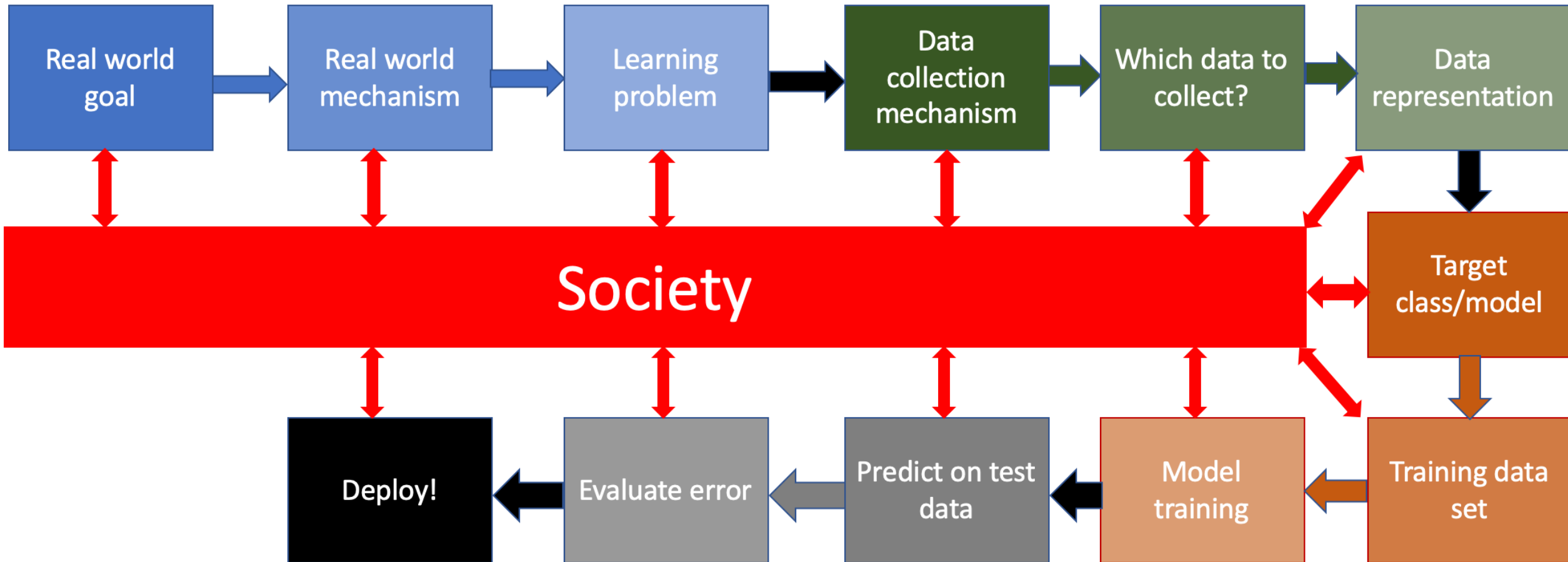
Reason #2: proxy is an over-simplification



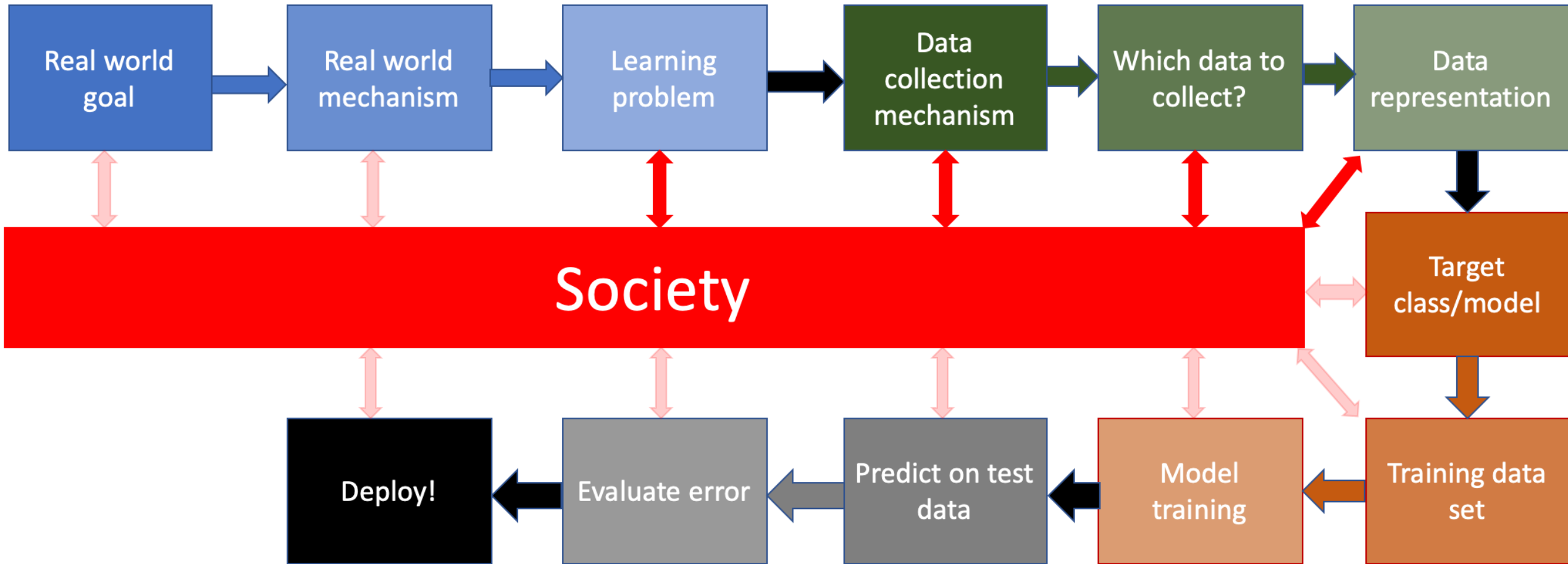
Example: Predictive policing



Back to the ML pipeline



What are the relevant interactions?





How do we handle measurement bias?

How we handle measurement bias in the ML pipeline?

Unlike the case of [handling historical bias](#) and (to a lesser extent) [handling representation bias](#), handling measurement bias is something that **has** be handled in the ML pipeline since picking the proxy variables is fair and square part of the ML pipeline.

When picking the target and input variables, you should first check if any of the picked variable is a proxy for the actual variable you are trying to measure. If so, you should re-think about your variable choice. Perhaps another variable is closer to the actual quantity you are trying to measure? If you cannot think of a better alternative you should be cognizant of the measurement bias you are introducing and at the very least this downside of the ML pipeline should be made clear to the actual users of the pipeline.



Passphrase for today: **Danielle Citron**



[Home](#) [Publications](#) [Media](#) [Professional Activities](#) [Bio](#) [Blog](#) [Contact](#) [Search](#)



Forbes



Bio

Danielle Citron is a Professor of Law at the Boston University School of Law where she teaches and writes about information privacy, free expression, and civil rights. She previously taught at the University of Maryland Carey School of Law where she received the 2018 “UMD Champion of Excellence” award for teaching and scholarship. Professor Citron has been a Visiting Professor at Fordham University School of Law (Fall 2018) and George Washington Law School (Spring 2017). In the future, she will do visiting stints at the University of Chicago School of Law and Harvard Law School.



TED Talk: How Deepfakes Undermine Truth and