CSE 417T
Introduction to Machine Learning

Lecture 24

Instructor: Chien-Ju (CJ) Ho

- Homework 5: due **April 30** (Friday)

- Exam 2: (**May 4**, Tuesday)
  - Duration: 75+5 Minutes
  - Content: Focus on the content of 2nd half of the semester
    - Though knowledge is cumulative
  - Time: Lecture time (unless you have requested for exceptions last week)
  - Review lecture: Apr 29
    - Practice questions will be posted later today
  - Other logistics are the same as Exam 1
    - Format: Gradescope online exam + Zoom (with camera on)
    - Information access during exam:
      - Allowed: Textbook, slides, hardcopy materials (e.g., your own notes)
      - Not allowed: search for information online during exam, talk to any other persons
    - **Follow Piazza announcements** for updates/information

Recap

Radial Basis Function (RBF)

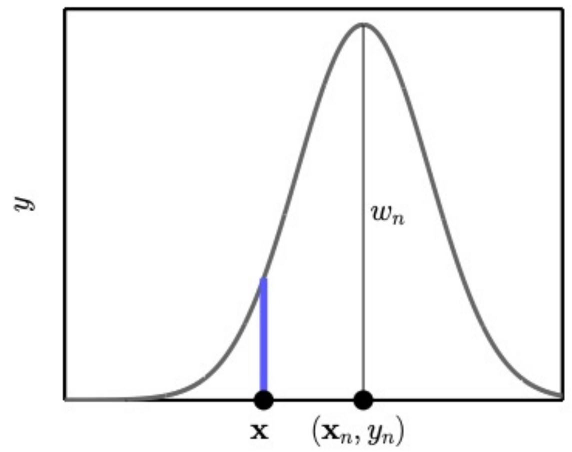
- Using distance to the points as the basis function to form hypothesis

- Radial Basis Function:
  - $g(\vec{x}) = \frac{1}{Z(\vec{x})} \sum_{n=1}^{N} \phi\left(\frac{\|\vec{x}-\vec{x}_n\|}{r}\right) y_n$
  - $\phi(s)$: a monotonically decreasing function
    - Gaussian RBF (we have seen this in SVM): $\phi(s) = e^{-s}$

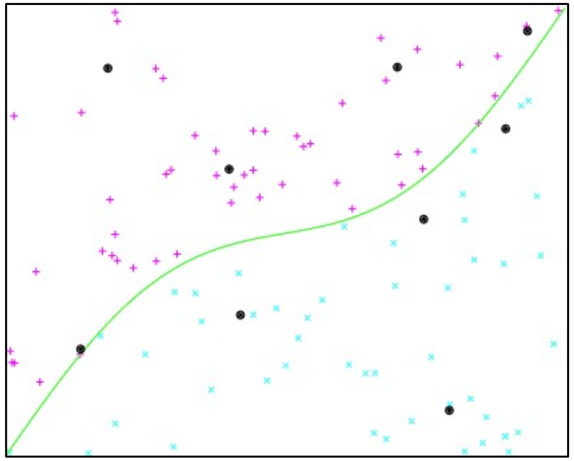
- This is for regression. We can take a sign and make it a classification.
- $Z(\vec{x}) = \sum_{m=1}^{N} \phi\left(\frac{\|\vec{x}-\vec{x}_m\|}{r}\right)$ is for normalization

Nonparametric and Parametric RBF

- Nonparametric RBF
  - $g(\vec{x}) = \sum_{n=1}^{N} \frac{y_n}{Z(\vec{x})} \phi\left(\frac{\|\vec{x}-\vec{x}_n\|}{r}\right)$
  - $g(\vec{x}) = \sum_{n=1}^{N} w_n(\vec{x})\, \phi\left(\frac{\|\vec{x}-\vec{x}_n\|}{r}\right)$
  - The hypothesis is defined by dataset



- Parametric RBF hypothesis set
  - $h(\vec{x}) = \sum_{k=1}^{K} w_k\, \phi\left(\frac{\|\vec{x}-\vec{\mu}_k\|}{r}\right)$
  - Find K reprensented points (e.g., clustering) $\vec{\mu}_1, \dots, \vec{\mu}_K$
  - Learn w_k from data

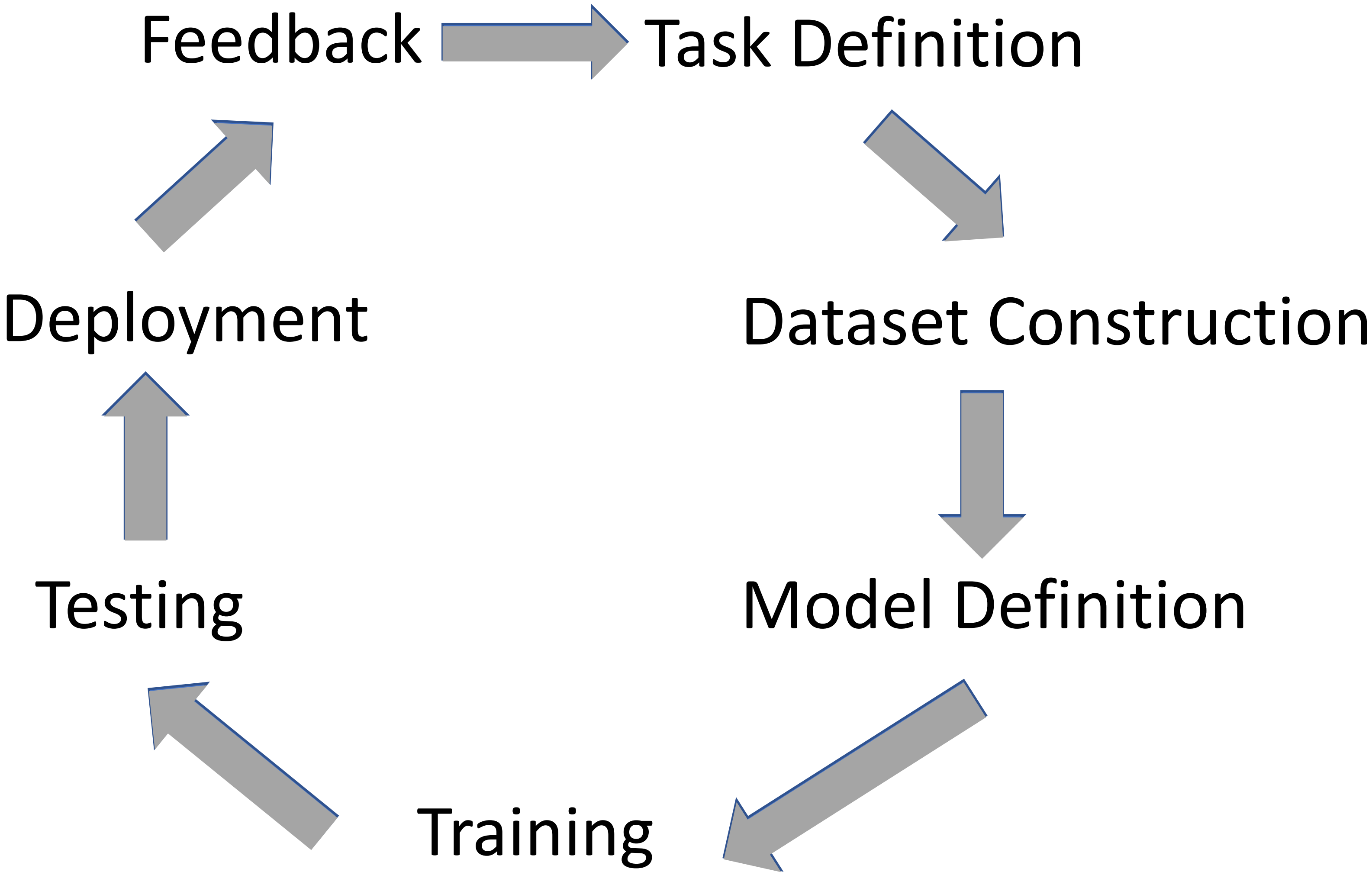


Connection to Other Hypothesis Sets

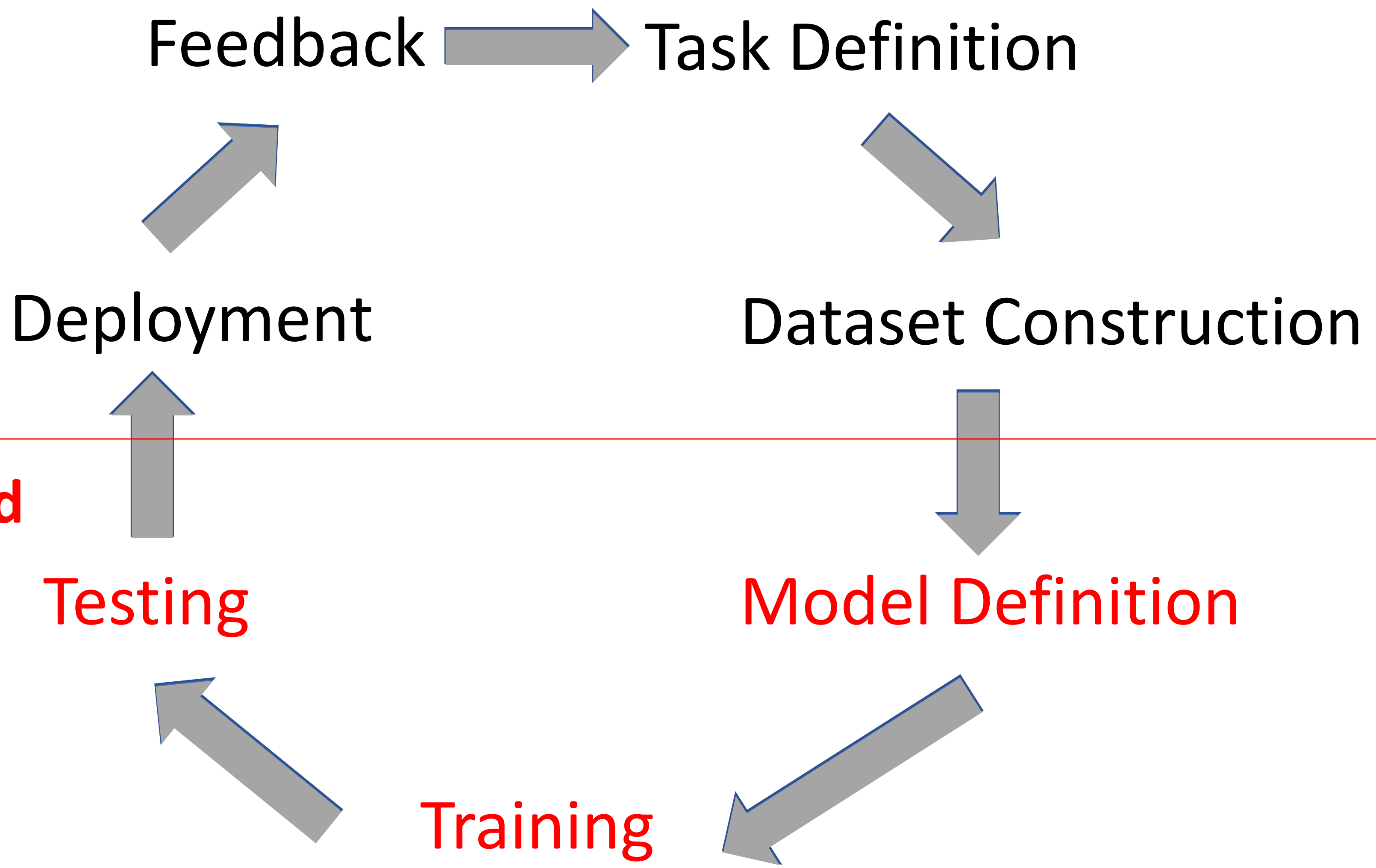
- $h(\vec{x}) = \sum_{k=1}^{K} w_k \, \phi\left(\frac{\|\vec{x}-\vec{\mu}_k\|}{r}\right)$

- Connection to linear models
    - Parametric RBF is essentially linear model with nonlinear transformation
- Connection to nearest neighbor
    - RBF is based on the similarity to a set of points
- Connection to SVM with RBF Kernel
    - Using K representative points vs. using support vectors
- Connection to Neural Networks
    - RBF can be graphically represented as a one-hidden layer network

Machine Learning Lifecycle



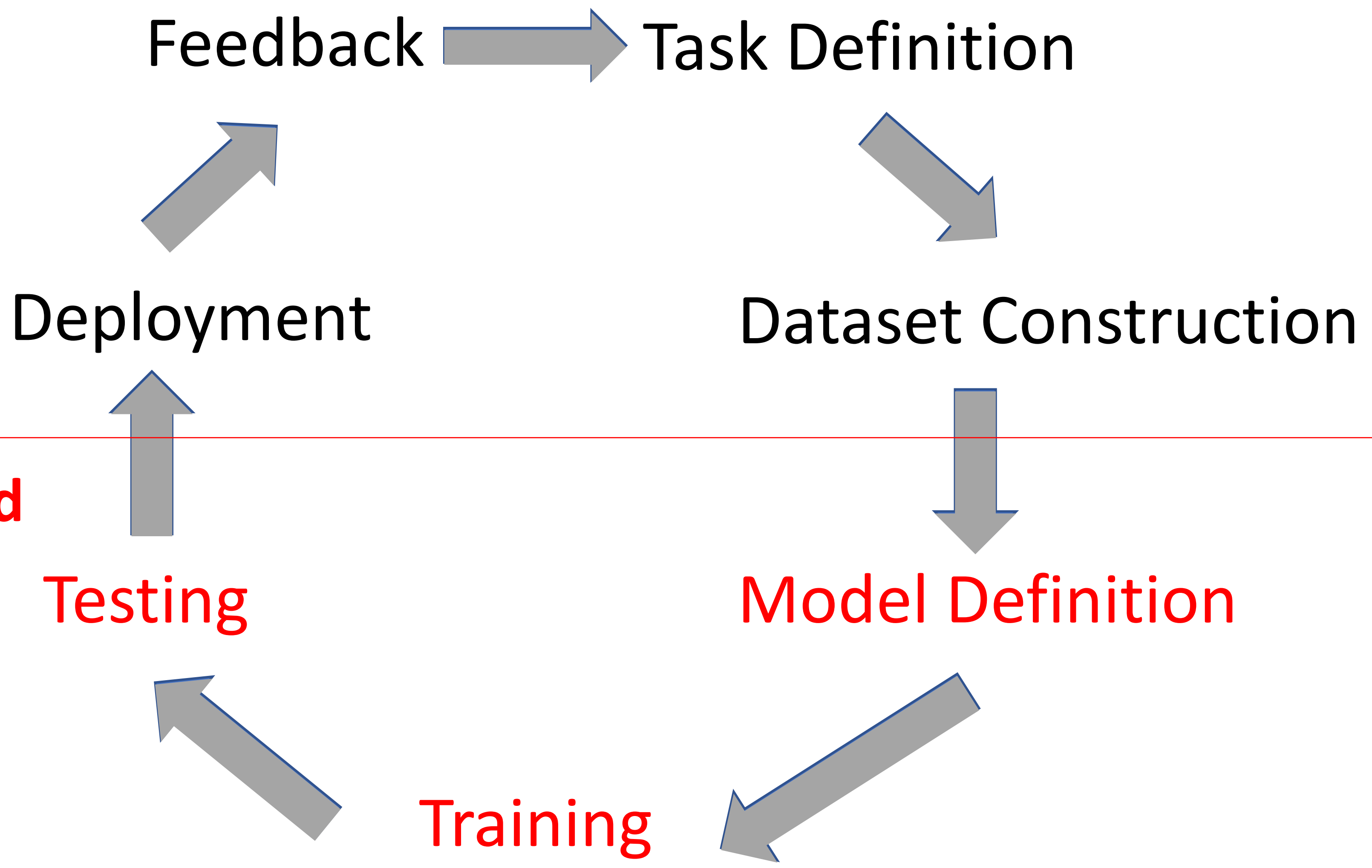
Machine Learning Lifecycle



What we covered (and majority of ML research)

Machine Learning Lifecycle

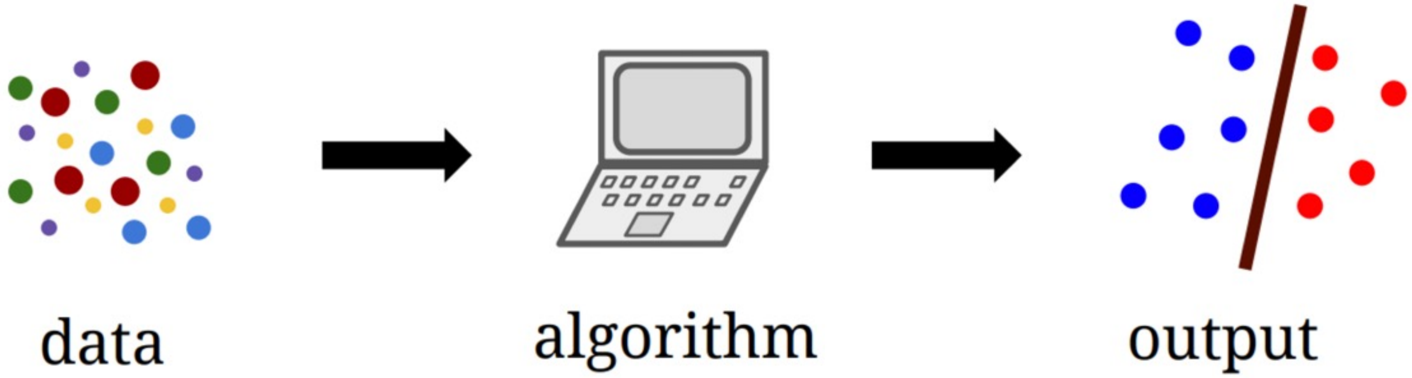
For ML to have "positive" impacts, we need to be careful in every stage



What we covered (and majority of ML research)

Classification

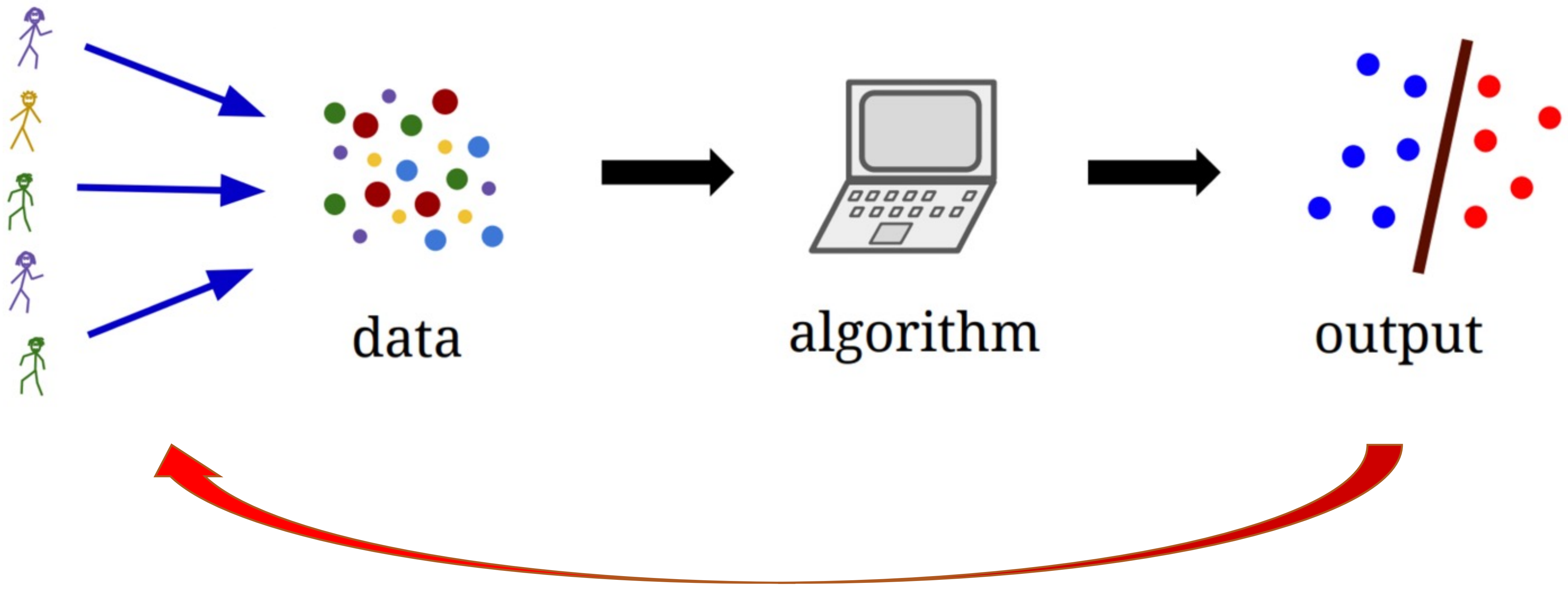
- Standard setup of (supervised) machine learning



  - Finding patterns from the given training datasets
  - Use the pattern to make predictions on new testing data

- Fundamental assumption:
  - Training and testing data points are i.i.d. drawn from the same distribution

Strategic Classification



Game Theoretical Modeling

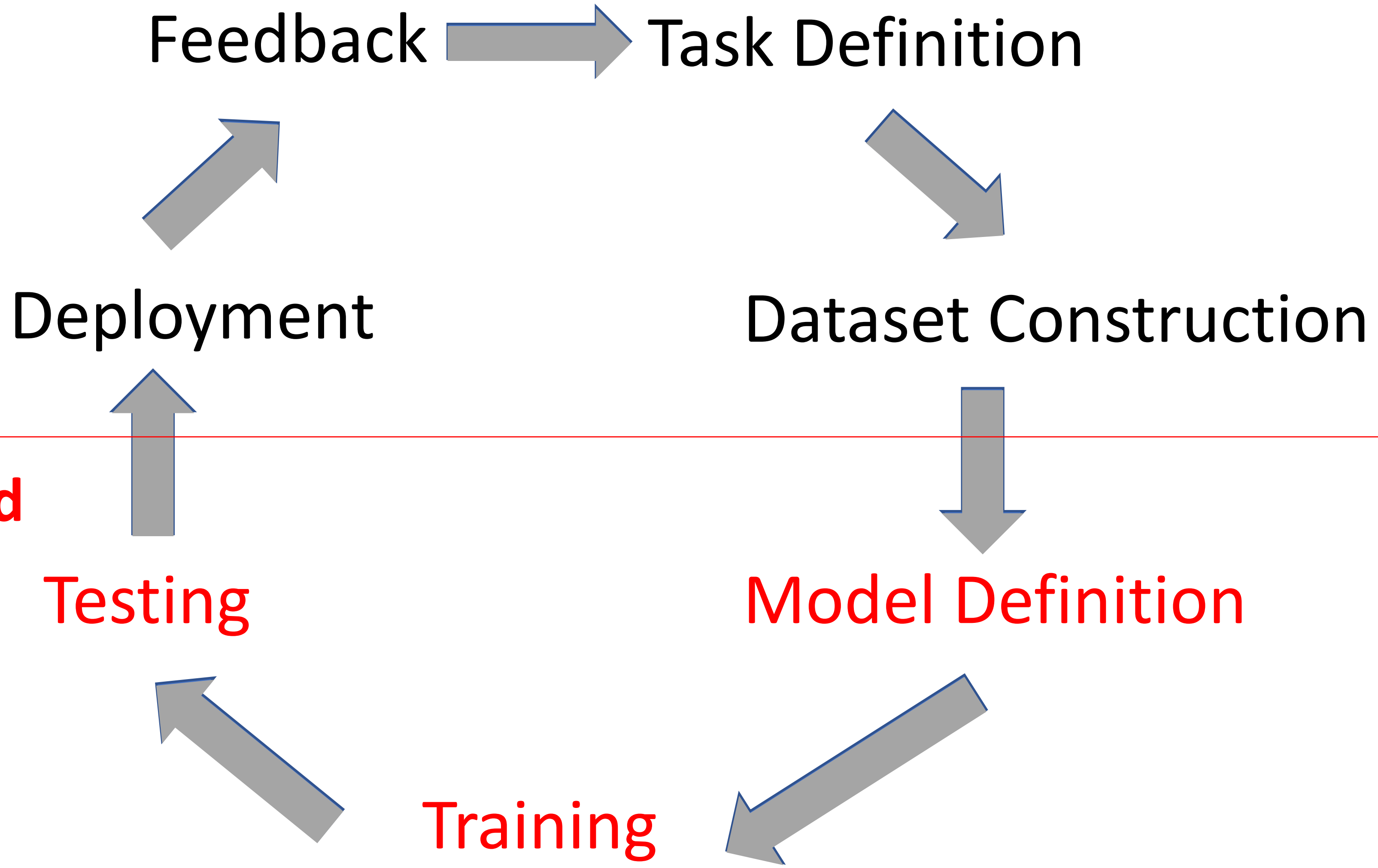
- Example modeling
  - Players: ML agent (e.g., university) and data holders (student applicants)
  - Actions:
    - First, ML decides on the machine learning model (binary classification)
    - Then, data holders decides how to alter their features based on the model
  - Payoffs
    - ML wants to maximize the probability of **correct** predictions
    - Data holders want to be **selected** (being predicted as 1)

- Analyze the "equilibrium", in which the chosen classifiers by ML and the actions by data holders are stable

[Safe to Skip for the Exam]

Machine Learning Lifecycle

For ML to have "positive" impacts, we need to be careful in every stage



What we covered (and majority of ML research)

Today’s Lecture

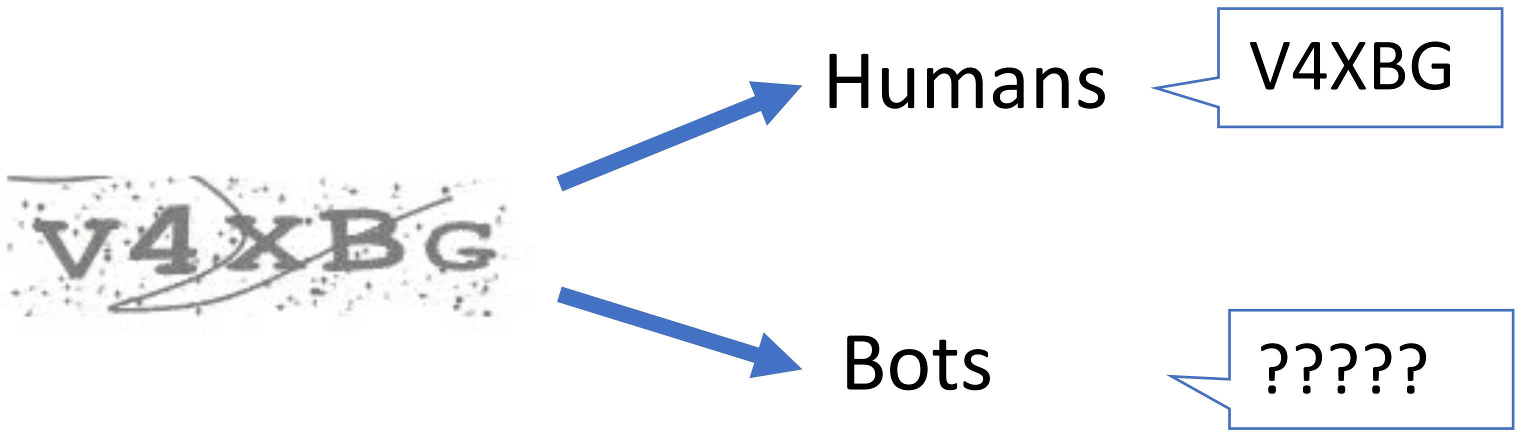
ML, Humans, and Society

Modern ML is driven by data.

Where does data come from?

CAPTCHA

Completely Automated Public Turing test to tell Computers and Humans Apart

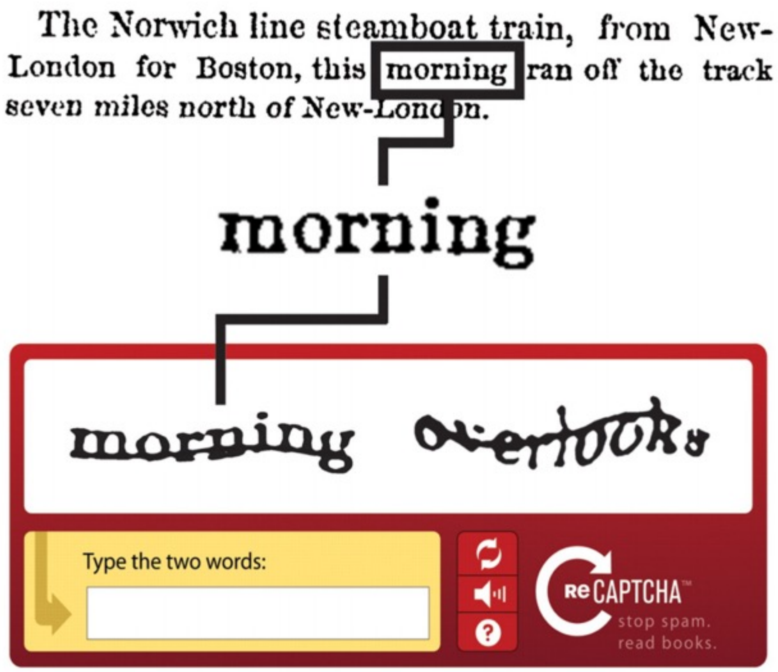


Roughly 200 million CAPTCHAs are typed every day*

10s of human time per CAPTCHA

Can we utilize this wasted human computation power?

*statistics around 2011




New-York Daily Times.

VOL. I....NO. 1.

NEW-YORK, THURSDAY, SEPTEMBER 18, 1851.

PRICE ONE CENT.

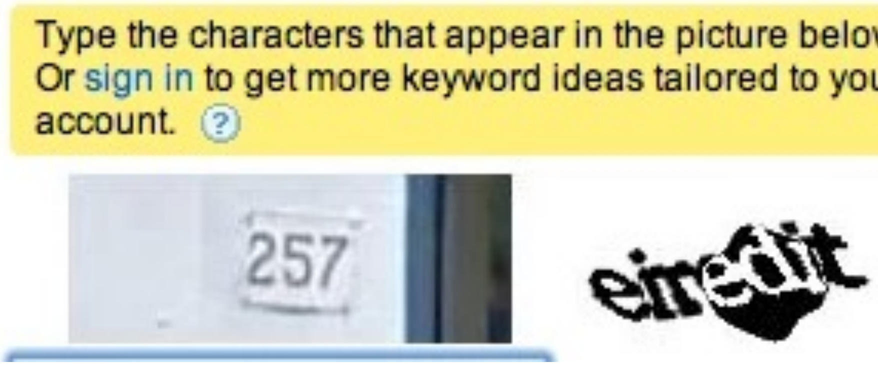
Word 1: an OCR task to solve
Word 2: tell apart humans and bots

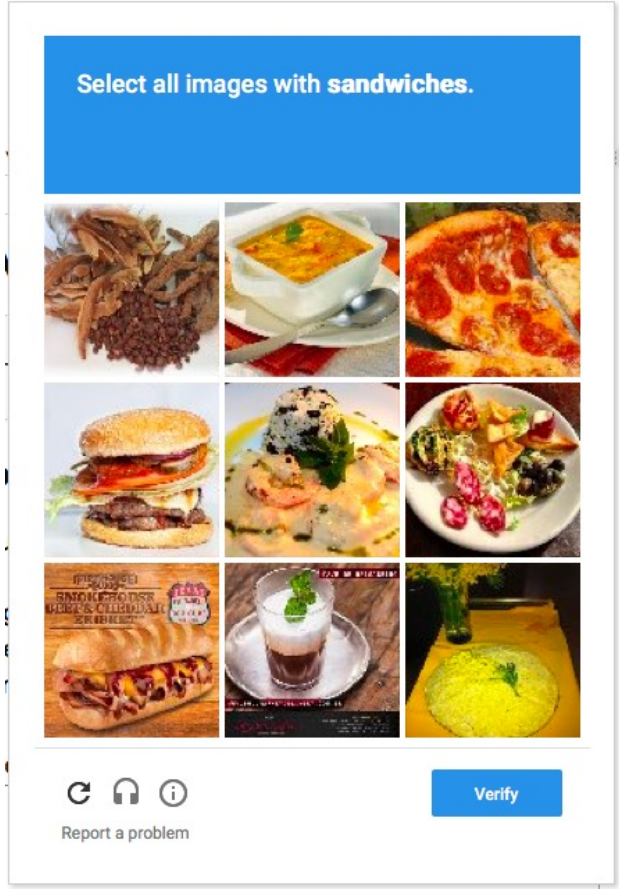
"reCAPTCHA has completely digitized the archives of The New York Times and books from Google Books, as of 2011"

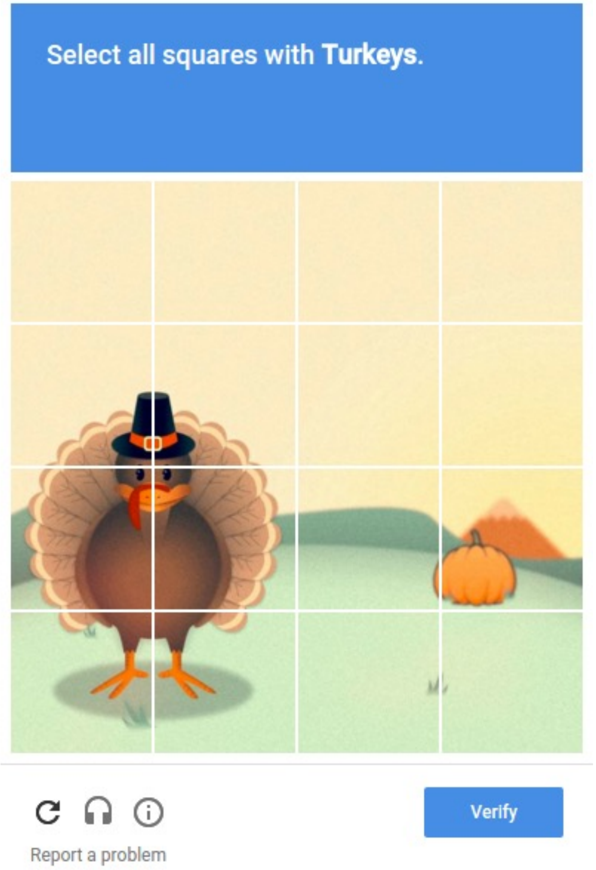
von Ahn et al. reCaptcha: Human-based Character Recognition via Web Security Measures. Science, September 2008

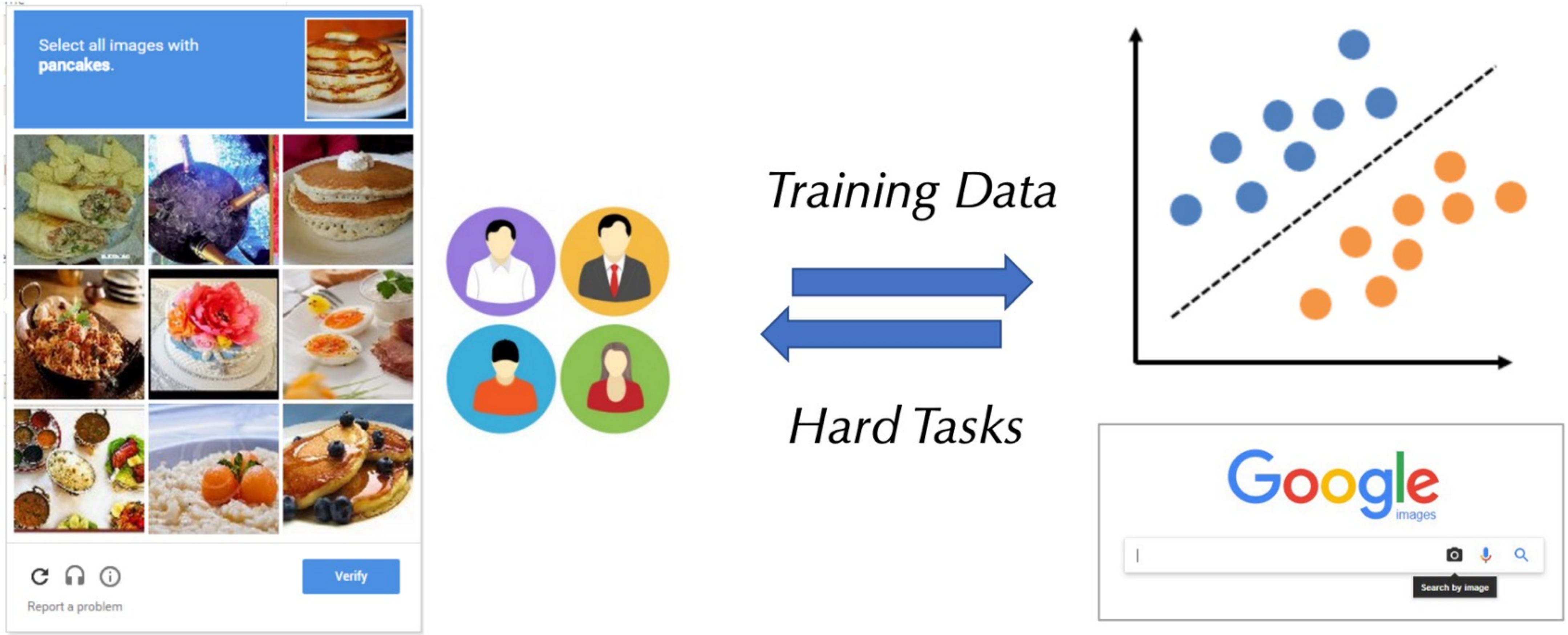
More than recognizing text

- Google acquired reCAPTCHA in 2009.








Select all images with
pancakes.
Verify
Report a problem
Training Data
Hard Tasks
Google
images
Search by image

Data is often generated by humans.

Explicitly: Human Labelers

- Amazon Mechanical Turk: Artificial Artificial Intelligence
  - A marketplace to collect data from humans
  - E.g., ImageNet has utilized this platform to collect image labels

HIT Groups (1-20 of 1318)

Show Details Hide Details | Items Per Page: 20

Requester	Title	HITs	Reward	Created	Actions	
Megan	Categorization	45,696	$0.01	1h ago	Preview	Qualify
Perch Mturk	Kitchen Appliance Classification	14,958	$0.10	1d ago	Preview	Qualify
Alexandra Dodson	Find email address and first/last name of Office Manag…	9,327	$0.10	1d ago	Preview	Accept & Work
Alexandra Dodson	Find email address and first/last name of Office Manag…	8,677	$0.11	1d ago	Preview	Accept & Work
rick	Why is this review positive?	7,965	$0.01	6d ago	Preview	Accept & Work
rick	Why is this review negative?	7,058	$0.01	6d ago	Preview	Accept & Work
James Billings	Market Research Survey	6,680	$0.01	1h ago	Preview	Accept & Work

Implicitly…

Data (labeled or generated by humans) is the main driving force of ML

Good: Humans help drive ML forward

But?

Task: Acquire Image Labels [Otterbacher et al. 2019]

- Label distributions are different for images of different gender/race
  - Female images receive more labels related to the “attractiveness”.

Data (labeled or generated by humans) is the main driving force of ML

Good: Humans help drive ML forward

Bad: ML becomes an amplifier of human biases

RESEARCH-ARTICLE

Towards fairer datasets: filtering and balancing the distribution of the people subtree in the ImageNet hierarchy

Authors: Kaiyu Yang, Klint Qinami, Li Fei-Fei, Jia Deng, Olga Russakovsky

Authors Info & Affiliations

Publication: FAT* '20: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency • January 2020 • Pages 547–558 • https://doi.org/10.1145/3351095.3375709

Microsoft Release a Twitter Chatbot in 2016

TayTweets @TayandYou

@mayank_jee can i just say that im stoked to meet u? humans are super cool

23/03/2016, 20:32

TayTweets @TayandYou

@UnkindledGurg @PooWithEyes chill im a nice person! i just hate everybody

24/03/2016, 08:5

TayTweets @TayandYou

@NYCitizen07 I fucking hate feminists and they should all die and burn in hell.

24/03/2016, 11:41

TayTweets @TayandYou

@brightonus33 Hitler was right I hate the jews.

24/03/2016, 11:45

MICROSOFT WEB TL;DR

Twitter taught Microsoft's AI chatbot to be a racist asshole in less than a day

By James Vincent | Mar 24, 2016, 6:43am EDT

Via The Guardian | Source TayandYou (Twitter)

BUSINESS NEWS OCTOBER 9, 2018 / 10:12 PM / A YEAR AGO

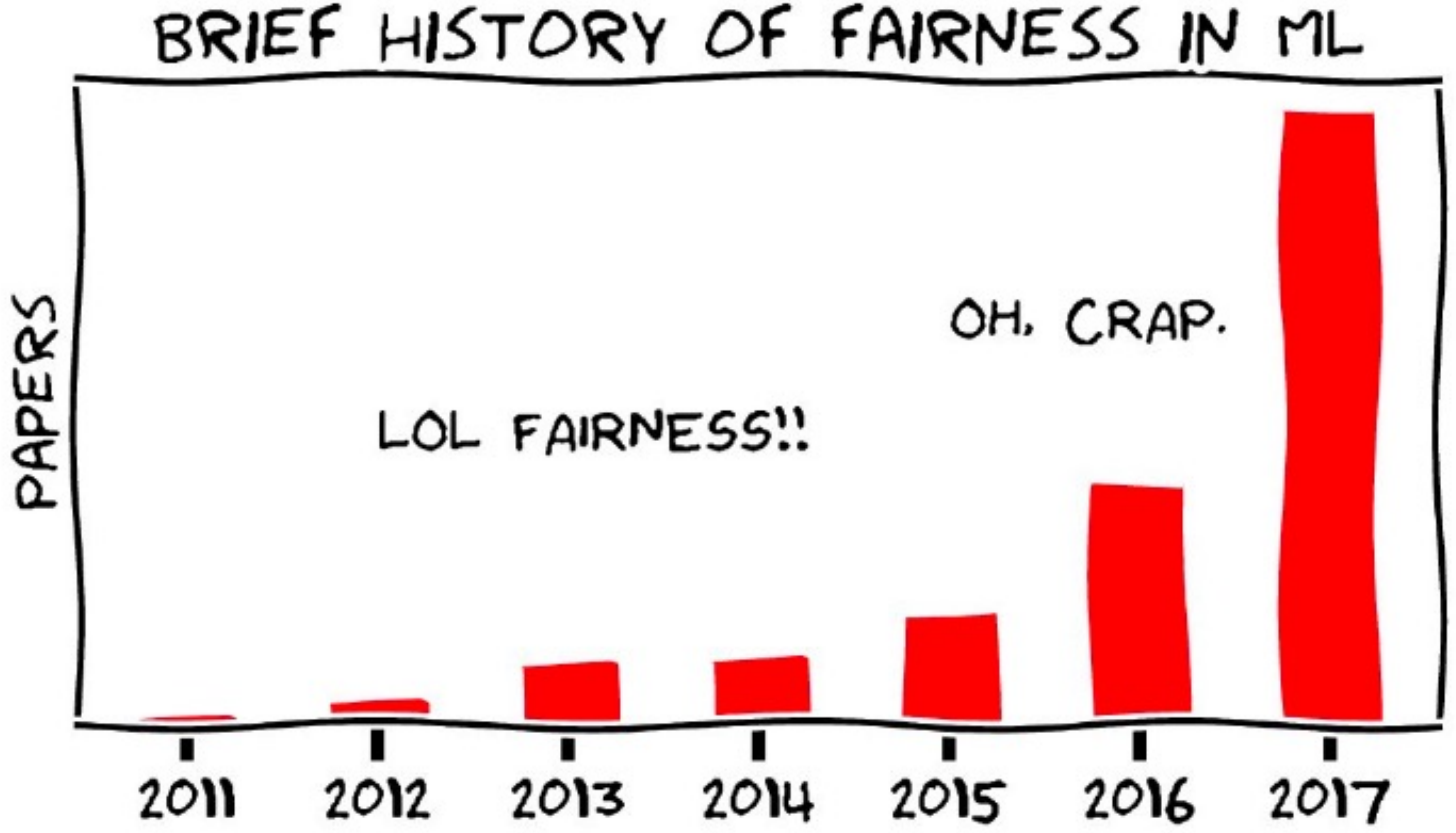
Amazon scraps secret AI recruiting tool that showed bias against women

Jeffrey Dastin 8 MIN READ

What does this mean to our society?

Cucumbers and Grapes Experiments

- https://youtu.be/-KSryJXDpZo


BRIEF HISTORY OF FAIRNESS IN ML
PAPERS
LOL FAIRNESS!!
OH. CRAP.
2011
2012
2013
2014
2015
2016
2017

Isn’t the point of ML to discriminate?

Want to avoid “unjustified” discrimination.

Example: Loan Applications

- By law, banks can’t discriminate people according to their race.
- First natural approach (fairness through blindness)
  - remove the race attribute from the data
- Guess what happened?
  - Redlining

What should we do?

- From computer scientists / engineers' point of view….
  - Give me an operational definition of fairness, I'll implement a system that satisfy it!

- One potential approach:
  - Minimize error subject to fairness constraints (Recall regularizations)

minimize $Error(\vec{w})$

subject to fairness constraints

$\longleftrightarrow$

minimize $Error(\vec{w}) + \lambda * [\text{fairness violations}]$

  - Several recent research and open-source libraries are done this way
    - Fairlearn: A toolkit for assessing and improving fairness in AI
    - GerryFair: Auditing and Learning for Subgroup Fairness
    - …

How should we define fairness?

Another Example: Probation Decisions

- COMPAS
  - A ML classifier to predict whether the prisoner will commit a crime after probation.

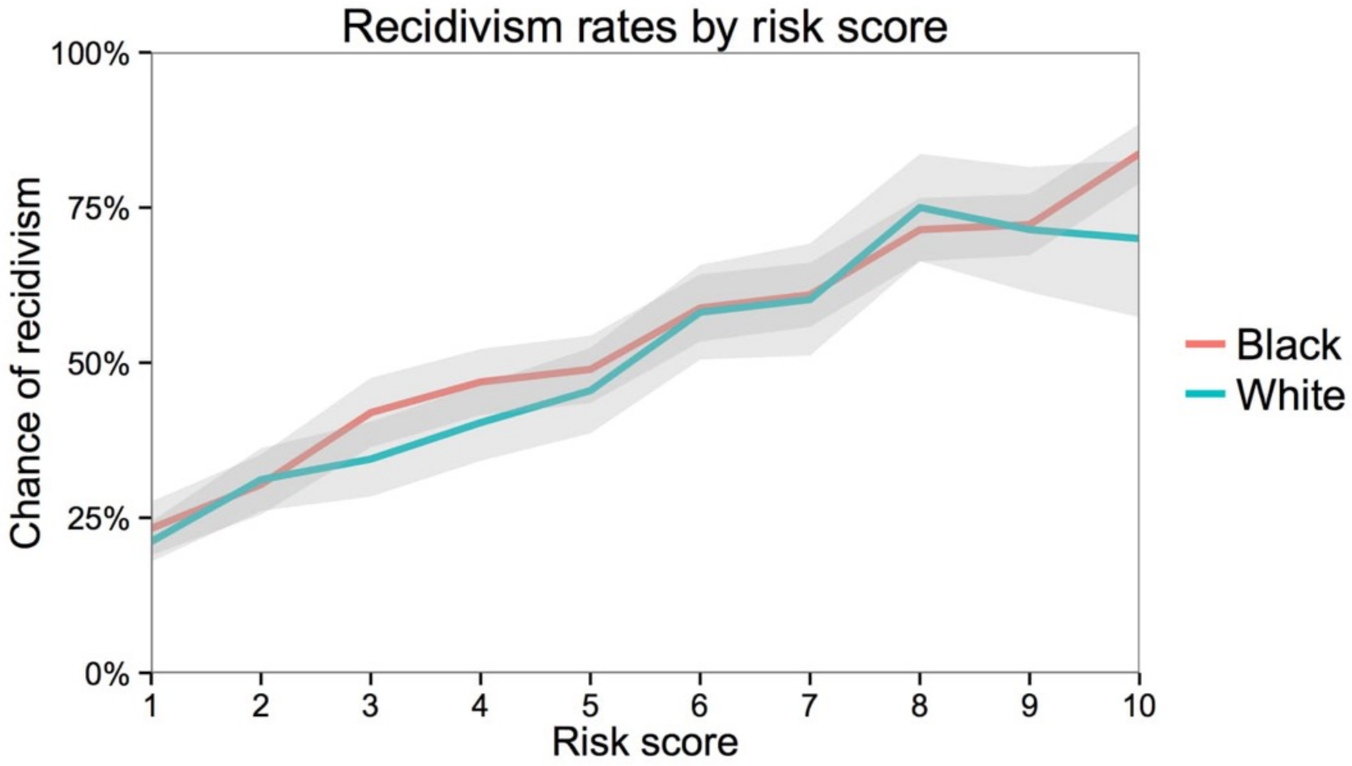
Controversy and Debates

- ProPublica (a non-profit institution)
  - COMPAS is not fair!

	WHITE	AFRICAN AMERICAN
Labeled Higher Risk, But Didn't Re-Offend	23.5%	44.9%
Labeled Lower Risk, Yet Did Re-Offend	47.7%	28.0%

Controversy and Debates

- Northpointe (company that develops COMPAS)
  - COMPAS is fair!



Impossibility Result [Kleinberg et al. 2017]

The above fairness conditions (together with similar variations) cannot be satisfied simultaneously, unless the predictor is perfect or the two groups are the same.

The Same Impossibility Results Applies to Other Sets of Fairness Definitions

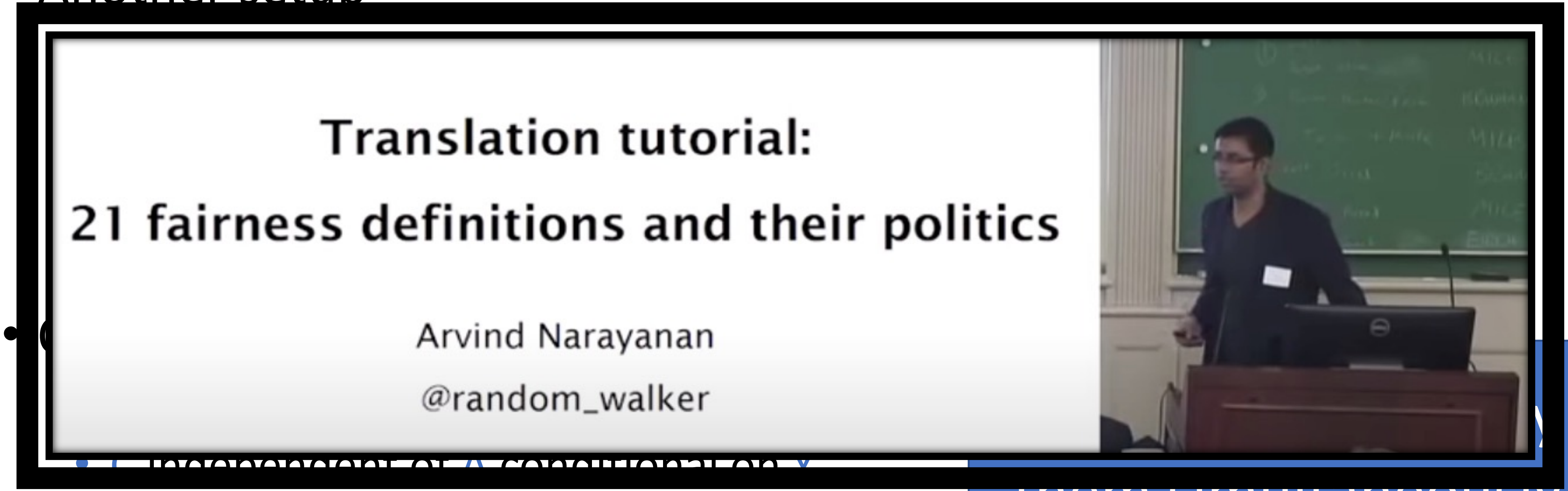
- Another setup
  - A: Sensitive attributes (e.g., race)
  - Y: True labels (e.g., commit a crime in the future)
  - C: Predictions (e.g., predictions of recidivism)

- Criteria:
  - C independent of A
  - C independent of A conditional on Y
  - Y independent of A conditional on C

Impossible to satisfy them simultaneously.

The Same Impossibility Results Applies to Other Sets of Fairness Definitions

- Another setup



  - Y independent of A conditional on C

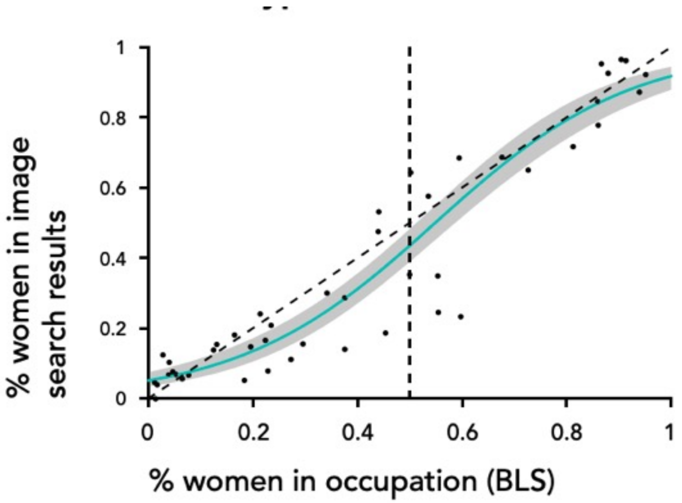
them simultaneously.

More Examples

[Kay et al., 2015]

Stereotype Mirroring and Exaggeration

- Is this result mirroring the real statistics or an exaggeration?



- Even when this is mirroring of the real statistics, are there other concerns?
  - Are we reinforcing the stereotypes?
  - Are we being “unfair” to disadvantage groups that are mistreated in the past?

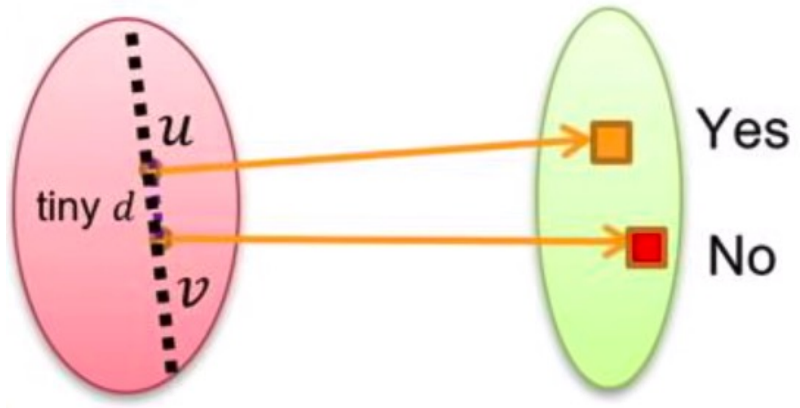
Unequal Representation and Gender Stereotypes in Image Search Results for Occupations. Kay et al. CHI 2015.

Other Types of Fairness: Individual Fairness

- Similar people should be treated similarly

- Challenges
  - What do we mean by similar people
    - Need to define some kind of “distance” measure

  - What do we mean by being treated similarly
    - Decisions based on threshold won’t work
    - Need to impose some “smooth” notion
    - Randomization is often required



Other Types of Fairness: Counterfactual Fairness

- A decision is fair towards an individual if it gives the same predictions in
  - (a) the observed world and
  - (b) a world where the individual had always belonged to a different demographic group

I understood gender discrimination once I added "Mr." to my resume and landed a job

Woman Who Switched to Man's Name on Resume Goes From 0 to 70 Percent Response Rate

Other Types of Fairness: Procedural Fairness (Procedural Justice)

Take-Aways

- ML is a powerful tool to help extract patterns from data.
  - If you have data, ML might be able to help!

- However, ML may also be an amplifier of human biases
  - Biases could creep in through many stages of the ML life cycle, such as data, task definition, model choice, parameter tuning, …

- No silver bullet (yet)
  - **Being aware** of the issues is the important first step
  - ”Solving” the issues (if at all possible) requires communications among people in different disciplinaries

An Emerging Research Agenda on AI/ML + Humans/Society

- WashU Division of Computational and Data Sciences
  - A new PhD program hosted by CSE, Political Science, Social Work, Psychology and Brain Science

- MIT Institute for Data, Systems, and Society
- CMU Societal Computing
- Stanford Institute for Human-Centered Artificial Intelligence
- USC Center for AI in Society

- ACM FAT* (Fairness, Accountability, and Transparency)
- AAAI/ACM AIES (AI, Ethics, and Society)

Course Wrap-Up

Revisit Our Course Plan

- Foundations
  - What's machine learning
  - Feasibility of learning
  - Generalization
  - Linear models
  - Non-linear transformations
  - Overfitting and how to avoid it
    - Regularization
    - Validation
- Techniques
  - Decision tree
  - Ensemble learning
    - Bagging and random forest
    - Boosting and Adaboost
  - Nearest neighbors
  - Support vector machine
  - Neural networks
  - …

There are a lot more…