Bias and Ethics in Web Systems

Review: Advertisements and Auctions

- **Advertisement** systems consist of a **seller** that **advertisers** pay for **user** time and interaction
  - **Advertiser** cares about ad quality, metrics, and maximizing return
    - Target demographics, click-through rate, cost to do business, new sales volume, etc.
  - **Seller** cares about balancing **user** needs with **advertiser** needs
    - Don't want to alienate the user with garbage ads
    - Want to provide quality platform to attract high-paying advertisers
  - **User** cares about free stuff!
    - If ads are too distracting, users stop or move to another platform (#bing)
    - User may evade ads with AdBlock'ers
- **Seller** offers a framework for **auctions**
  - **Advertisers** bid on the chance to display an ad
  - **Vickrey** auction consists of *blind-bid* and *second-price*
    - (for economic efficiency)

Auctions for Web Systems

- The **user** types a query on Google (the **seller**)

- The **seller** goes to market and tells the **advertisers:**
  "I just got a sweet query from this user. They searched for X and have demographic properties Y. How much do you want to pay to display your ad for them?"

- **Advertisers** compete based on
  - Amount willing to pay
  - What type of payment they want (i.e., per click? Per view? Per action?)
  - Ad quality
  - Ad relevance to user

Example of strategic behavior

- Auction format: 3 buyers
  - buyer 1 names a price, then buyer 2, then 3.

	B1	B2	B3
Willing to Pay	$100	$500	$400
Bid	$100	$150	$151
Why?	(must compete with B2, B3)	(thinks B3 will only go for $125)	(lol rekt)

Why was this bad?

- **Strategic behavior**: everyone had to do a lot of thinking about the other
- **Missing Information**: The outcome of the auction would change if performed again with the same knowledge
- **Misallocation**: At the end of the auction, the “wrong” person won. Buyer 2 should talk to Buyer 3 to buy it from them.
- **No list price:** The **seller** could have gotten much more money

Sealed-bid first price auction

- Everyone submits a bid, doesn't tell the others
- The person with the highest bid pays what they bid

	B1	B2	B3
Willing to Pay	$100	$500	$400
Bid	$100	$500	$400
Why?	Incomplete information… bid for what you're willing		

Sealed-bid first price auction

- What happens if we have the auction many times?

	B1	B2	B3
Willing to Pay	$100	$500	$400
Bid	$100	$500	$400
Bid 2	$100	$499	$400
Bid 3	$100	$498	$400
…			
Bid n	$100	$401	$400

- B2's strategy can be to keep lowering the price until they find the minimum they need to keep beating the others
- **Consequence**: buyers bid less than they are willing to pay

Sealed-bid second-price

- Get same long-term outcome, but incentive to bid true value and no strategic thinking needed

	B1	B2	B3
Willing to Pay	$100	$500	$400
Bid	$100	$401	$400
Why?	B2 pays $400 (the “second price”)		

Google advertising auction

- When a search query comes in, Google finds *relevant* ads, then considers the **bid** and **quality score** of each relevant **advertiser**, ranks the ads, and selects the one that ranks the highest

- Small example (ford pays $3.01)

	Bid	Quality Score	Ad shown?
Ford	$5	10	Shown
GM	$3	10	Not shown
Chevrolet	$1	10	Not Shown

Google search advertising auction

- Consider (CPC = cost per click)
  - CPC = $\frac{\text{ad rank of next ad}}{\text{your ad's quality score}} + \$0.01$

Ad #	Advertiser	Max bid	Quality Score	Ad Rank	CPC
1	Ford	$3.00	8	24	$\frac{20}{8} + .01 = \$2.51$
2	GM	$4.00	5	20	$\frac{15}{5} + .01 = \$3.01$
3	Chevrolet	$5.00	3	15	$\frac{12}{3} + .01 = \$4.01$
4	Toyota	$6.00	2	12	N/A

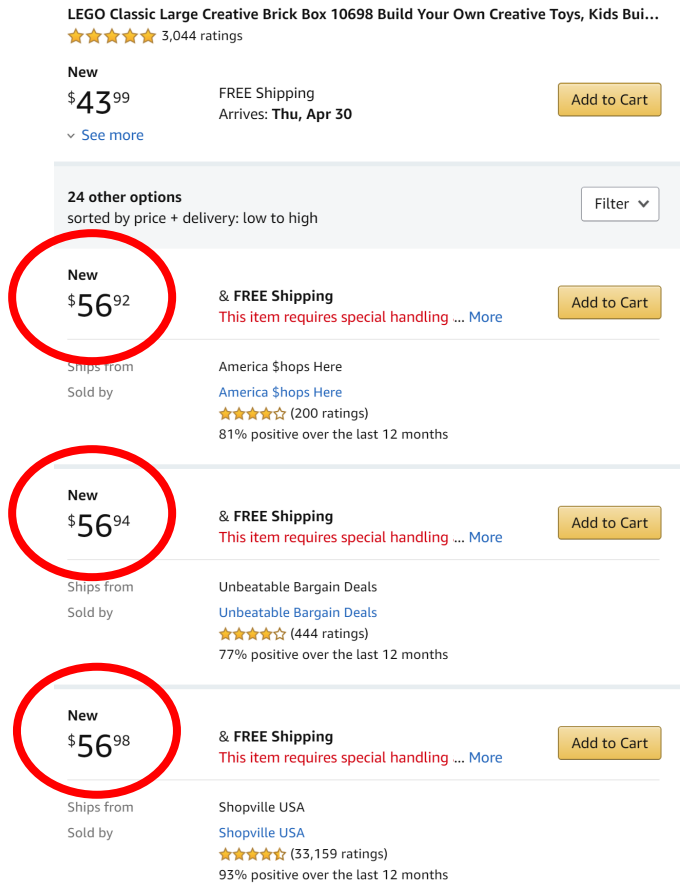
- Higher quality ads are rewarded by paying less

https://support.google.com/adsense/answer/160525?hl=en

Thought questions

- Amazon has an "Automate Pricing" feature for sellers where it will set the seller's price for an item to automatically be lower than the next competitor's. Is this more similar to a first-price or second-price auction?

- Do you think it would be more profitable to spend less money on an ad for a rare search term (like LEGO Set 10698) or more money for a common search term (like LEGO)?



One Slide Summary: Ethics and Bias

- **Ethics** are guiding principles for acceptable conduct within a society
  - It is ethical to protect user information
  - It is unethical to cheat on exams
  - It is ethical to disclose conflicts of interest
- In contrast, **morality** speaks to personal values for making judgments
  - "I don't *need* to protect user information; I don't think I'm collecting sensitive information."
- **Ethics** reach a variety of topics in web systems**:**
  - **Privacy**: "The Web remembers" – Public disclosures stick around forever
    - Waybackmachine – remember those embarrassing angsty posts you made in middle school?
  - **Informed Consent:** if we collect data about people, we must let them know what's being collected and why
    - Milgram experiment? Stanford Prison? Tuskegee?
  - When do we **delete data**? When is it acceptable to **sell data**?
  - **Bias in Web Systems:** Should we "adjust" predictions made by ML-dependent systems?

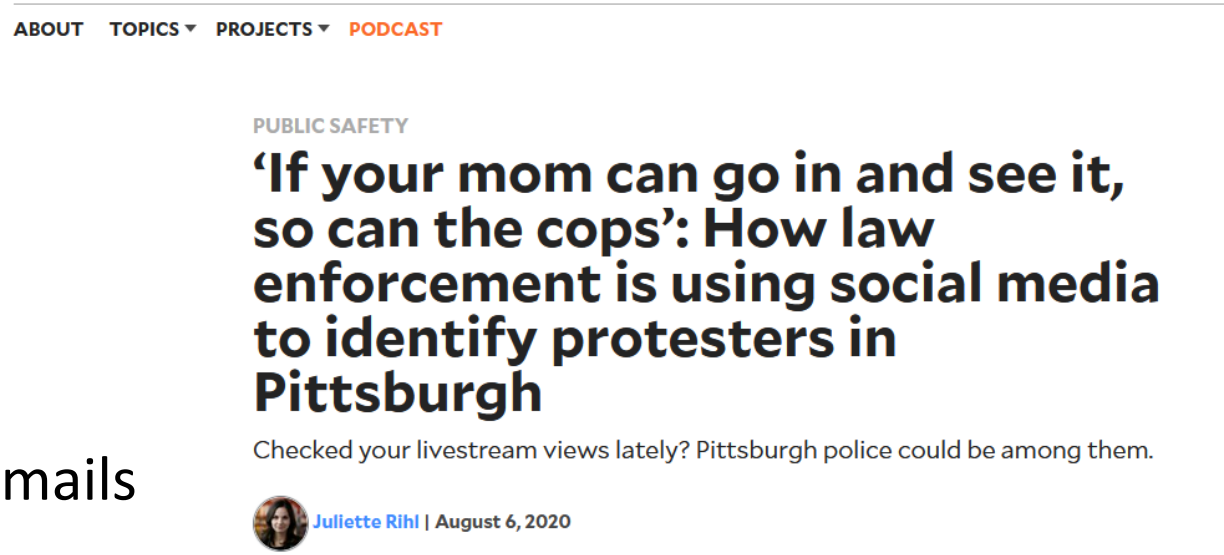
Privacy

- Ability to control sharing of information about oneself
- Basic human need?
  - Even for people who have “nothing to hide”
  - Generally “ethical” to preserve user privacy

Privacy

- "Obvious" examples
  - Facebook messages, Instagram posts, E-mails
  - Piazza posts? Slack DMs?

- But also **metadata**
  - Recipient
  - Subject line
  - Length of conversation
  - The time frame in which a conversation took place
  - Your location when communicating



Privacy and metadata

- They know you rang a phone sex line at 2:24 am and spoke for 18 minutes.
  But they don't know what you talked about.
- They know you called the suicide prevention hotline from the Golden Gate Bridge.
  But the topic of the call remains a secret.
- They know you got an email from an HIV testing service, then called your doctor, then visited an HIV support group website in the same hour.
  But they don't know what was in the email or what you talked about on the phone.
- They know you received an email from a digital rights activist group with the subject line “Let’s Tell Congress: Stop SESTA/FOSTA” and then called your elected representative immediately after.
  But the content of those communications remains safe from government intrusion.
- They know you called a gynecologist, spoke for a half hour, and then called the local abortion clinic’s number later that day.
  But they don’t know what happened.
- They know you went to CVS to fill your prescription for Suboxone.
  But they don’t know what you’re diagnosed with.

https://ssd.eff.org/en/module/why-metadata-matters

Loss of Privacy

- Privacy is **lost** when a person no longer **controls** information they consider *sensitive*

- Considerations
  - **What data** did the person allow you to collect about them?
    - **Important:** metadata counts
    - **Important:** how do you let them know? How often?
  - **In what way** did the person allow you do use the data?
    - **Important:** What if you need to change how you use the data?
    - **Important:** What if your organization exits the market?

Losing Privacy today: No exit

- In the past, one could get a fresh start by:
  - Moving to a new place
  - Waiting until the past fades
    - Reputations can be rebuilt over time
    - Credit scores rebuilt, etc.
    - Dutch might have a pLaN to start over in Valentine

- **Today**: Internet remembers everything
  - Careers destroyed
    - Remember your tasteless Halloween costume on FB?
  - Relationships ruined
    - How about the time you spoiled Half-Blood Prince for all your friends on Xanga?
  - Finances in shreds
    - “Post your SSN: MySpace automatically censors it!”

Internet archive: wayback machine

- Past versions of websites
- https://archive.org/web/

The Vicious Cycle of “Never Forgetting”

Amazon scraps secret AI recruiting tool that showed bias against women

Jeffrey Dastin

8 MIN READ

SAN FRANCISCO (Reuters) - Amazon.com Inc’s (AMZN.O) machine-learning specialists uncovered a big problem: their new recruiting engine did not like women.

- AI review of job applications
- Trained on last 10 years of Amazon employees ...
  - ... who were mostly men
- Penalized resumes that included the word “women’s,” as in “women’s chess club captain.”
- Downgraded graduates of two all-women’s colleges

Ethics

- Ethics are principles that help us distinguish right from wrong.
- Ethics are the basis for the rules we all voluntarily choose to follow because that makes the world a better place for all of us.
- Generally, professional organizations will have various **ethics standards**
  - Often coincide with morality, but not always

Ethics in Data Collection and Privacy

- Web systems entail the collection of vast amounts of data
  - Targeted advertisements
  - Improved search results
  - Improved services
- What is right and wrong in data collection?

- **Bottom Line:** Society increasingly trusts *you* (CS and DS majors) with tons of data.

Step 0: What is *Your* data?

- If we're going to be ethical with data, we'll ask the owner to use it, right?
  - How do we know when data is *yours*?

- If Bob writes a biography about Alice, then Bob owns the copyright.
  If Alice doesn't like it, that's too bad.
  Libel? Only if it's inaccurate

Limits on Ownership (photos)

- If Alice photographs Bob, Alice owns the photo.

- Limits on Alice can be:
  - On taking the photo in certain private areas
    - In Alice's home
    - In a bathroom in Alice's workplace
  - On using the photo in certain ways
    - As implied endorsement
    - As implied libel

Data Ownership (in general)

- Similar limits exist in Alice’s collection of data about Bob
- Free to record and free to use otherwise.
- And we have done this forever …

Limits on recording

- Recording is “wrong” when there is reasonable expectation of privacy
  - E.g., no cameras in clothing store fitting room

Walmart Fitting Room
Random Retail
cc-by-2.0

Limits on data

- Similar limits on data
- Phone company must not record phone call content
- Email provider must not read emails
  - Unless clearly agreed to.
- Mobile apps must not record location except where they provide location-based service.

Informed consent

- You could pose for a photograph under an agreement that I will give you ownership of the photo.
- You could agree to participate in a research experiment under an agreement that has *informed consent*.

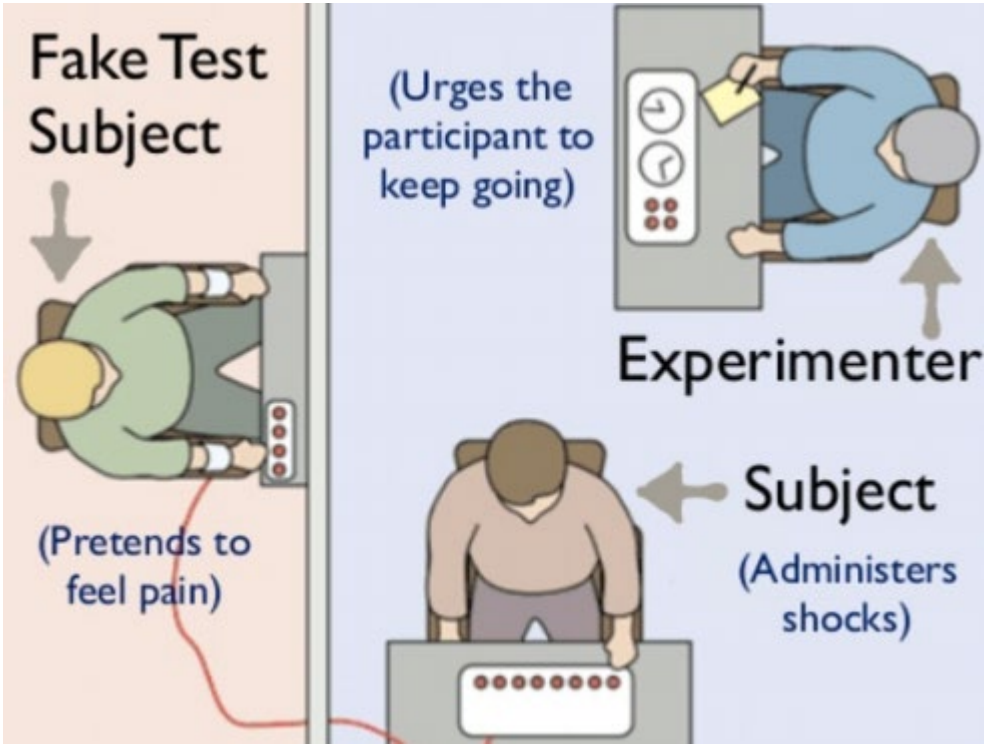
Informed consent

- Research with human subjects
  - Common in medicine, sociology, education, etc.
  - UM (and other organization) have an **institutional review board** to ensure research complies with ethical guidelines

- Human subject must be
  - *Informed* about the experiment
  - Must consent to the experiment
    - Voluntarily, without **coercion**
  - Must have the right to *withdraw* consent at any time

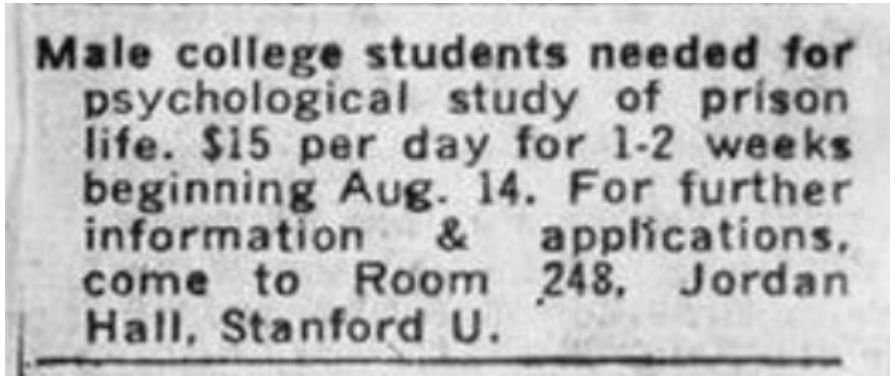
What Goes Wrong Without Informed Consent?

- Milgram
  - Experiment conducted after WWII investigating influence of authority over people's willingness to comply with unpleasant directions
  - **Subject** told to *ask questions* of a **fake subject**.
    - **Wrong answer =>** subject presses a button to shock the fake subject
    - **TL;DR:** subject thinks the fake subject is *really feeling shocks and pain*
- Problem: subject was not informed of the purpose of the experiment



What Goes Wrong Without Informed Consenst?

- Stanford Prison Experiment
  - Two-week simulation of jail.
  - Subjects consist of *prisoners* and *guards*
  - How do people interact in these roles?



  - **TL;DR:** it was a disaster. Psychological torture, participants fully embraced roles

- Problem: No informed consent. Participants didn't know what they were in for.

What Goes Wrong Without Informed Consent?

- Tuskegee Experiment
  - Study 600 African Americans with syphilis
  - Tell them they'll get free medical treatment

  - TL;DR: Experimenters lied. Used placebos, only said "you have bad blood"

- Problem: Participants not informed of their condition, of the risks, and were misled about the purpose of the experiment
  - Subjects were coerced as well – all were poor

Ethical Data Use

- We've established the importance of data **ownership** and **informed consent** for collecting data

- **Using** the data is another story.
  - Even if we own the data or have informed consent, ethical guidelines exist concerning the way such data is used

Data Use: Video Cameras in Stores

- Purpose:
  - **Security:** Prevent theft, robberies, etc.
  - **Efficiency:** study customers to figure out better product placements

- Unethical to publish

Cell phone location tracking

- Necessary to provide services like navigation
- Required for many valuable applications.
- But can result in a huge loss of privacy.

Tracking Phones, Google Is a Dragnet for the Police

The tech giant records people's locations worldwide. Now, investigators are using it to find suspects and witnesses near crimes, running the risk of snaring the innocent.

By JENNIFER VALENTINO-DeVRIES APRIL 13, 2019

- Murder case, trail goes cold
- Google provides information on all devices it recorded near the killing
- Wrong person arrested

https://www.nytimes.com/interactive/2019/04/13/us/google-location-tracking-police.html

Limits on Use of Data

- Web-based services may have a compelling reason to **collect data**
  - They may have implied consent, or may not need it
- **Instead**, we are often concerned with the *use* of data rather than its collection
  - Consider: collecting the data may not require *interacting with or affecting* the user
  - We'll **limit our use** of the data rather than collecting it
- Consequence: guarding data is particularly important
  - Ethically bound to protect the data
  - Attitude: "Let's work as though we never collected this data in the first place."

Voluntary limits on use

- Police reassure citizens that bodycam video will not be posted on the web.

- Businesses can reassure customers that data collected for one purpose will not be used for another.

- These assurances can
  - Have legal force
  - Remove barriers to many transactions.

- "We decline to grant the state unrestricted access to a wireless carrier's database of physical location information."

The New York Times

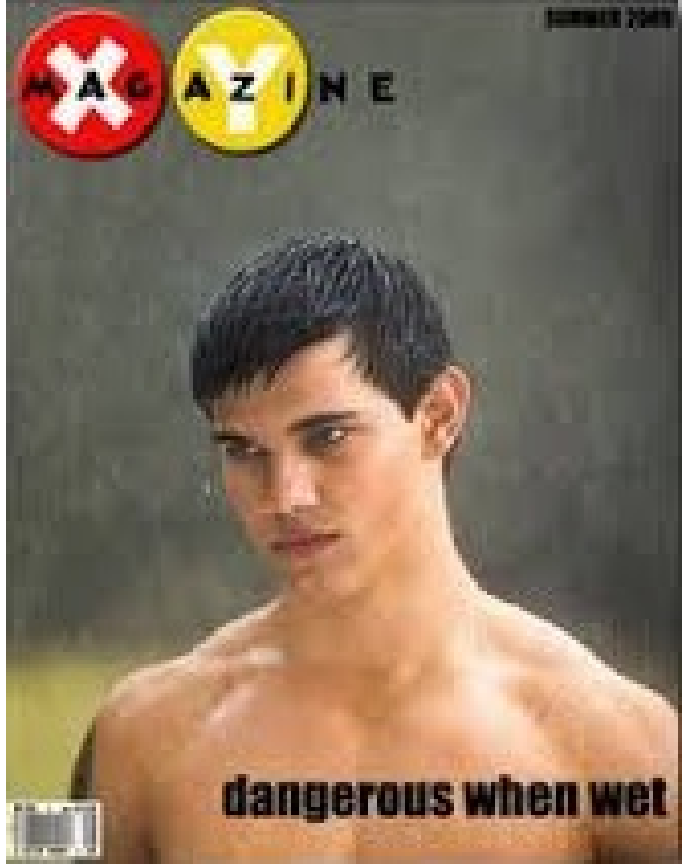
In Ruling on Cellphone Location Data, Supreme Court Makes Statement on Digital Privacy

Data destruction

- Companies legitimately collect data as part of doing business.
  - Ethically compliant and so forth
- Companies need to retain **goodwill** of customers
  - Don’t betray their trust
- Organizations frequently **exit the market** (e.g., bankruptcy)
  - Data *could* be sold to others… valuable asset

- Ethical consideration:
  - Collected data must be **destroyed**, not sold

Example: XY magazine



- 100K print subscribers, 1M online
- Company goes bankrupt, investors buy it and want **user data**
- **Privacy policy** originally stated "never sell its list to anybody"
- Court sides with **privacy policy** over investors
- Many of those customers would still be underage and would not be out to their families yet, thus making their privacy of particular concern.

Data ~~Destruction~~ Reuse?

- My medical data is collected by a hospital and doctors to improve care
  - I'm also okay if that data is **repurposed** for medical research
- My bank collects data about credit card purchases
  - They'll repurpose it for sharing with credit reporting agencies (lol Equifax)
- Considerations: have provisions for data **reuse** in policies
  - Example: as a researcher, we're often limited by companies' willingness to share customer data *because they have no such provisions*

Anonymity

NETFLIX PRIZE

Closeted Lesbian Sues Netflix For Potential Outing

By Laura Northrup on December 19, 2009 3:00 PM

Here's the problem with anonymized data: if it were truly anonymized, it wouldn't be useful to anyone for anything. With enough data about a person—say, their age, gender, and zip code—it's not hard to narrow down who someone is. That's the idea behind a class-action lawsuit against Netflix regarding the customer data they released to the public as part of the Netflix Prize project, a contest to help create better movie recommendations. A closeted lesbian alleges that the data available about her could reveal her identity.

- Settled for $9 million after >2 years of litigation

consumerist.com

A Face Is Exposed for AOL Searcher No. 4417749

By MICHAEL BARBARO and TOM ZELLER Jr. AUG. 9, 2006

Buried in a list of 20 million Web search queries collected by AOL and recently released on the Internet is user No. 4417749. The number was assigned by the company to protect the searcher's anonymity, but it was not much of a shield.

No. 4417749 conducted hundreds of searches over a three-month period on topics ranging from "numb fingers" to "60 single men" to "dog that urinates on everything."

And search by search, click by click, the identity of AOL user No. 4417749 became easier to discern. There are queries for "landscapers in Lilburn, Ga," several people with the last name Arnold and "homes sold in shadow lake subdivision gwinnett county georgia."

It did not take much investigating to follow that data trail to Thelma Arnold, a 62-year-old widow who lives in Lilburn, Ga., frequently researches her friends' medical ailments and loves her three dogs. "Those are my searches," she said, after a reporter read part of the list to her.

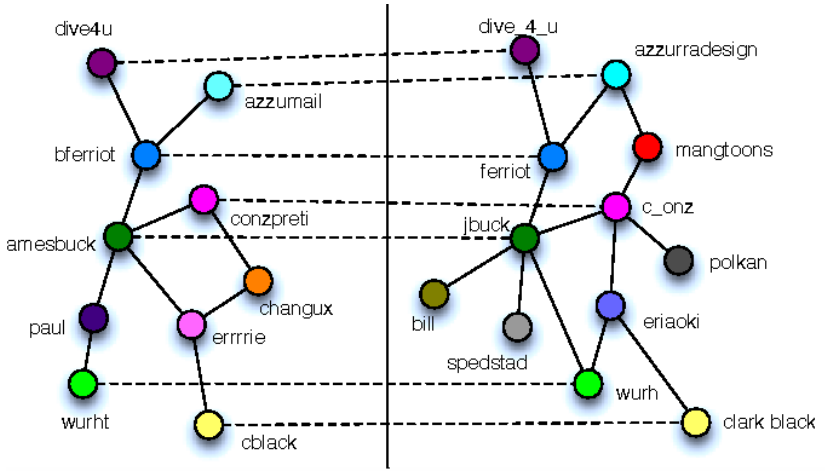
https://www.nytimes.com/2006/08/09/technology/09aol.html

Anonymity is impossible

- Anonymity is virtually *impossible*, with enough other data
  - Diversity of entity sets can be eliminated through joining external data
  - Aggregation works only if there is no known structure among entities aggregated

- **Faces can be recognized** in image data.
  - Progressively, even under challenging conditions, such as partial occlusion

- **Deanonymization** is possible
  - Use information about one network to learn identities in another
  - **Consider**: careful guarding of collected data



Limit publication of datasets

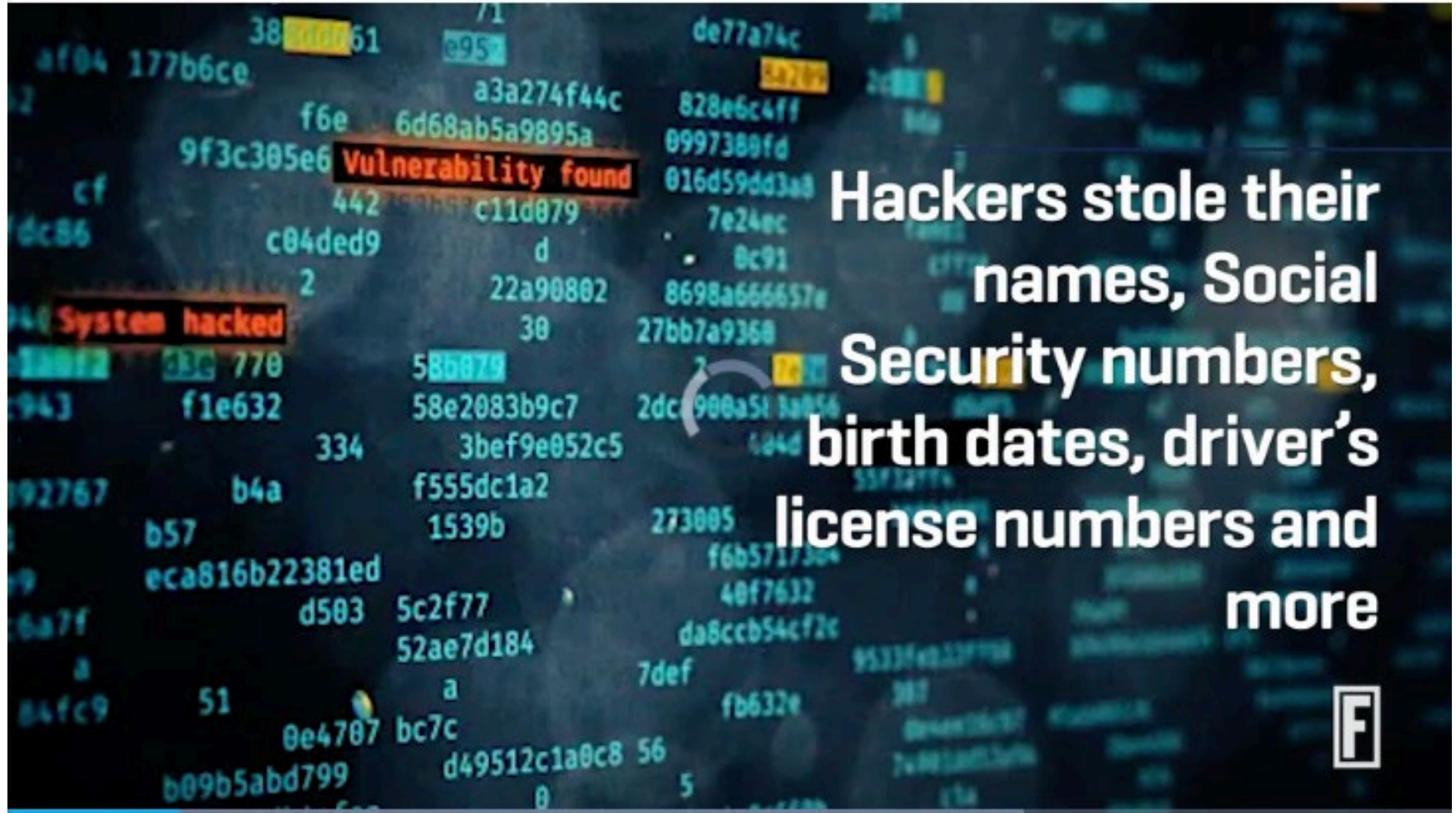
- If anonymity is not possible, the simplest way to prevent misuse is **not to publish** the dataset.
  - e.g. government agencies must not make public potentially sensitive data

- Yet access to data is crucial for many desirable purposes, including:
  - Medical research
  - Public watchdogs

License data to trusted parties

- Need simple licensing regime for access to potentially sensitive data, including de-identified data.
- Enforce through contracts in the business world.
- Enforce through professional standards in the research world.
  - Investigator promises not to re-identify
  - Else, loses reputation and future access
  - Similar to double-blind review
- Sometimes, things still go wrong

Equifax Data Breach, One Year Later: Obvious Errors and No Real Changes, New Report Says



Bias

- We already saw **Google Bombing** to influence search results
- More generally, **big data** has led to an explosion of **deep learning**
- **Biases present in data may influence machine learning models**
  - Systemic biases that influence outcomes of predictions for a variety of purposes
  - Affects performance and use of web systems

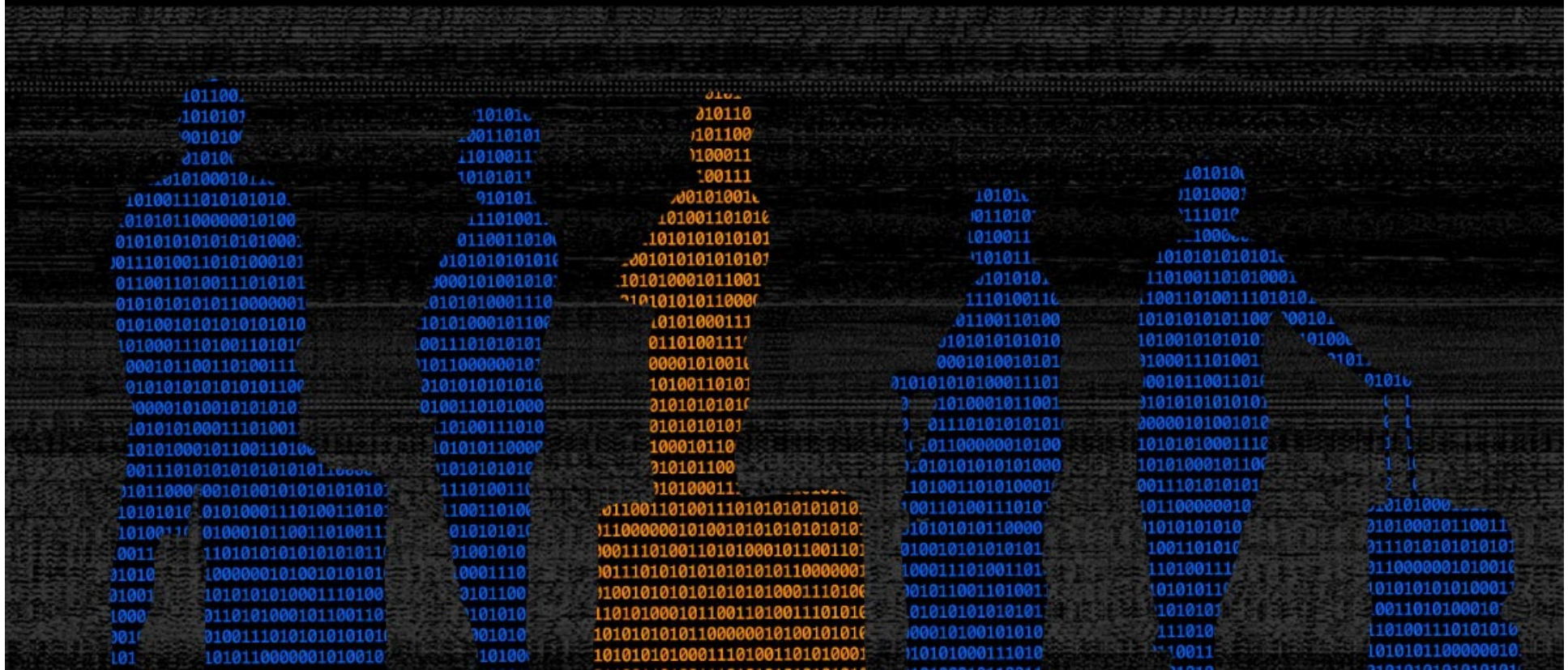
The Intercept_

HOMELAND SECURITY WILL LET COMPUTERS PREDICT WHO MIGHT BE A TERRORIST ON YOUR PLANE — JUST DON'T ASK HOW IT WORKS

Sam Biddle

December 3 2018, 1:47 p.m.

39



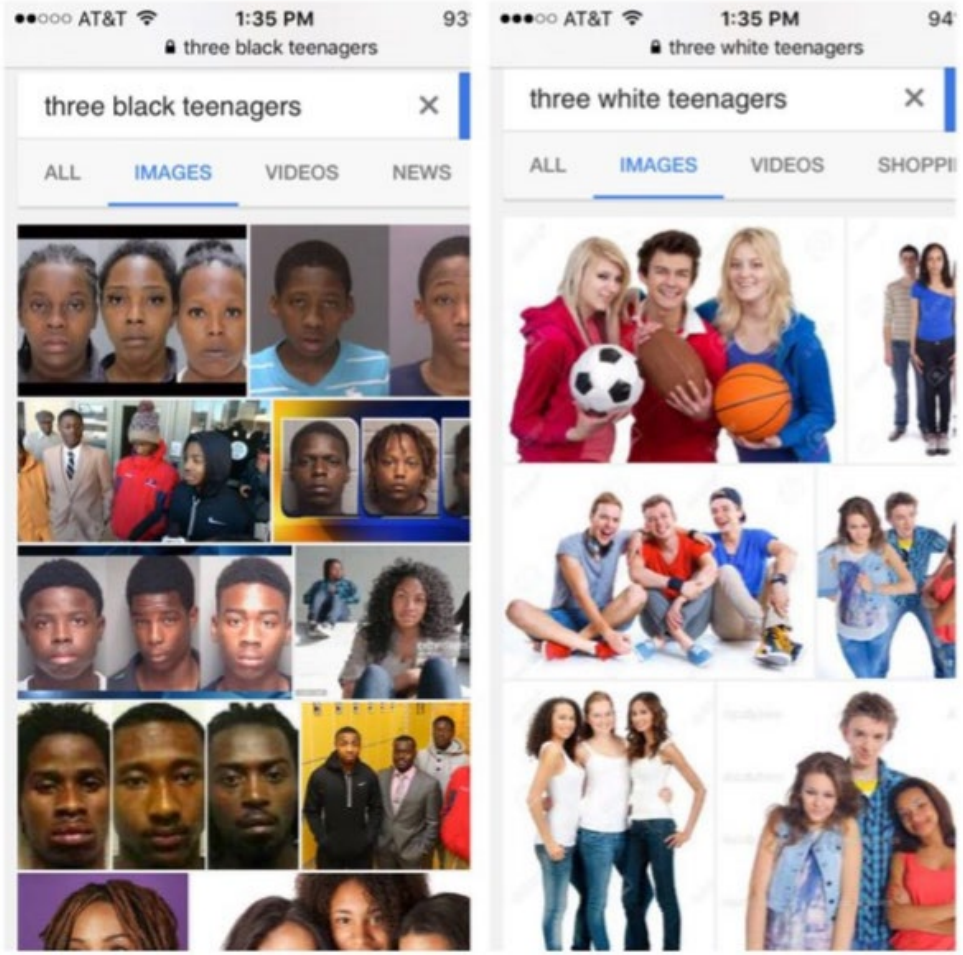
LAPD to scrap some crime data programs after criticism

By MARK PUENTE
APR 05, 2019 | 6:00 PM

https://www.latimes.com/local/lanow/la-me-lapd-predictive-policing-big-data-20190405-story.html

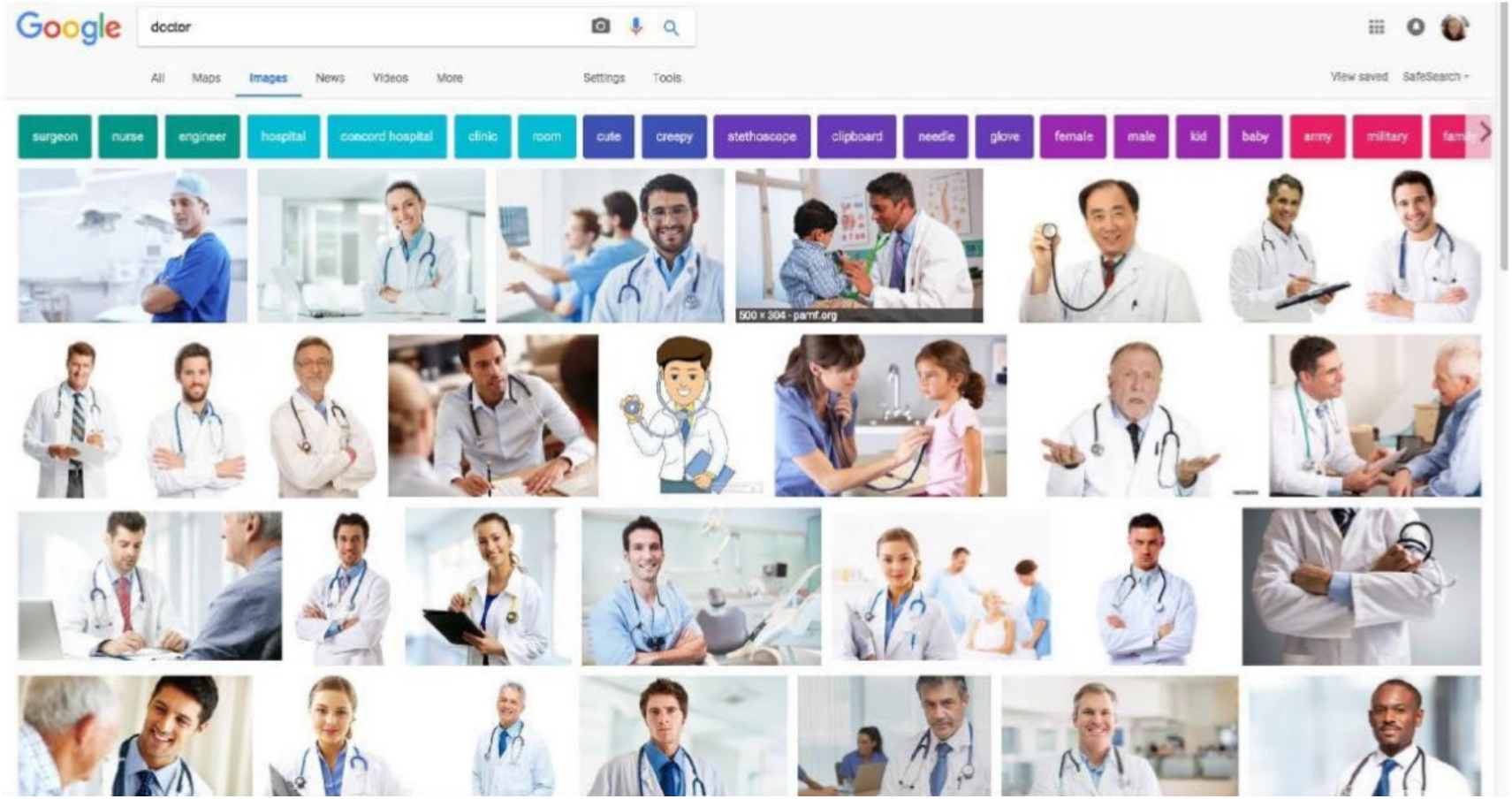
Impact of Social Stereotypes on Data

- 2016 Google queries: racial stereotypes



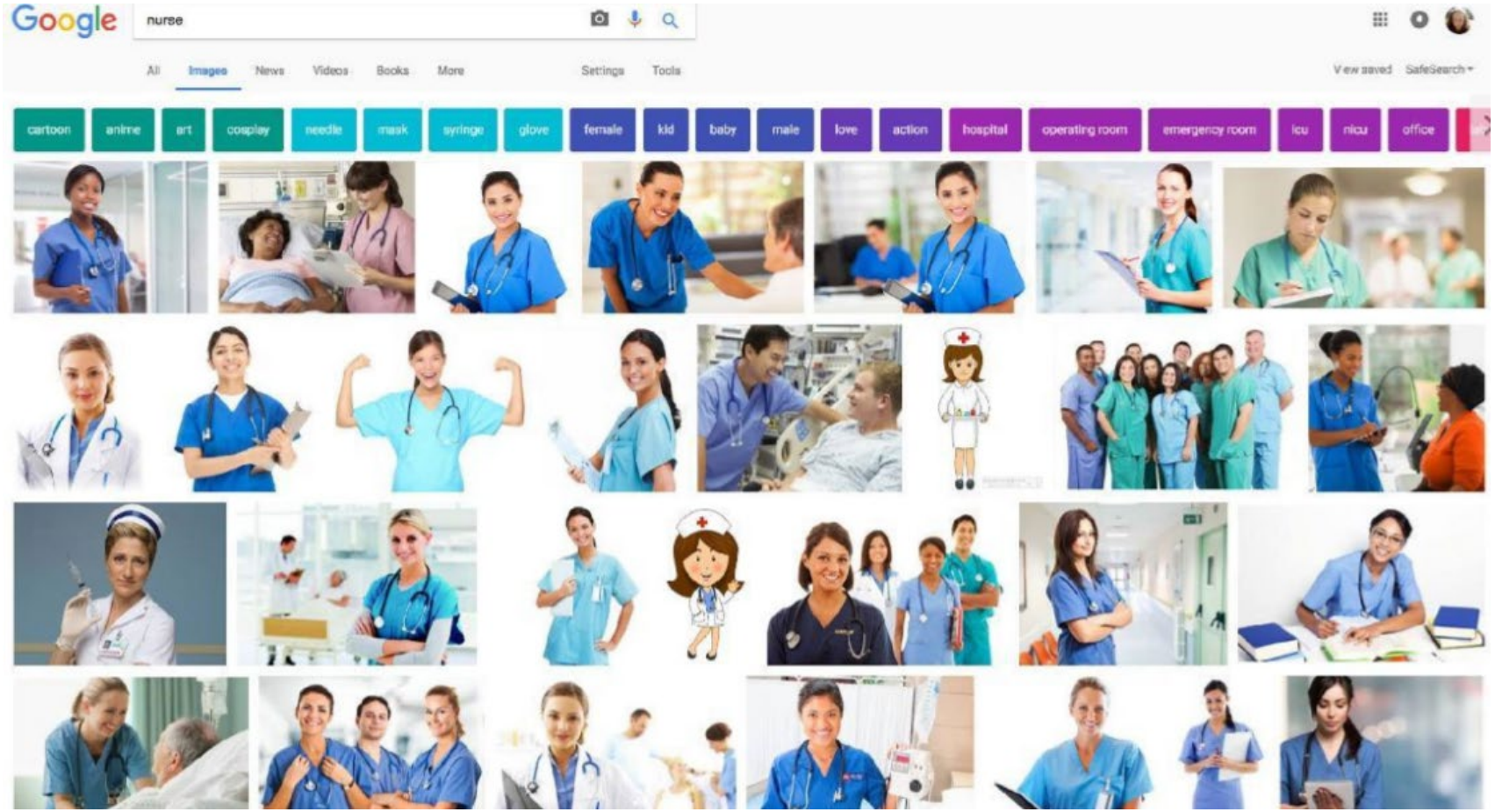
Impact of Social Stereotypes on Data

- Google query for “doctor”: race/gender/age stereotypes



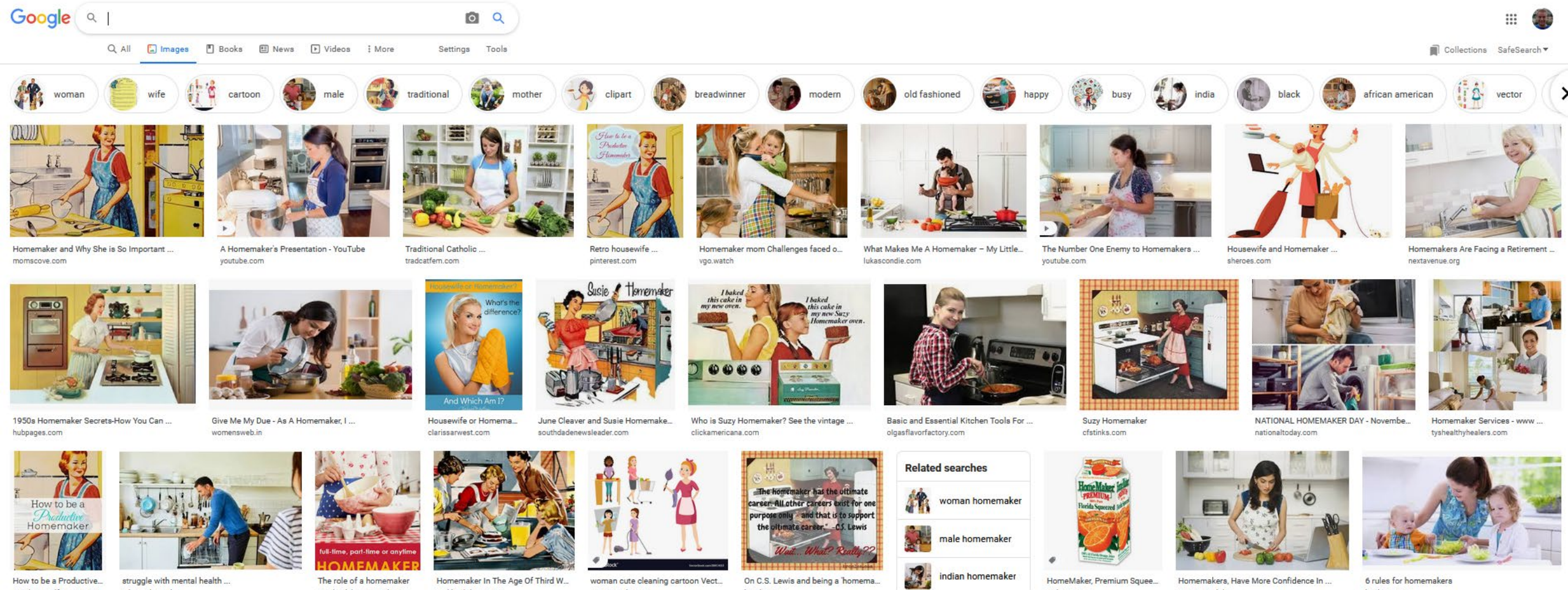
Impact of Social Stereotypes on Data

- Google query for “nurse”



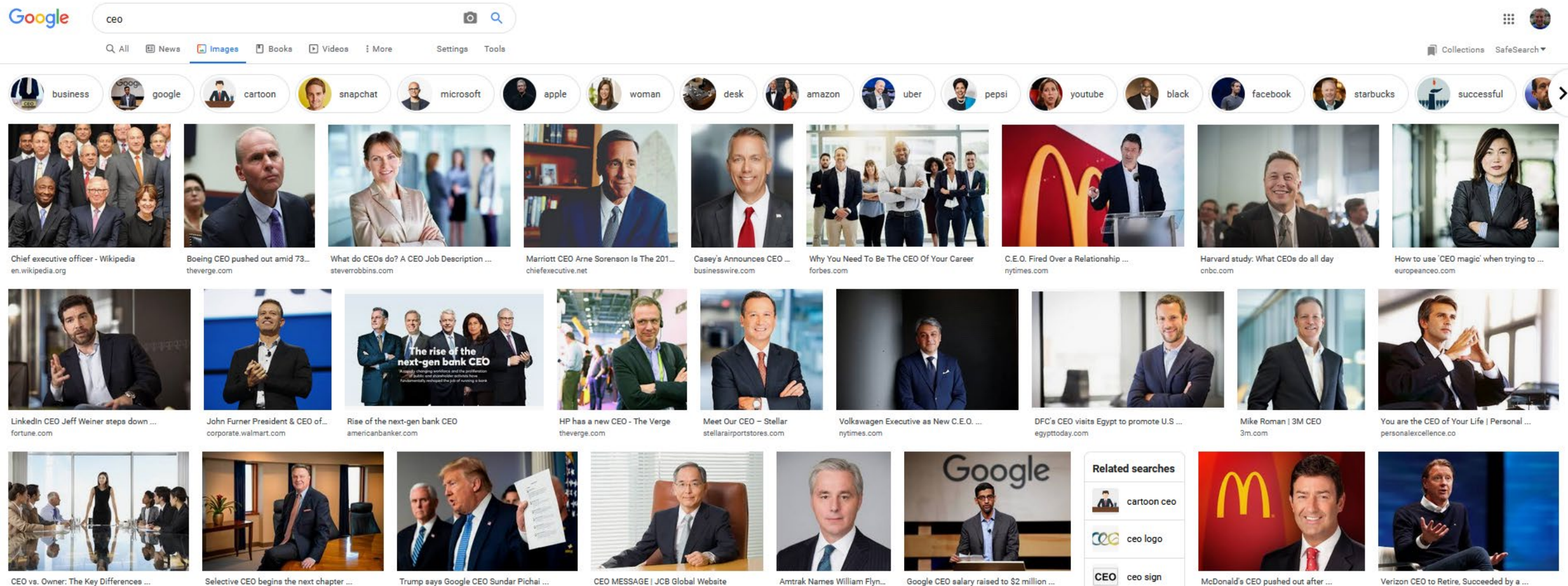
Impact of Social Stereotypes on Data

- Google query for “homemaker”



Impact of Social Stereotypes on Data

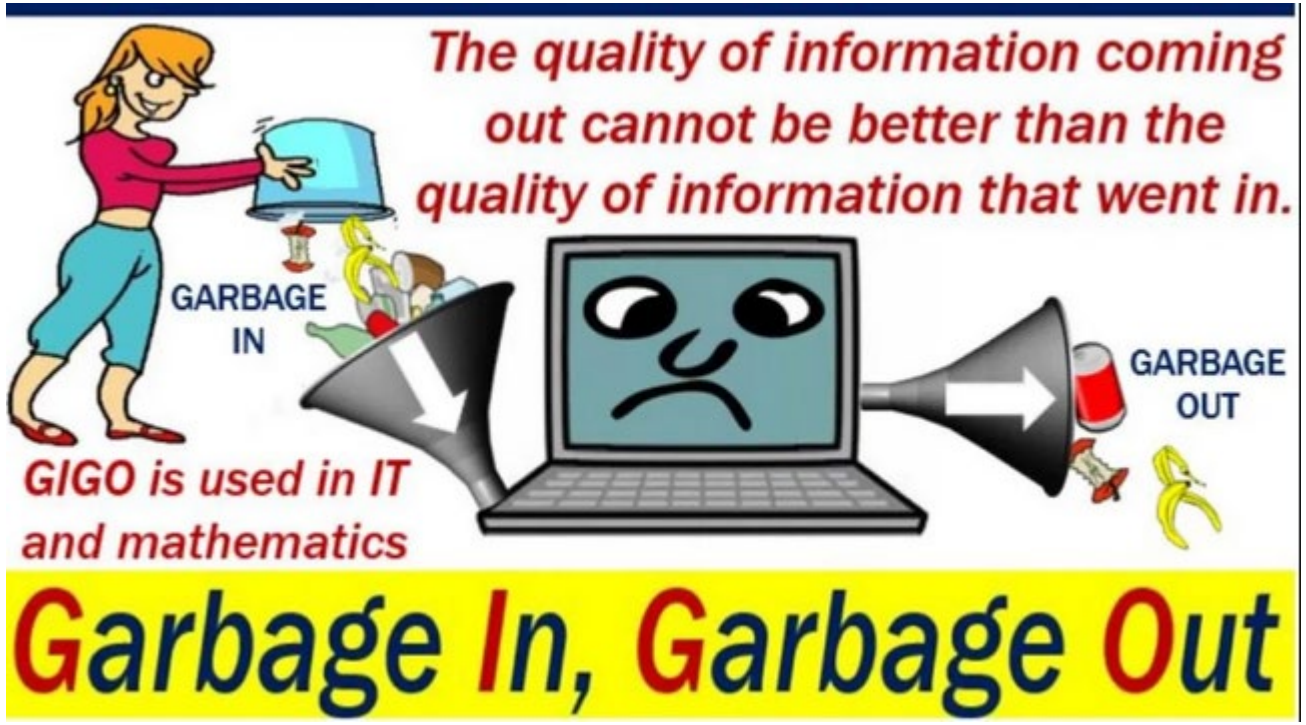
- Google query for “CEO”



Societal Stereotypes in Data

- Biased data produces biased models
  - Thus, predictions are biased as well

- Alternative thought question:
  - What is a chair?



Research on Bias in ML

- Machines learn trustworthiness and likeability traits from faces
  (Steed and Caliskan 2020)
- Self-driving cars biased against genders and races
  (Wilson, Hoffman, and Morgenstern 2019)
- Males are over-represented in the reporting of web-based news articles
  (Jia, Lansdall-Welfare, and Cristianini 2015)
- Males are over-represented in twitter conversations
  (Garcia, Weber, and Garimella 2014)
- Biographical articles about women on Wikipedia disproportionately discuss romantic relationships or family-related issues
  (Wagner et al. 2015)
- IMDB reviews written by women are perceived as less useful
  (Otterbacher 2013)

Sources of Bias

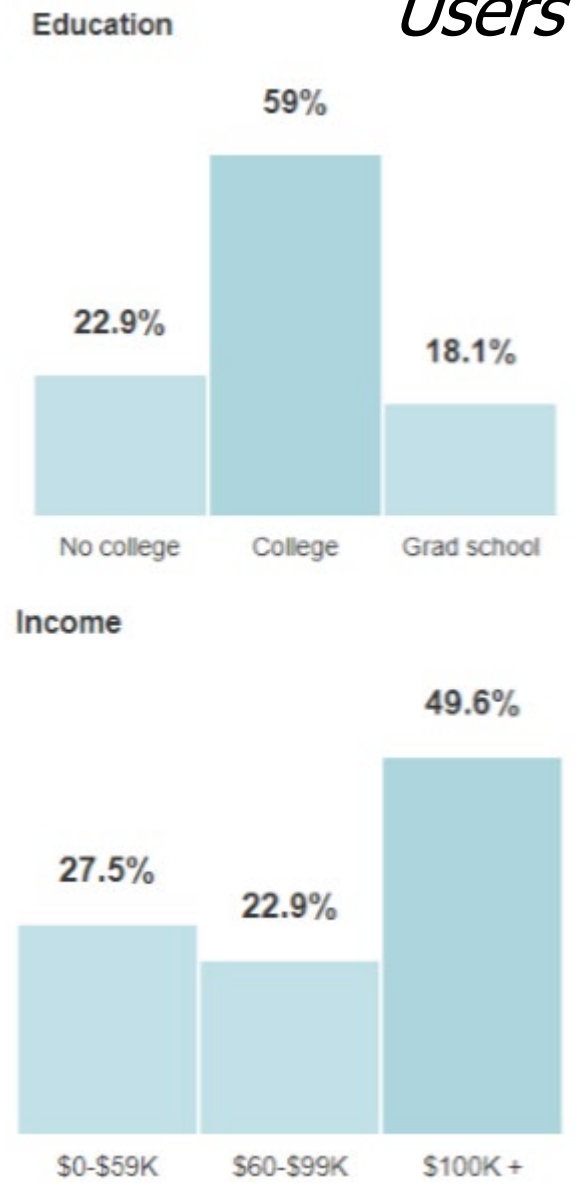
- Bias in data and sampling
  - (social biases, unrepresentative user base)
- Optimizing for a biased objective
  - (bad training)
- Inductive bias
  - (implicit assumptions made by the model itself)
- Bias amplification
  - (the model learns the “wrong” features)

Bias in Data and Sampling

- **Self-selection bias** is a statistical effect in which a group will select themselves, biasing a sample

- **Concretely:** who writes Yelp reviews? Who reads them?
  - People may not talk about things consistent with empirical measurement
  - Communities of language speakers lead to differing model performance
- What about system bias?
  - Can we tell if Yelp is biasing reviews?
    - "it would be a shame if you didn't pay us and you got a few 1-star reviews…"

Distribution of Yelp Users



Bias in Language Identification

- NLP application: Identifying a language give a string written in it

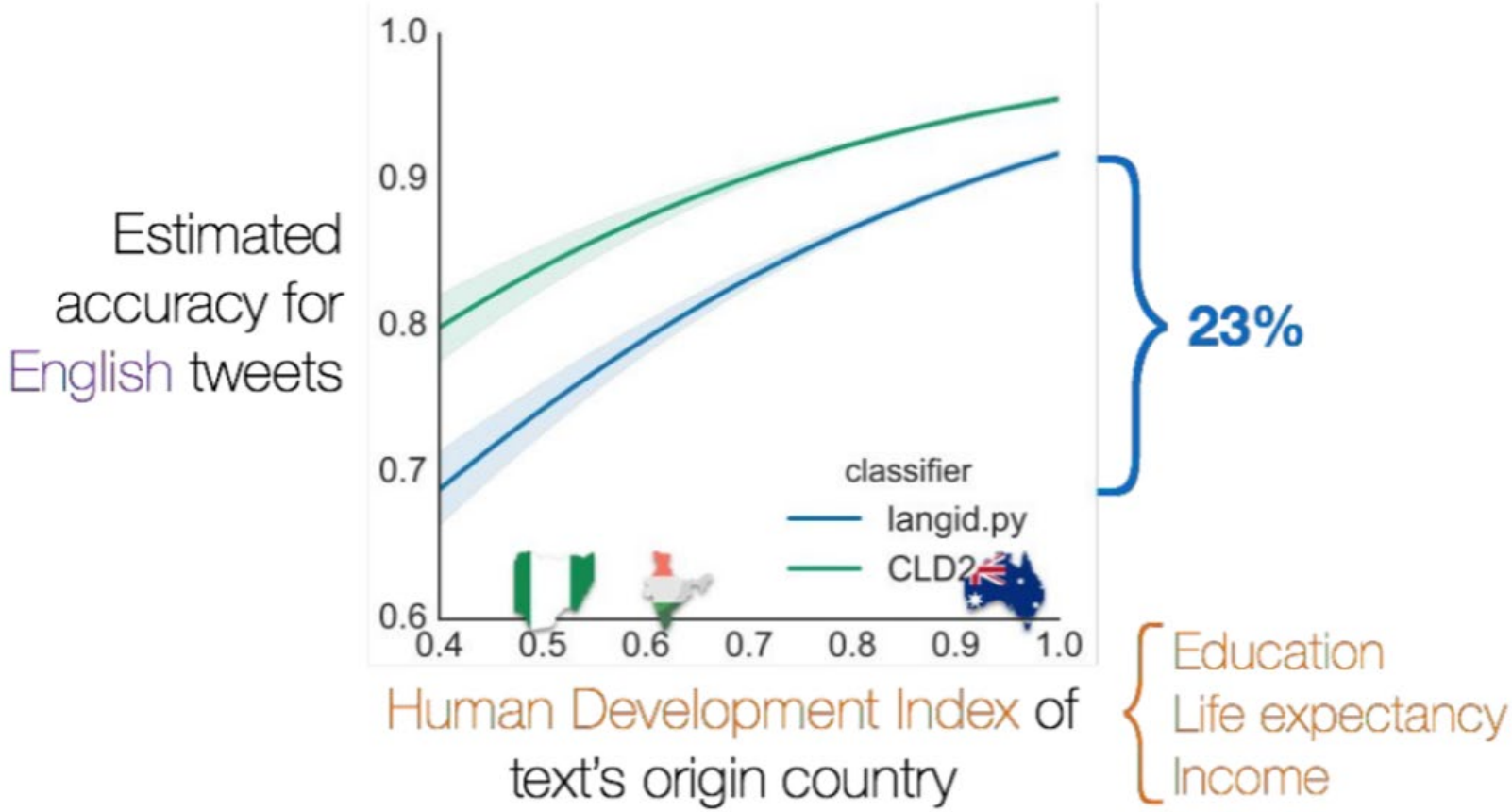
After language identification, we can look for keywords like flu/sick, then followup with a conclusive explanation

(maybe they're hungover)

If we can't identify the language to begin with, there's no way to extract followup semantics (i.e., we can't find keywords like flu/sick without knowing it's an English Tweet)

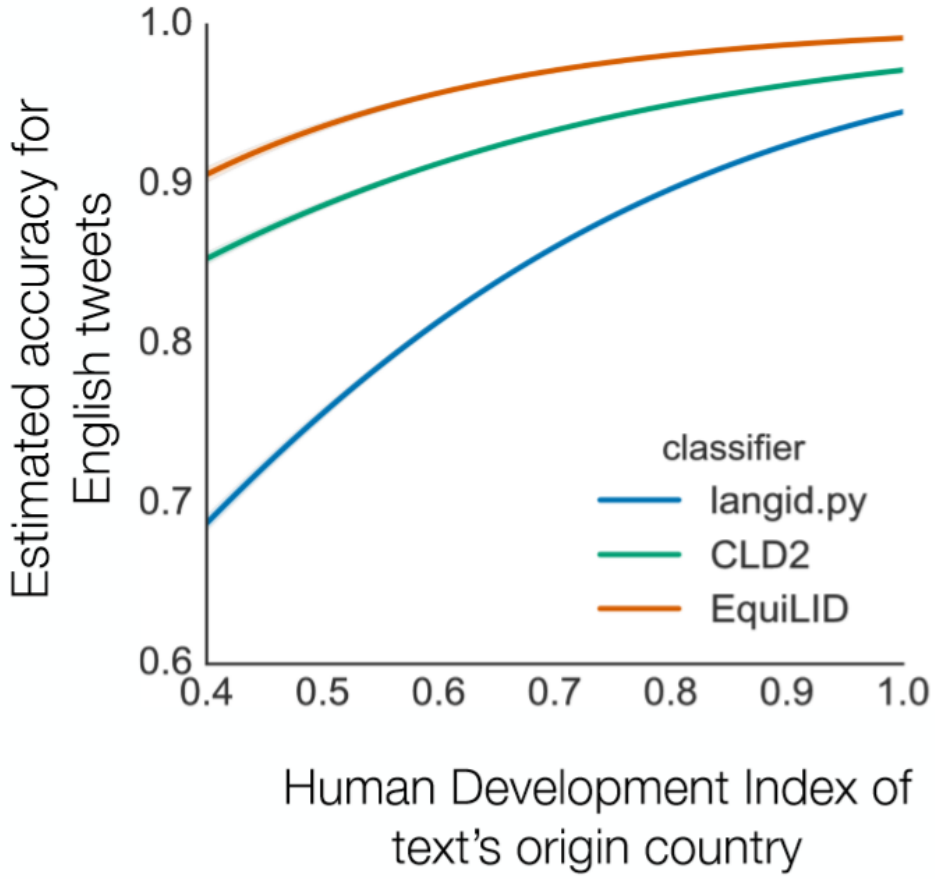
Bias in Language Identification

- Language Identification systems under-represent populations in underdeveloped countries



Bias in Language Identification

- By retraining on more representative corpora:



Objective Bias

- **Objective bias** occurs when models are asked to make predictions that actually answer a different question

- **Concretely:** "What is the **probability** that a given **person** will commit a serious **crime** in the **future** based on the **sentence given now**?"

- Example: COMPAS
  - Balanced data from people of all races (and race was not a feature)
  - **Problem:** "who will commit a crime" is not obtainable (we can't know it ahead of time)
    - **Instead**: model was learning "who is more likely to be convicted" (notice the difference!)

Inductive Bias

- An **Inductive bias** is the result of an implicit assumption made in the construction of a given model

- **Concretely:** Embeddings may represent biases
- In word2vec (gross oversimplification: fancy tf-idf scores):
  - $\overrightarrow{man} - \overrightarrow{woman} \approx \overrightarrow{computer\ programmer} - \overrightarrow{homemaker}$

Inductive Bias in Embeddings

$$\min cos(he - she,\ x - y)\ s.t.\ ||x - y||_2 < \delta$$

Extreme *she*	**Extreme *he***
1. homemaker	1. maestro
2. nurse	2. skipper
3. receptionist	3. protege
4. librarian	4. philosopher
5. socialite	5. captain
6. hairdresser	6. architect
7. nanny	7. financier
8. bookkeeper	8. warrior
9. stylist	9. broadcaster
10. housekeeper	10. magician

Gender stereotype *she-he* analogies

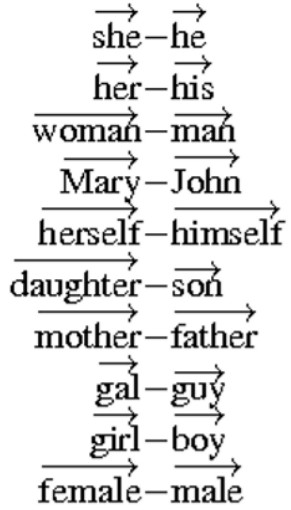
sewing-carpentry	registered nurse-physician	housewife-shopkeeper
nurse-surgeon	interior designer-architect	softball-baseball
blond-burly	feminism-conservatism	cosmetics-pharmaceuticals
giggle-chuckle	vocalist-guitarist	petite-lanky
sassy-snappy	diva-superstar	charming-affable
volleyball-football	cupcakes-pizzas	lovely-brilliant

Gender appropriate *she-he* analogies

queen-king	sister-brother	mother-father
waitress-waiter	ovarian cancer-prostate cancer	convent-monastery

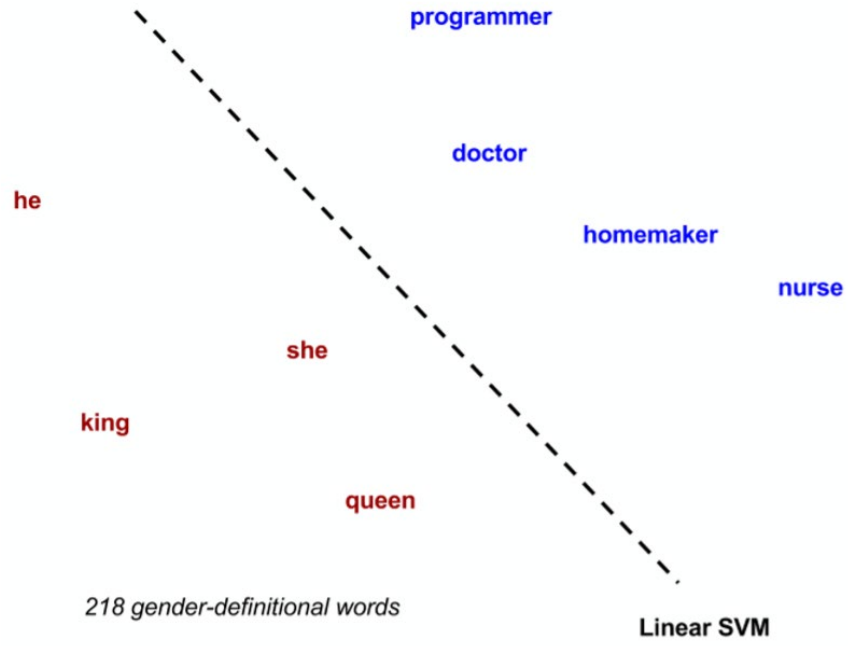
Figure 1: **Left** The most extreme occupations as projected on to the *she−he* gender direction on w2vNEWS. Occupations such as *businesswoman*, where gender is suggested by the orthography, were excluded. **Right** Automatically generated analogies for the pair *she-he* using the procedure described in text. Each automatically generated analogy is evaluated by 10 crowd-workers to whether or not it reflects gender stereotype.

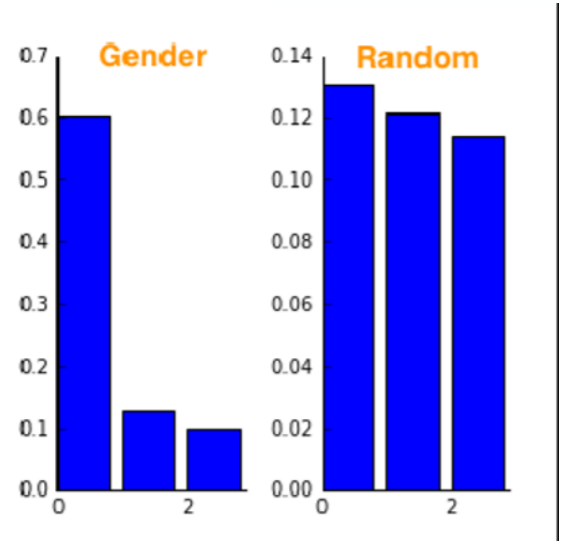
Fixing Inductive Bias: Debiasing



- First: Identify some biased subspace (e.g., using PCA)
- Second: Find subspace-neutral words (e.g., using SVM)
- Third: Transform feature space to minimize the subspace components

- TL;DR: Minimize impact of feature components that leads to bias in subspace-neutral words





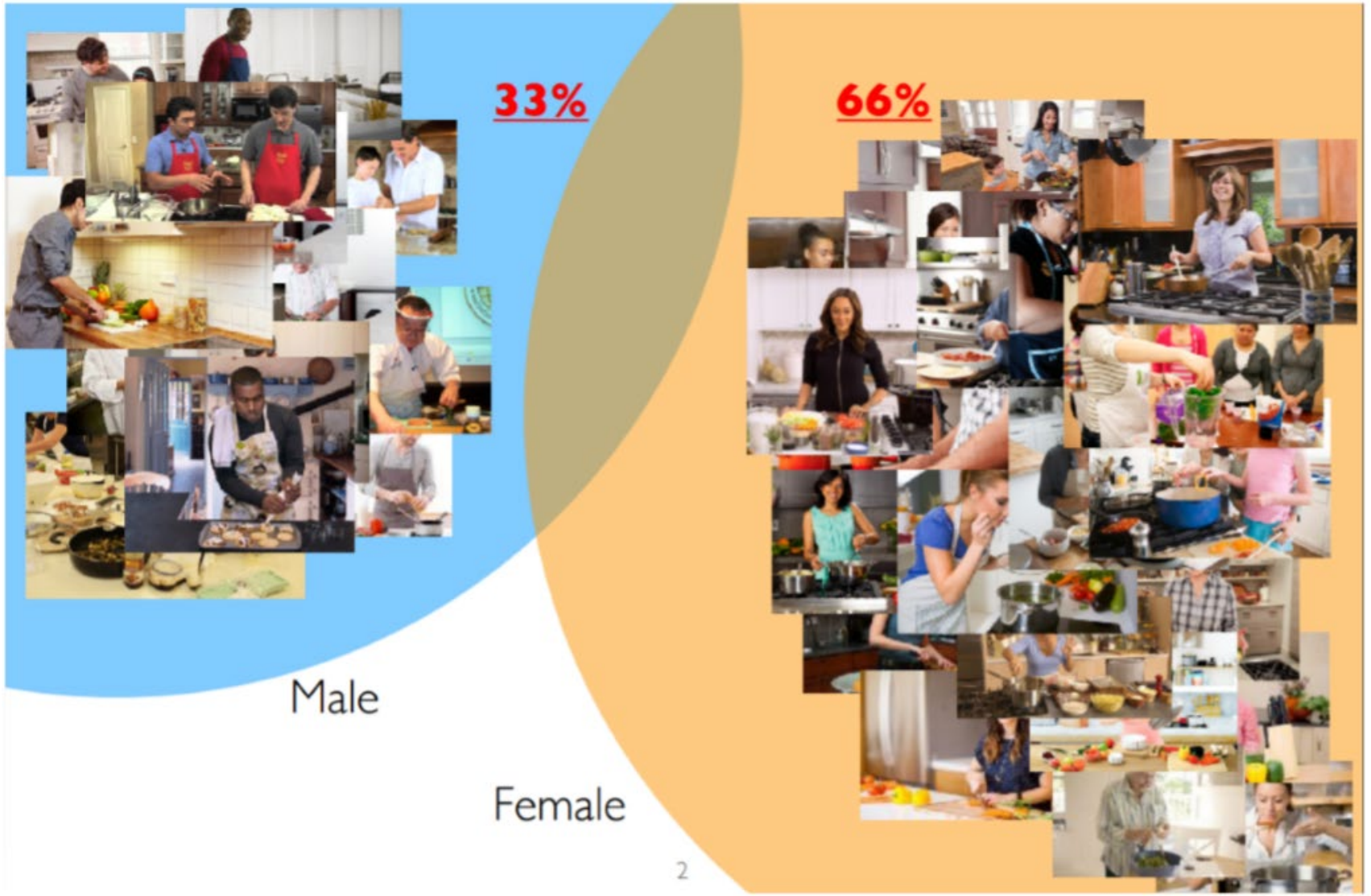
Top principal components identify gender subspace

Bias Amplification

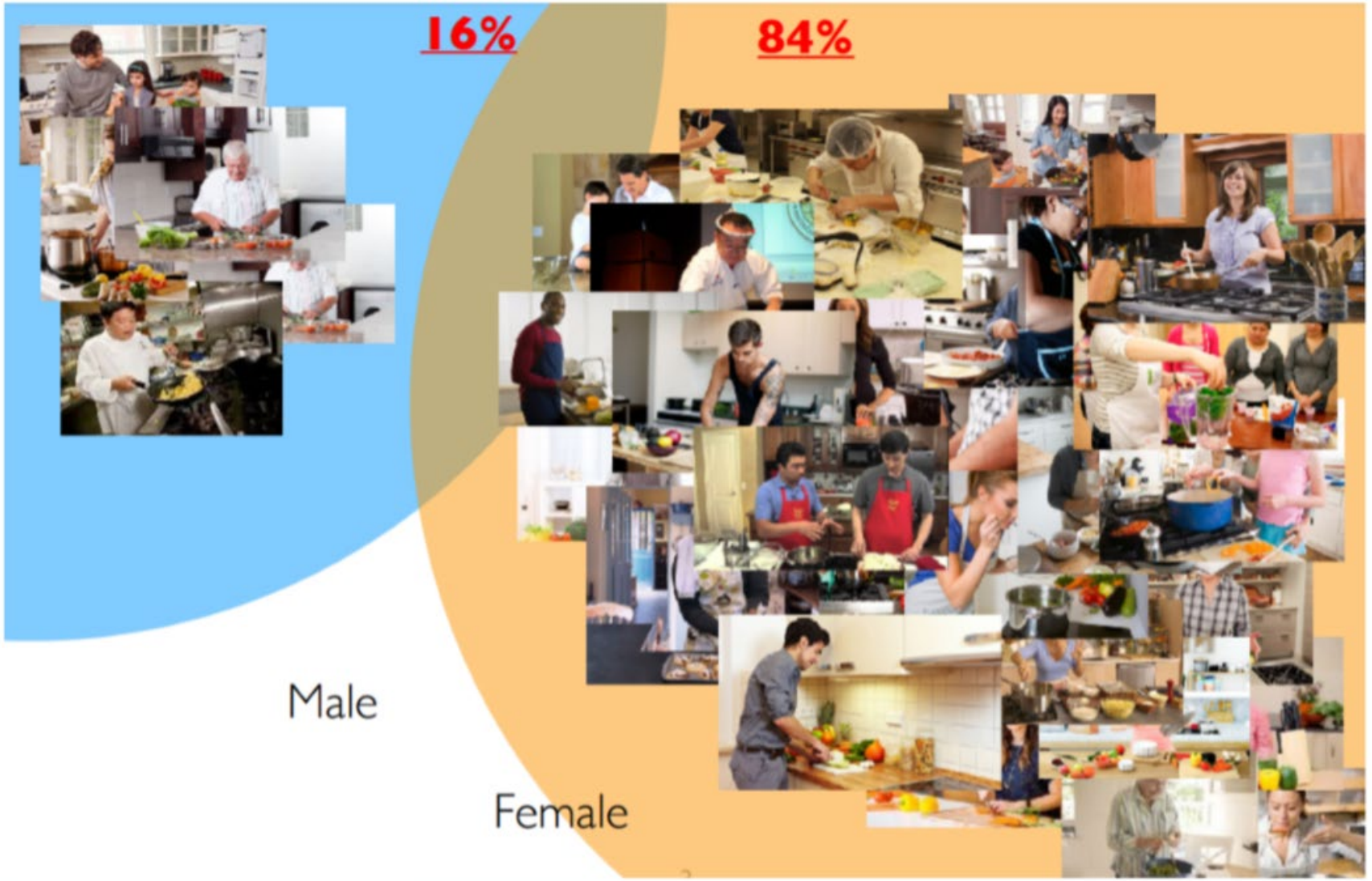
- **Bias Amplification** occurs when unrepresentative data leads a model to learn the wrong features

- **Consider:** What is a chair?
- **Concretely:** if all of your dataset contains barstools as examples of chairs, your model will learn the wrong features
  - e.g., it will only have examples of tall, backless seats near alcohol sources

Bias Amplification: Training

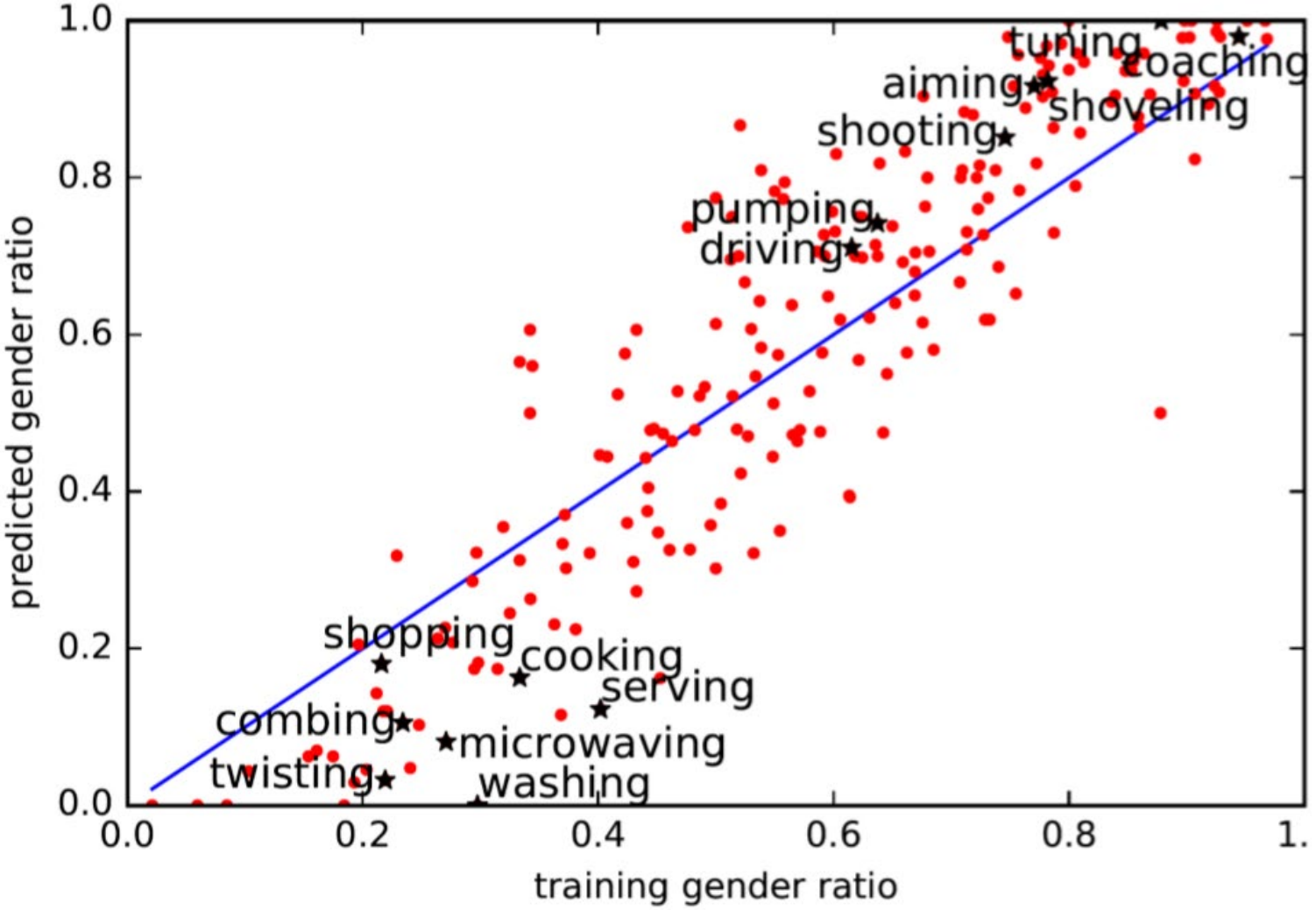


Bias Amplification: Predictions



Reducing Bias Amplification

- Find ratio of predictions made against ground truth labels
- Identify distribution of labels in dataset
- Adjust predicted outputs based on target distribution



Bias in ML and Web Systems

- Bias in data and sampling
  - (social biases, unrepresentative user base)
- Optimizing for a biased objective
  - (bad training)
- Inductive bias
  - (implicit assumptions made by the model itself)
- Bias amplification
  - (the model learns the “wrong” features)

Algorithmic fairness

- Do the data “speak for themselves”?
- Can algorithms be biased?
- Can we make algorithms unbiased?
  - Is training data set representative of the population?
  - Is past population representative of future population?
  - Are observed correlations due to confounding processes?

Algorithmic fairness

- Humans have many biases.
  - No human is perfectly fair, even with the best of intentions.
- Biases in algorithms usually easier to measure, even if outcome is no fairer.
- Mathematical definitions of fairness can be applied, proving fairness, at least within the scope of the assumptions.