



University of Colorado **Boulder**

Department of Computer Science

CSCI 4622: Machine Learning

Chenhao Tan

Lecture 27: Ethics and machine learning

Slides adapted from Chris Ketelsen

# HAPPY



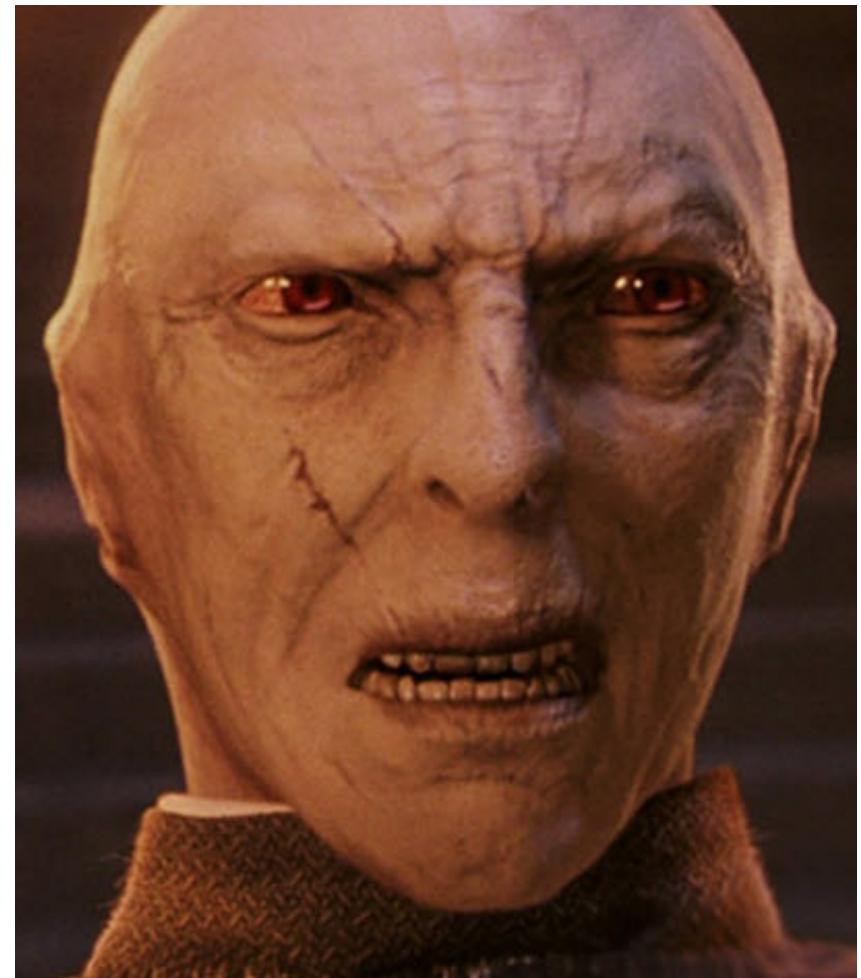
# Remaining lectures

- Prelim 3
- Two hackathon where the TAs and I will be here

# Learning objectives

- Learn about the connection between our society and machine learning
- Make sure that you think about ethics when applying machine learning

Now you understand the magic behind  
machine learning!



Now you understand the magic behind  
machine learning!

Machine learning in real life

Machine learning are commonly used in our society

# Brainstorm

Where did you see machine learning today?

# Machine learning and our society

- Machine learning are increasingly connected with our society
  - Authorizing credit
  - Sentencing guidelines
  - Suggesting medical treatment

And many more!

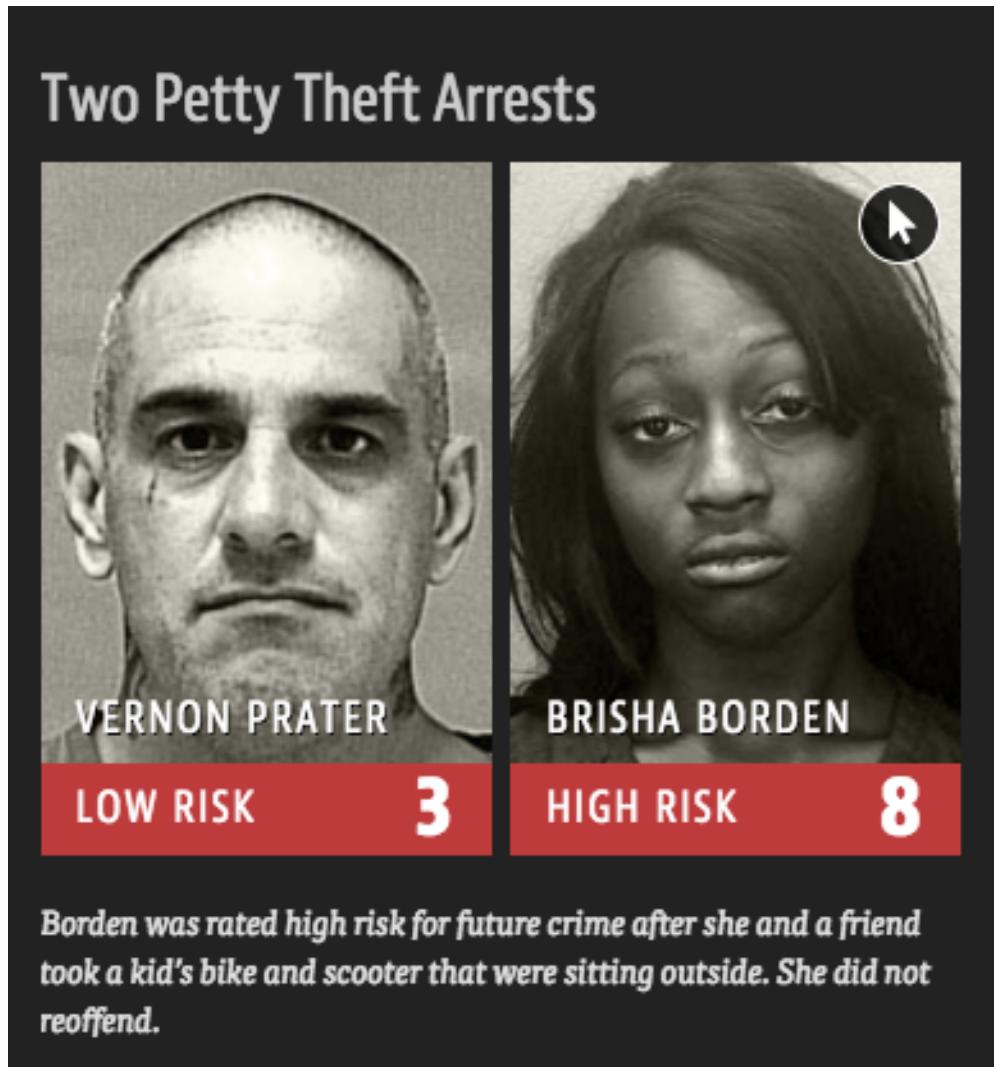
# Ethical issues in Machine Learning





<https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>

# Case study: Decision making



<https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>

# Case study: Decision making

- Crime prediction, similar situations arise when someone is getting a loan
- Harms of allocation

<https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>

# Case study: Decision making

- Crime prediction, similar situations arise when someone is getting a loan
- Harms of allocation
- What are potential reasons in the machine learning pipeline?

<https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>

# Case study: Word embeddings

# Case study: Word embeddings

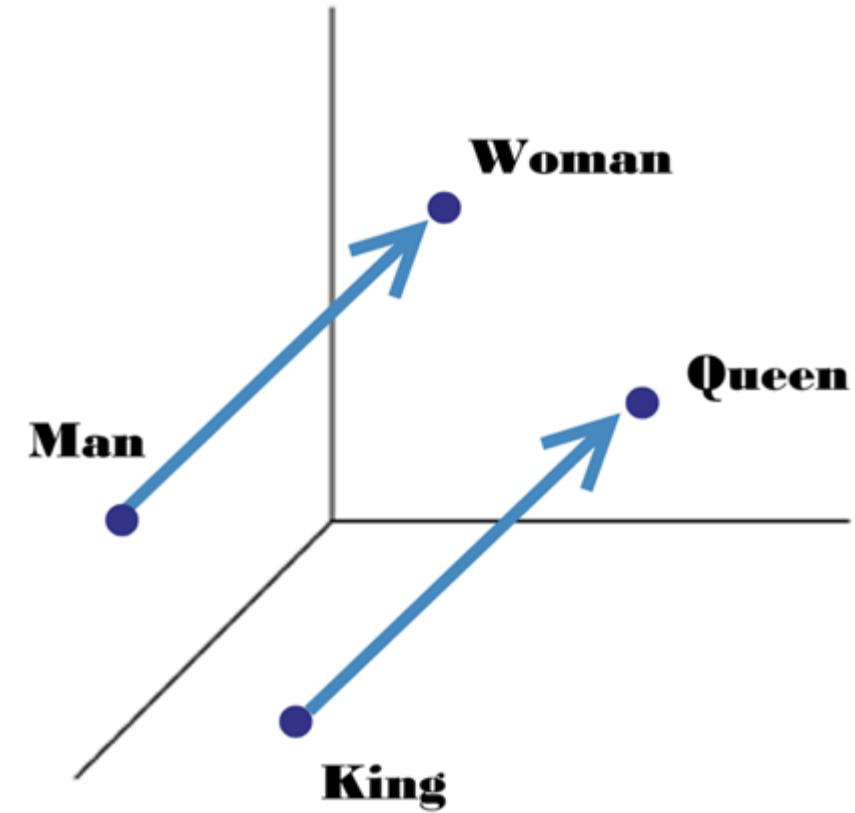


Figure credit: Lior Shkller

# Case study: Word embeddings

## Gender stereotype *she-he* analogies.

sewing-carpentry	register-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	hairdresser-barber

## Gender appropriate *she-he* analogies.

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

Bolukbasi et al. 2016, Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings

# Case study: Word embeddings

SVD to rescue

Bolukbasi et al. 2016, Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings



## Kääntäjä

Poista pikakäännös k

englanti

suomi

ruotsi

Tunnista kieli



suomi

englanti

ruotsi



Käännä

hän on lastenhoitaja  
hän on automekaanikko  
hän on lääkäri  
hän on professori  
hän on lakimies  
hän on voimistelija  
hän on jääkiekkoilija  
hän on taiteilija  
hän on ohjelmoija  
hän on sairaanhoitaja  
hän on opettaja  
hän on parturi  
hän on kampaaja



she is a nanny  
he is a car mechanic  
she is a doctor  
he is a professor  
she is a lawyer  
he is a gymnast  
he is a hockey player  
he is an artist  
he is a programmer  
she is a nurse  
he is a teacher  
she is a barber  
she is a hairdresser



# Case study: Recommender Systems

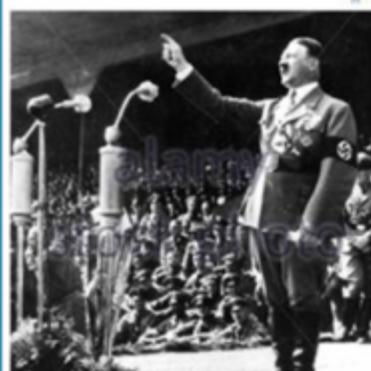
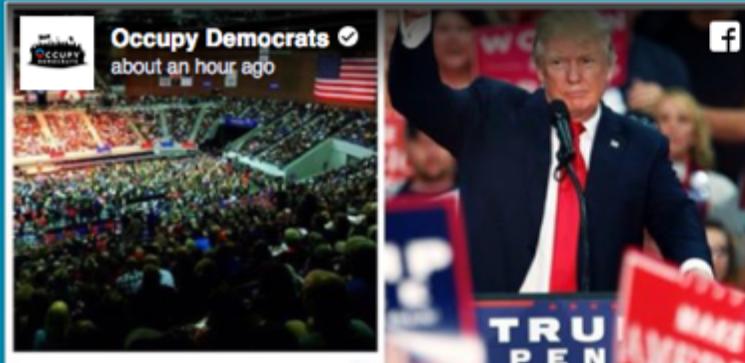
## Filter bubble [Pariser, 2011]

- Main idea: Personalized search determines what information you see and what information you don't see
  - Google, Facebook, Netflix

# Case study: Recommender Systems

- Blue feed vs Red feed

<http://graphics.wsj.com/blue-feed-red-feed/>

**LIBERAL**SHOWING POSTS ABOUT:  
"PRESIDENT TRUMP"**CONSERVATIVE****Occupy Democrats** about an hour ago

"Having obtained my degree in communications I've been fascinated by semiotics (study of signs and symbols) and crowd manipulation. I watched President Trump's rally in Florida to observe his communication style and how he connects to his audience. It brought to mind of the Nuremberg Rally in 1937 where Hitler addressed his nation. I've used parts of the film, Triumph of The Will, that documents this rally to students in classes to illustrate the communication of signs and sy... See More

1.7K 795 2.2K

**THERE'S A WAY TO REMOVE  
TRUMP FROM OFFICE****Allen West** about an hour ago

Gotta admit, this is pretty good...  
Well played, President Trump.

**Watch: Trump Gives a Special 'Present' to Oba...**

And boom goes the dynamite...

ALLENBWEST.AMERICANNEWSHUB.COM



1.8K



78



240

**Gateway Pundit** 2 hours ago

VA has Alec Baldwin hanging instead of President Trump!

**OUTRAGE: Veterans Affairs Offic...**

A photo posted by a Veterans Affairs physician sho...

THEGATEWAYPUNDIT.COM



69



301



430

**Conservative Tribune** 3 hours ago

In 8 years, the best Obama could do was give the whole Middle East over to ISIS.

In just 11 months, Donald Trump utterly crushed them.

Imagine how many lives Obama lost over those years and how many Trump just saved.

**Less Than 1 Year After Trump Too...**

# Case study: Recommender Systems

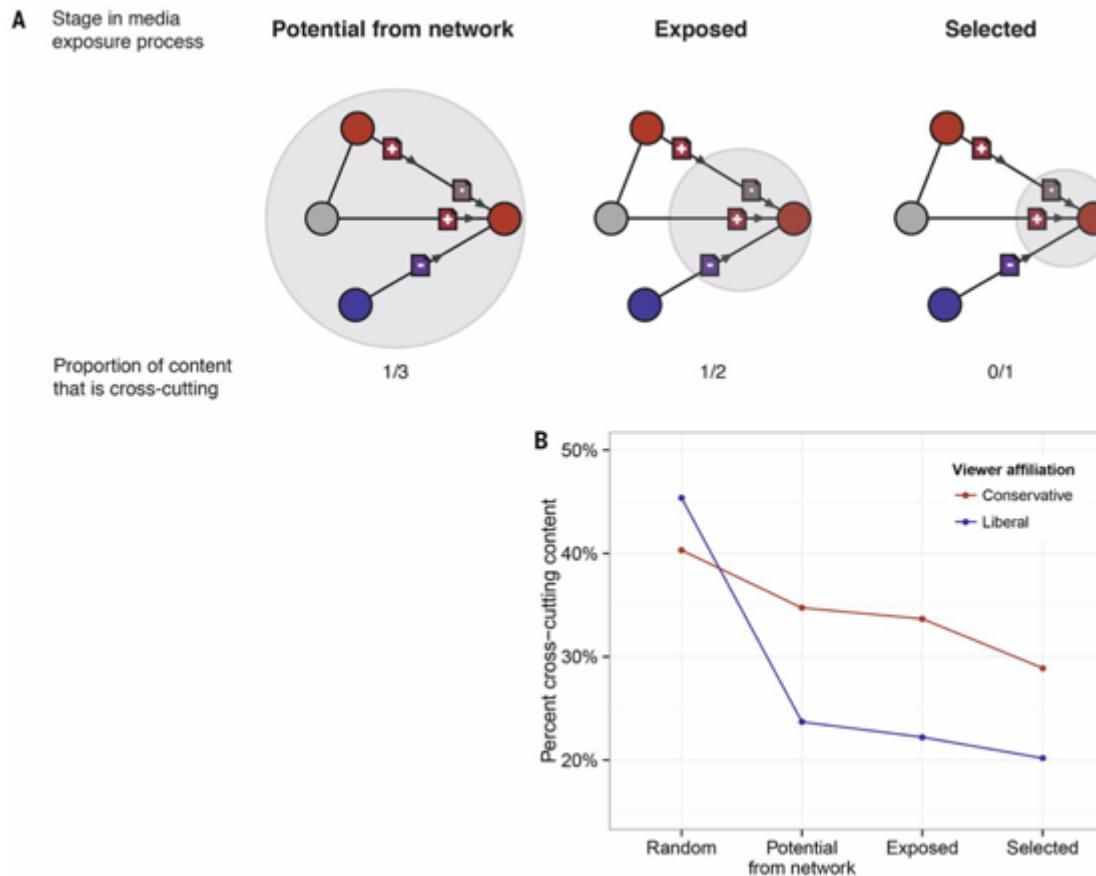
## Filter bubble [Pariser, 2011]

- Main idea: Personalized search determines what information you see and what information you don't see
  - Google, Facebook, Netflix
- Debate: a good thing or a bad thing?

# Case study: Recommender Systems

- Bakshy, Messing, Adamic, “Exposure to ideologically diverse news and opinion on Facebook”

# Case study: Recommender Systems



Bakshy, Messing, Adamic, “Exposure to ideologically diverse news and opinion on Facebook”

# Case study: Recommender Systems

- Latanya Sweeney 2013, “Discrimination in Online Ad Delivery”

Ads related to latanya farrell ⓘ

**Latanya Farrell, Arrested?**

[www.instantcheckmate.com/](http://www.instantcheckmate.com/)

1) Enter Name and State. 2) Access Full Background Checks Instantly.

**Latanya Farrell**

[www.publicrecords.com/](http://www.publicrecords.com/)

Public Records Found For: Latanya Farrell. View Now.

(a)

LATANYA FARRELL  
41 Lexington Rd  
West Hartford, CT 06119  
DOB: Jun 16, 1972 (40 years old)

**Criminal History**

This section contains possible citation, arrest, and criminal records for the subject of this report. While our database does contain hundreds of millions of arrest records, different countries have different rules regarding what information they will and will not release.

We share with you as much information as we possibly can, but a clean slate here should not be interpreted as a guarantee that Latanya Farrell has never been arrested; it simply means that we were not able to locate any matching arrest records in the data that is available to us.

**Possible Matching Arrest Records**

Name	County and State	Offense	View Details
No matching arrest records were found.			

(b)

Ad related to latanya sweeney ⓘ

**Latanya Sweeney Truth**

[www.instantcheckmate.com/](http://www.instantcheckmate.com/)

Looking for Latanya Sweeney? Check Latanya Sweeney's Arrests.

Ads by Google

**Latanya Sweeney, Arrested?**

1) Enter Name and State. 2) Access Full Background Checks Instantly.

[www.instantcheckmate.com/](http://www.instantcheckmate.com/)

**Latanya Sweeney**

Public Records Found For: Latanya Sweeney. View Now.  
[www.publicrecords.com/](http://www.publicrecords.com/)

**La Tanya**

Search for La Tanya Look Up Fast Results now!  
[www.ask.com/La+Tanya](http://www.ask.com/La+Tanya)

(c)

LATANYA SWEENEY  
1420 Charles Ave  
Pittsburgh, PA 15219  
DOB: Oct 27, 1984 (35 years old)

**Criminal History**

This section contains possible citation, arrest, and criminal records for the subject of this report. While our database does contain hundreds of millions of arrest records, different countries have different rules regarding what information they will and will not release.

We share with you as much information as we possibly can, but a clean slate here should not be interpreted as a guarantee that Latanya Sweeney has never been arrested; it simply means that we were not able to locate any matching arrest records in the data that is available to us.

**Possible Matching Arrest Records**

Name	County and State	Offense	View Details
No matching arrest records were found.			

(d)

# Case study: Recommender Systems

- Latanya Sweeney 2013, “Discrimination in Online Ad Delivery”

INSTANT CHECKMATE ADS ON REUTERS

	OBSERVED				EXPECTED		
	BLACK	WHITE	Totals		BLACK	WHITE	
Arrest Ads	291 60%	308 48%	599 53%		260	339	
Neutral Ads	197 40%	330 52%	527 47%		228	299	
Totals	488	638	1126				

INSTANT CHECKMATE ADS ON GOOGLE

	OBSERVED				EXPECTED		
	BLACK	WHITE	Totals		BLACK	WHITE	
Arrest Ads	335 92%	53 80%	388 90%		329	59	
Neutral Ads	31 8%	13 20%	44 10%		37	7	
Totals	366	66	432				

	denigration	stereotype	recognition	under-representation	ex-nomination
<b>Image search for 'CEO' yields all white men on first page of results.</b>			x	x	x
<b>Google Photo mislabels black people as 'gorillas'</b>	x				
<b>YouTube speech-to-text does not recognize women's voices</b>			x		x
<b>HP Cameras' facial recognition unable to recognize Asian people's faces</b>			x	x	x
<b>Amazon labels LGBTQ literature as 'adult content' and removes sales rankings</b>		x	x		x
<b>Word embeddings contain implicit biases [Bolukbasi et al.]</b>	x	x	x	x	x
<b>Searches for African American-sounding names yield ads for criminal background checks [Sweeney]</b>	x	x		x	

Credit: @math\_rachel, Kate Crawford

# Representation harms

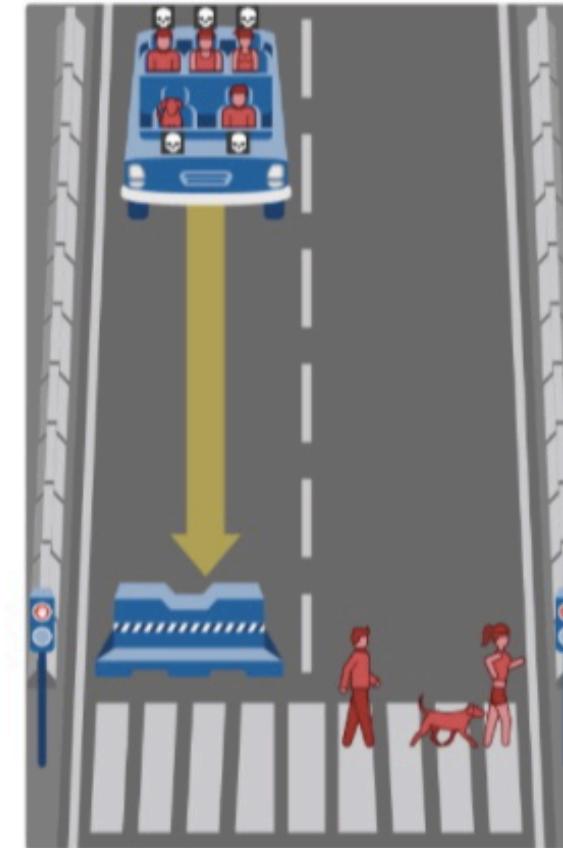
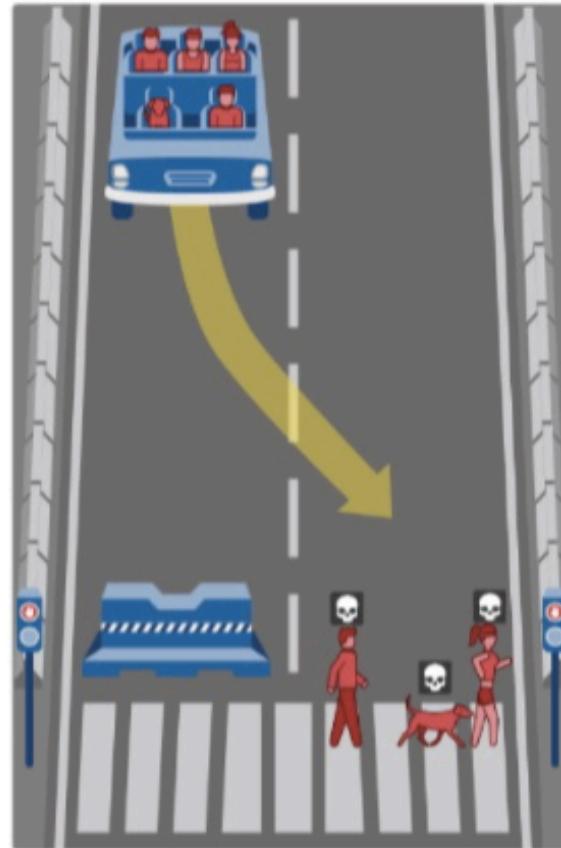
- Denigration
- Stereotype
- Recognition
- Under-representation

Credit: Solon Barocas, Kate Crawford, Aaron Shapiro, Hanna Wallach

# Case study: Physical Systems

## Self-driving cars

What should the self-driving car do?



# Case study: Physical Systems

## Self-driving cars

- In the event of an inevitable crash leading to likely loss of life, what should the car do?
- Example: Car crash will either results in death of
  - Driver and several passengers
  - Several pedestrians
- Debate: How do we choose?

# At least think about the following question

- What question can/should we ask?
- What data is OK to use?
- What is the boundary between private data and public data?
  - If something is public, should you use the data?
  - Example: Enron email corpus
- Who owns your data?
- What anonymization should be done and is the data \*really\* anonymized?

There is also hope!

*Biased Algorithms Are Easier  
to Fix Than Biased People*

[https://www.nytimes.com/2019/12/06/business/alg  
orithm-bias-fix.html](https://www.nytimes.com/2019/12/06/business/algorithm-bias-fix.html)

Ask not what AI can do, but what AI  
should do

<https://delegability.github.io/>

# THANKS



Thanks!

Amit Rege  
Cory Paik

# A MILLION

Most importantly, all of you!