



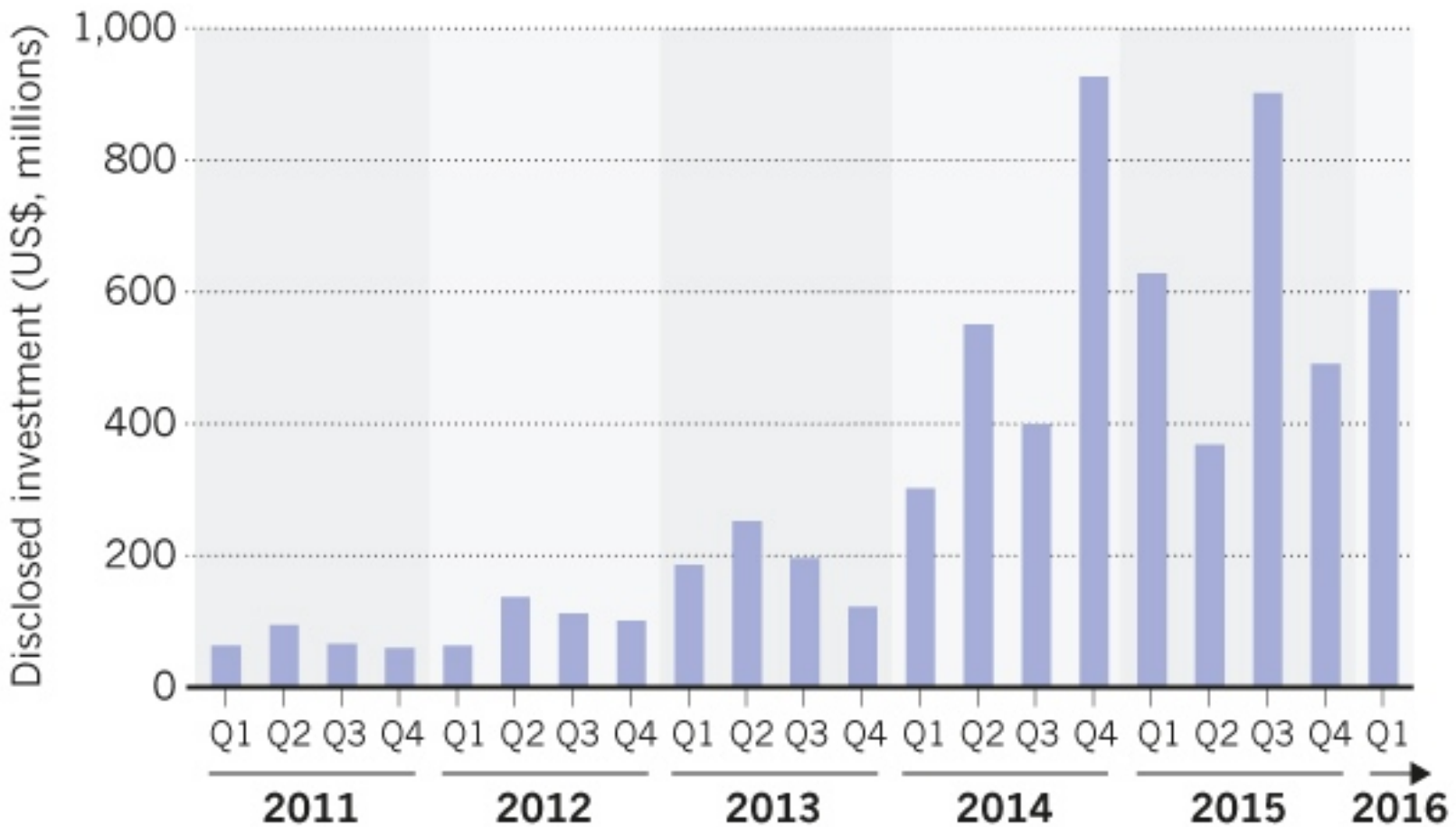
bias in a.i.

Hal Daumé III | University of Maryland | me@hal3.name | [@haldaume3](https://twitter.com/haldaume3)

the age of automated decision making

ON THE RISE

Investment in technologies that use artificial intelligence has climbed in recent years.



©nature

things can go really badly

Many Cars Tone Deaf To Women's Voices
Female voices pose a bigger challenge for voice-activated technology than men's voices

To predict and serve?
Kristian Lum, William Isaac
First published: 7 October 2016 [Full publication history](#)

Discrimination in Online Ad Delivery

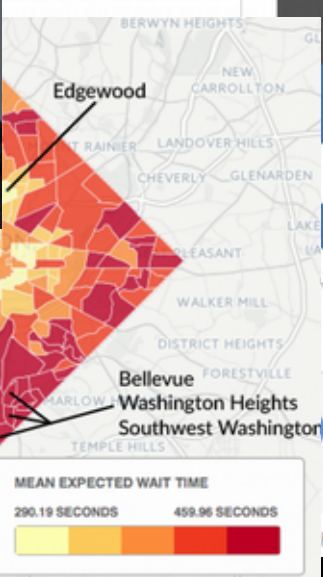
Latanya Sweeney
Harvard University
latanya@fas.harvard.edu

January 28, 2013¹



Machine Bias
There's software used across the country to predict future criminals. And it's biased against blacks.
by Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, ProPublica
May 23, 2016

Facebook Lets Advertisers Exclude Users by Race
Discrimination in areas raises some



as
has
inds of

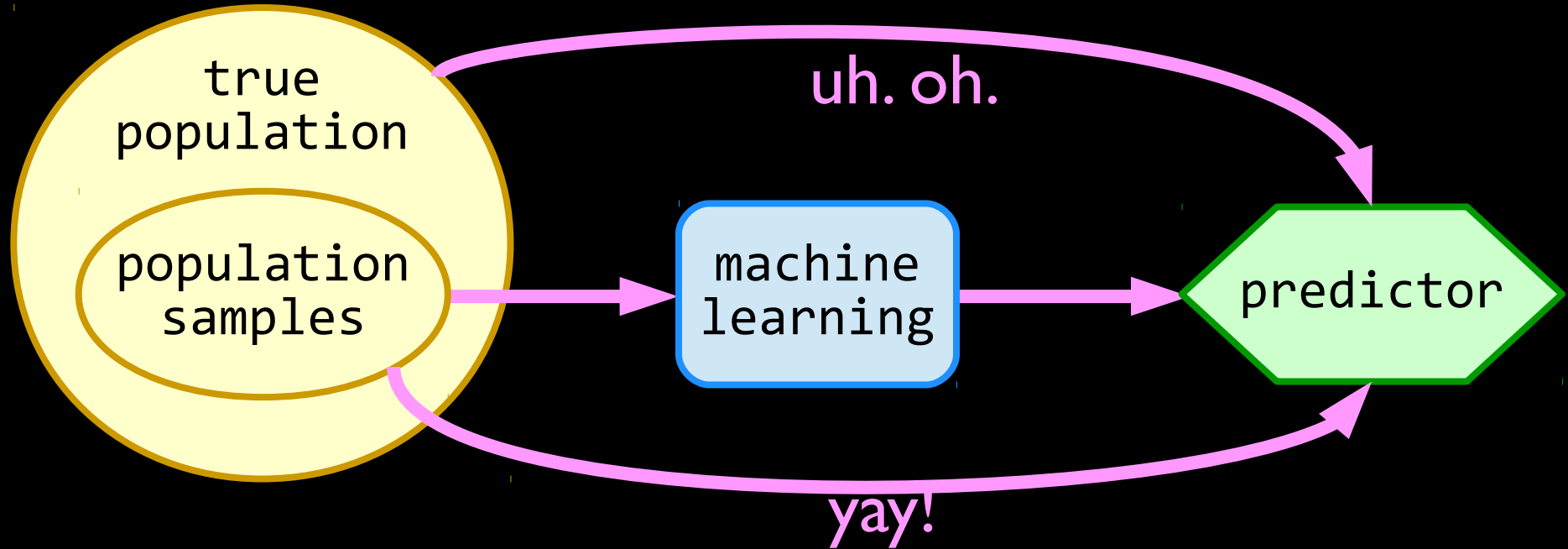
three (out of many) sources of bias

data collection

objective function

feedback loops

sample selection bias



Corinna Cortes,
*Domain adaptation
and sample bias
correction theory and
algorithm for
regression*
TCS, 2013

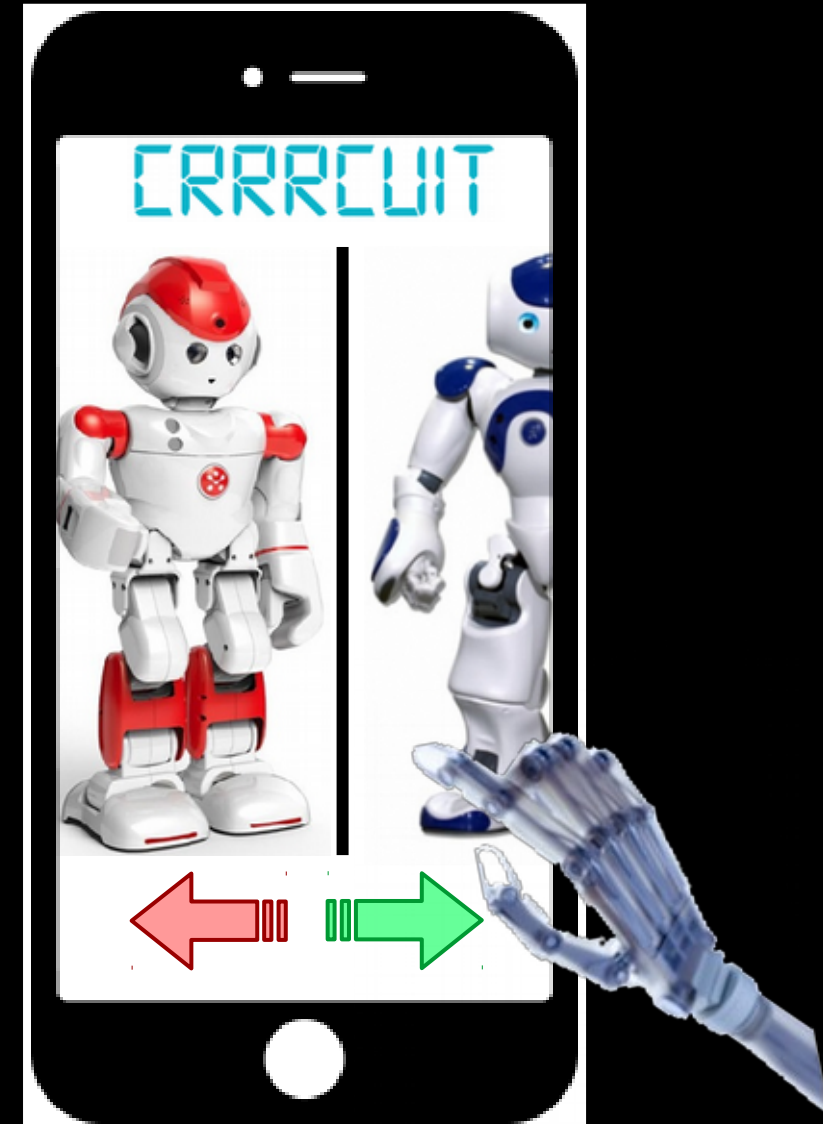
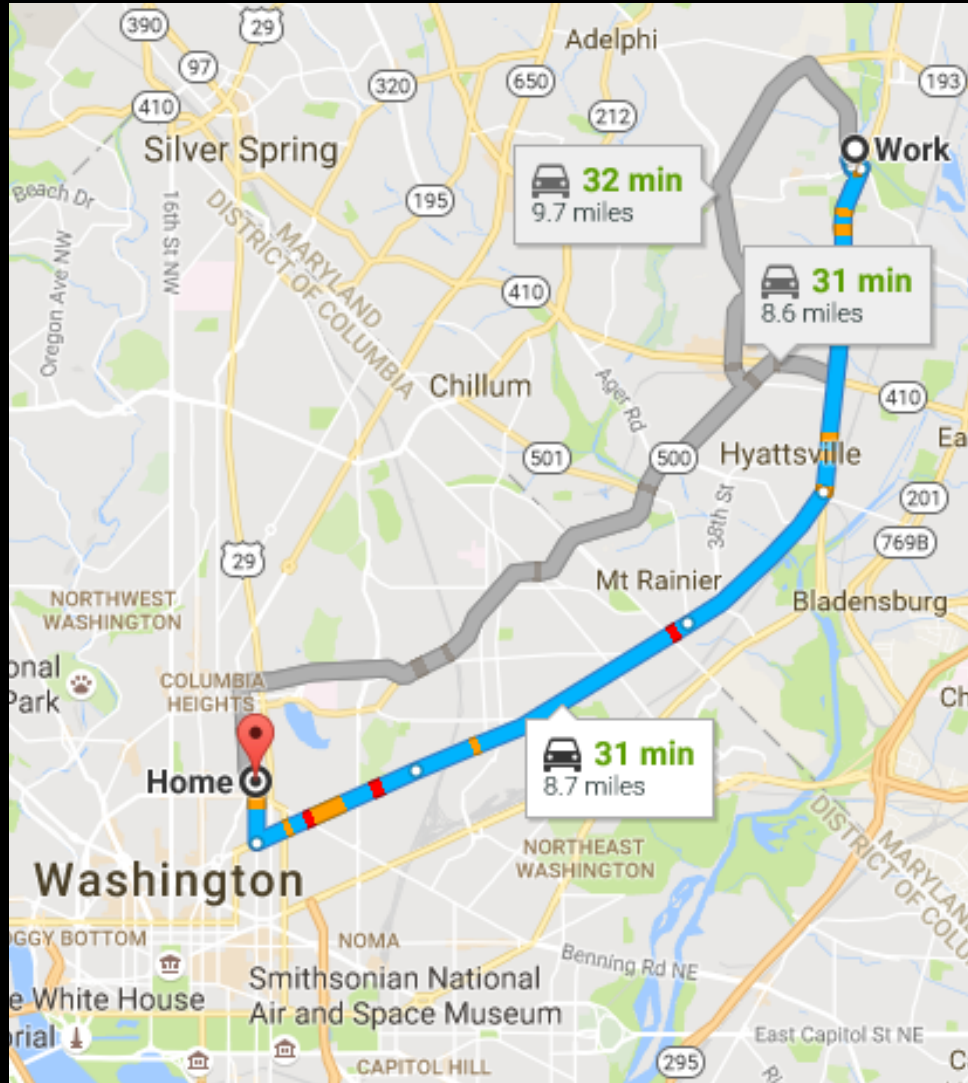
it's not just that error rate goes up...

train on electronics
rate of positive predictions

	book	dvd	elec	kit
book	1.00	1.53	1.94	1.85
dvd	1.17	1.00	1.55	1.43
elec	1.81	1.71	1.00	1.06
kit	2.19	2.29	1.55	1.00

book	.468
dvd	.509
elec	.505
kit	.564

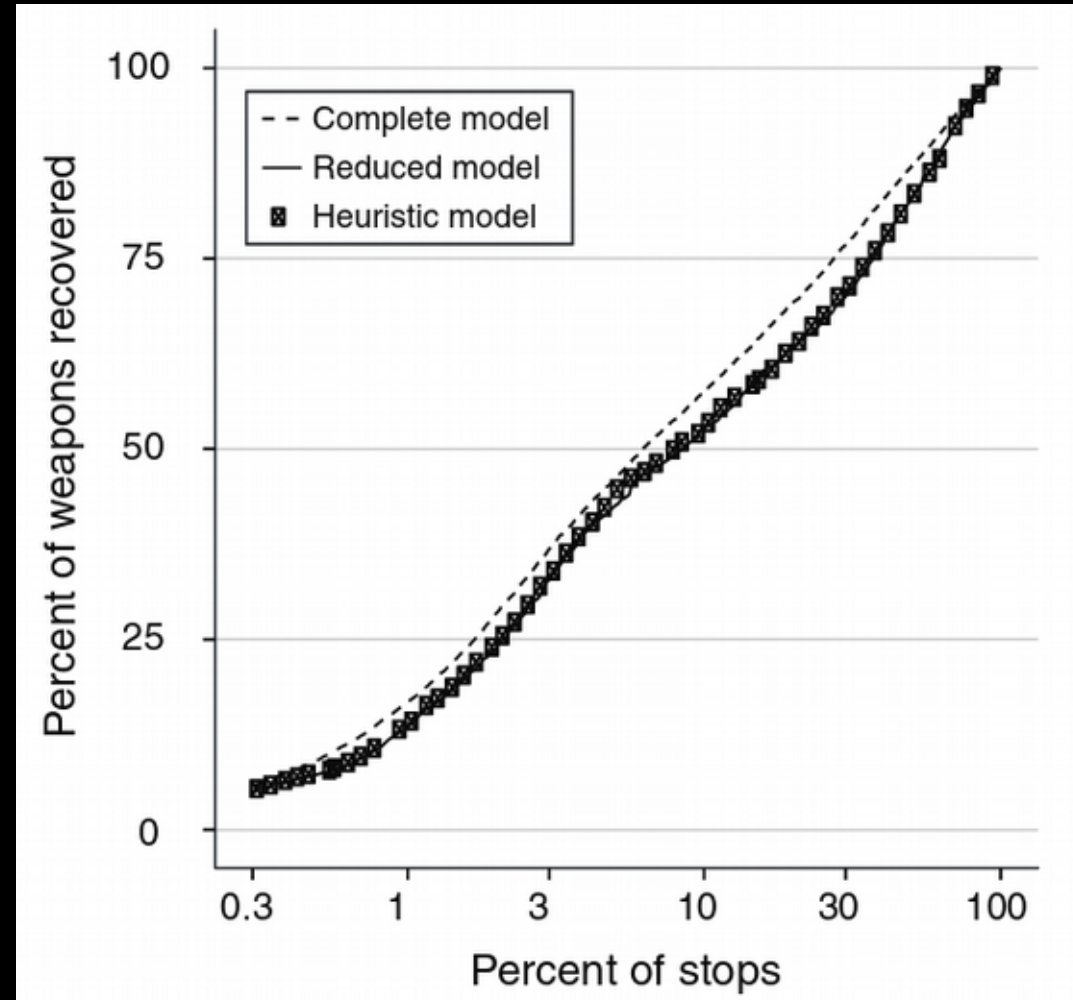
what are we optimizing for?



feedback loops in stop+frisk

Can we reduce the number of (and bias in) stops under a stop and frisk policy?

What happens if/when police officers start using this system?



Personalized risk assessments in the criminal justice system
Goel, Rao & Shroff; American Economic Review, 2016

three (out of many) sources of bias

data collection

objective function

feedback loops



Joanna Bryson
[@j2bryson](#)



Kate Crawford
[@katecrawford](#)



Nick Diakopoulos
[@ndiakopoulos](#)



Sorelle Friedler
[@kdphd](#)



Suresh Venkat
[@geomblog](#)



Hanna Wallach
[@hannawallach](#)

Fairness, Accountability &
Transparency in ML
[fatml.org](#)

Critical Algorithm Studies:
A Reading List
[socialmediacollective.org/reading-lists/critical-algorithm-studies](#)