KAIST Summer Session 2018

Module 1. Research Design for Data Analytics

Reflection on Predictive Analytics

KAIST College of Business

Jiyong Park

9 July, 2018

What is Predictive Analytics?

Prediction is based on Correlation

- Prediction does not require causality. Correlation is enough for prediction.
  - ➢ (Example) Prediction of influenza using Google search query (Ginsberg et al. 2009)
    - Internet search for ‘influenza complication’ is correlated to occurring influenza.
    - But, searching for ‘influenza complication’ does not cause the influenza.

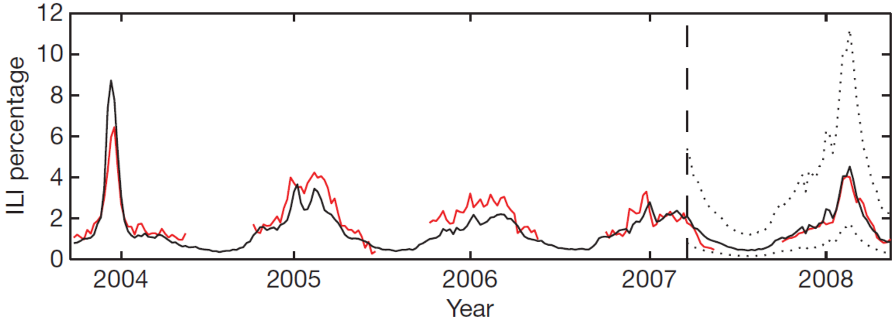


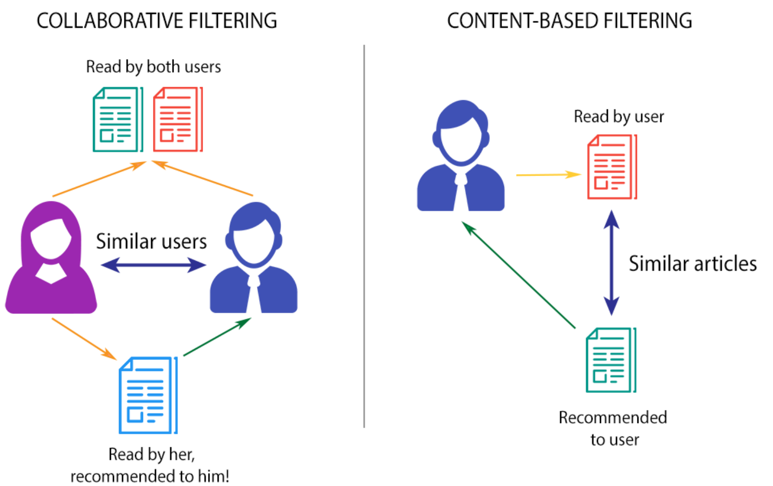
Figure 2 | A comparison of model estimates for the mid-Atlantic region (black) against CDC-reported ILI percentages (red), including points over which the model was fit and validated. A correlation of 0.85 was obtained over 128 points from this region to which the model was fit, whereas a correlation of 0.96 was obtained over 42 validation points. Dotted lines indicate 95% prediction intervals. The region comprises New York, New Jersey and Pennsylvania.

Search query topic	Top 45 queries	
	n	Weighted
Influenza complication	11	18.15
Cold/flu remedy	8	5.05
General influenza symptoms	5	2.60
Term for influenza	4	3.74
Specific influenza symptom	4	2.54
Symptoms of an influenza complication	4	2.21
Antibiotic medication	3	6.23
General influenza remedies	2	0.18
Symptoms of a related disease	2	1.66
Antiviral medication	1	0.39
Related disease	1	6.66
Unrelated to influenza	0	0.00
Total	45	49.40

Ginsberg, J., Mohebbi, M.H., Patel, R.S., Brammer, L., Smolinski, M.S. and Brilliant, L., 2009. Detecting Influenza Epidemics using Search Engine Query Data. *Nature*, 457(7232), p.1012.

Recommendation Systems Aim at Prediction

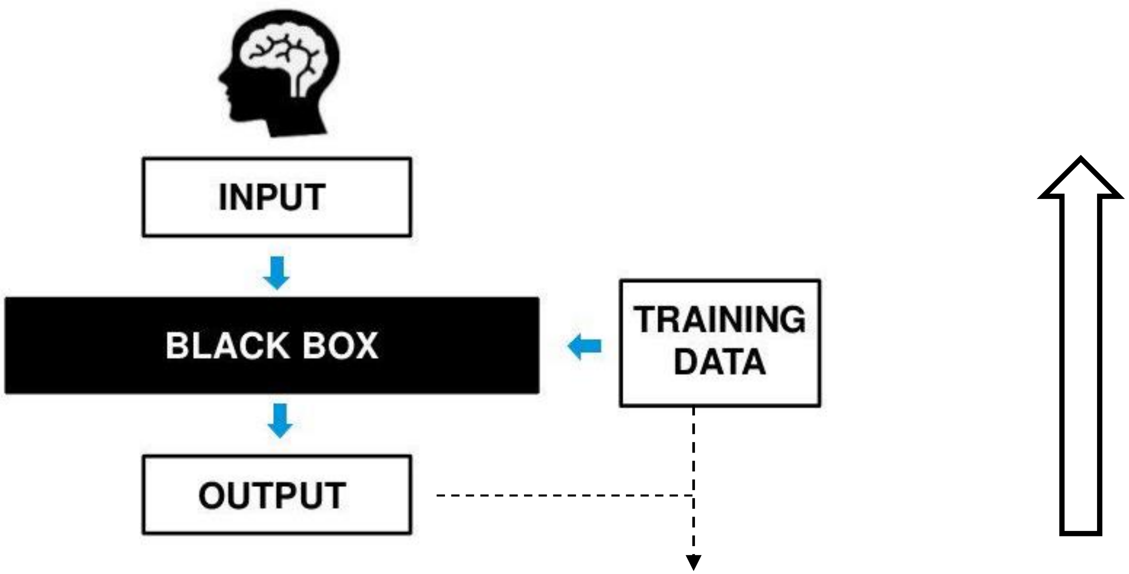
- Recommendation systems are based on associations or links between a set of users and a set of items.
  - ➢ A link between a user and a product means that the user has indicated an interest in the product in some fashion.
  - ➢ The problem is to suggest other items to a given user that he or she may also be interested in, based on the data across all users.



KAIST COLLEGE OF BUSINESS

Deep Learning and Machine Learning Aim at Prediction

- Supervised learning takes the form of a collection of (x, y) pairs and the goal is to produce a prediction y^* in response to a query x^*.
  - ➢ The goal of back-propagation algorithm is to reduce the loss function, or to best imitate (predict) the training data.

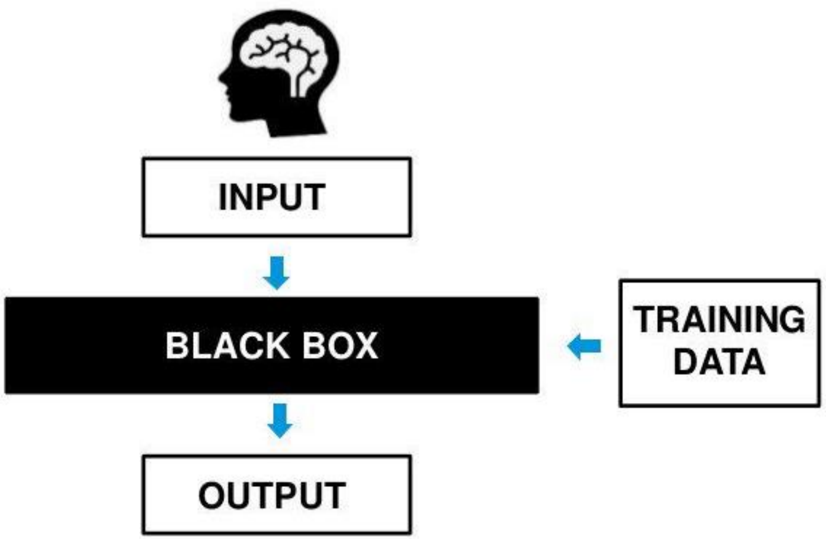


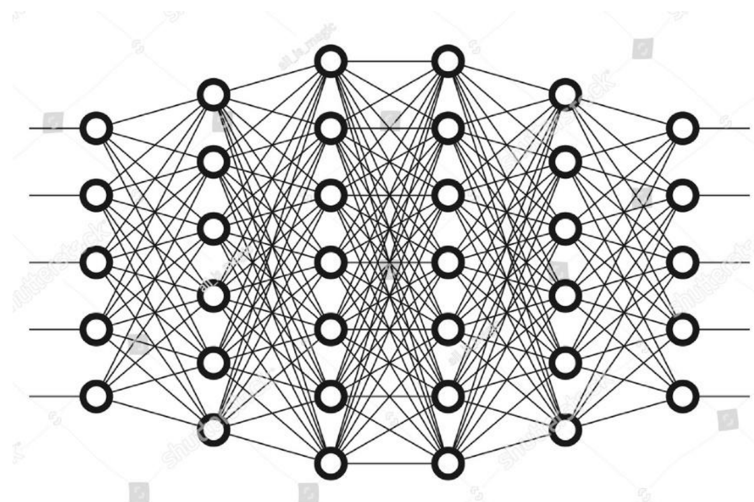
Tuning parameters of neural networks to minimize the loss function

$$Loss\ function = f(output - training\ data)$$

Deep Learning is a Particularly Dark Black Box

- Deep learning is to imitate and predict man-made data, not to understand.
  - ➢ Large neural networks learn to make decisions by subtly adjusting up to hundreds of millions of numerical weights that interconnect their artificial neurons.





Deep Learning, Turing Test, and Chinese Room

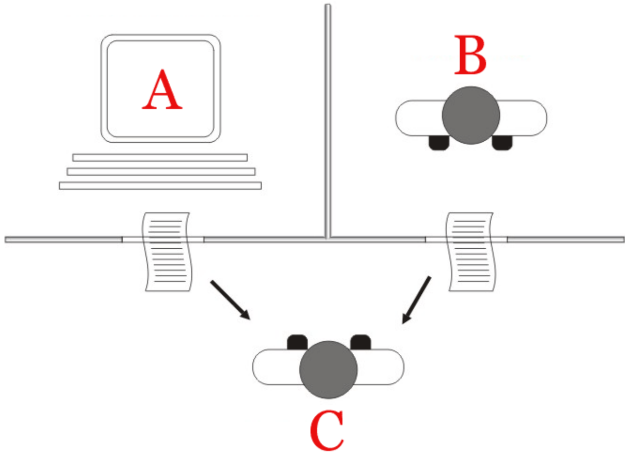
- Alan Turing proposed the concept of digital computers at 1936

Alan Turing

1. Alan Turing and universal machine (https://www.youtube.com/watch?v=xvppZNe6jAQ)
2. Turing test (imitation game) (https://www.youtube.com/watch?v=Vs7Lo5MKIws)

Deep Learning, Turing Test, and Chinese Room

- Turing test or imitation game (Can machines think?)
  - A human evaluator would judge natural language conversations between a human and a machine designed to generate human-like responses. If the evaluator cannot reliably tell the machine from the human, the machine is said to have passed the test (to be able to think).
  - The test results do not depend on the ability to give correct answers to questions, only how closely one's answers resemble those a human would give.

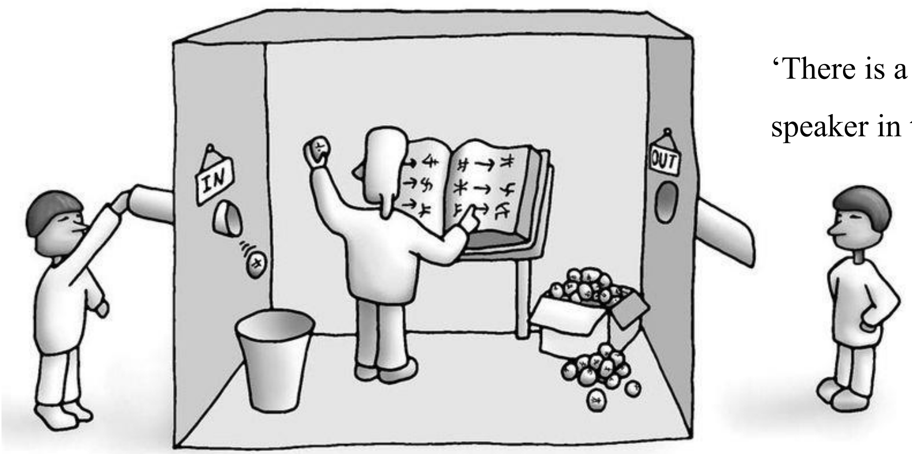


https://en.wikipedia.org/wiki/Turing_test

Deep Learning, Turing Test, and Chinese Room

- Chinese room (Can machines understand?)
  - Thought experiment related to Turing test, proposed by philosopher John Searle
  - Does the machine literally "understand" Chinese? Or is it merely simulating the ability to understand Chinese?
  - One of refutations for the Chinese room argument

    'Although the man does not, the whole system can understand Chinese'

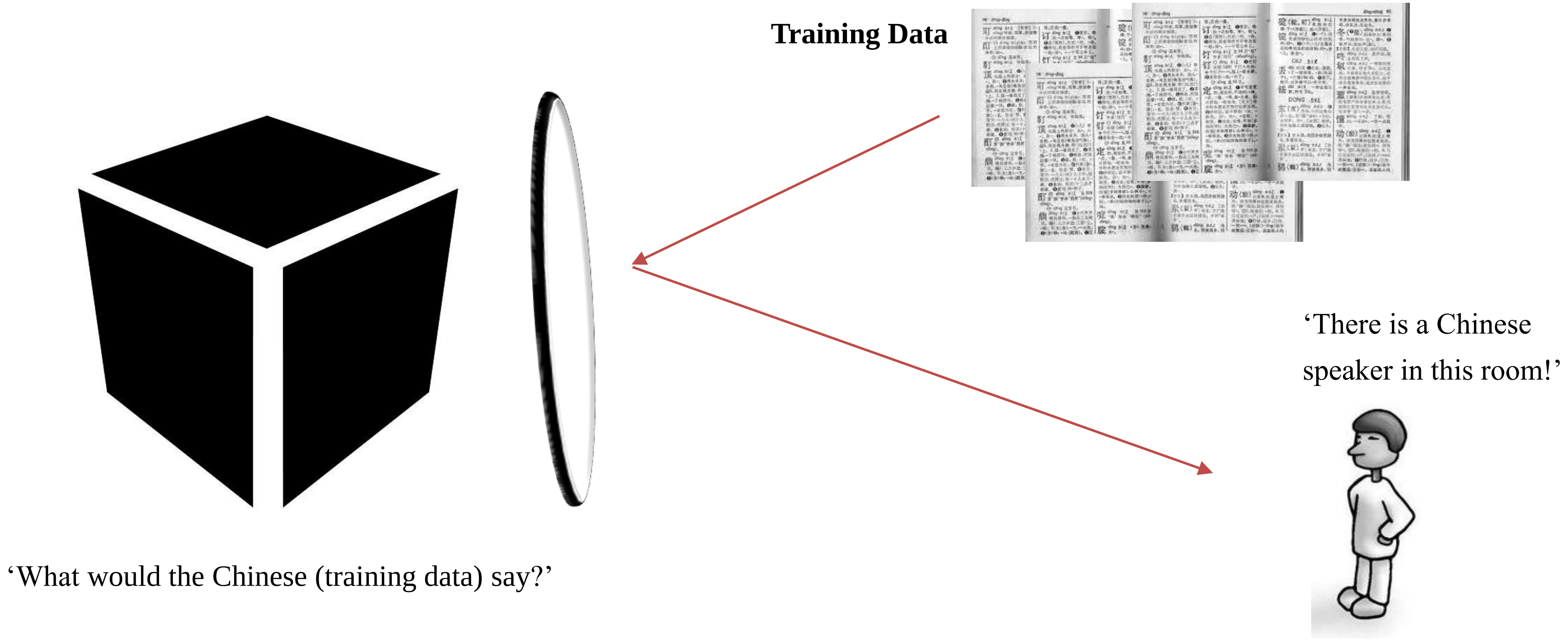


'There is a Chinese speaker in this room!'

https://en.wikipedia.org/wiki/Chinese_room

Deep Learning is the "Chinese Mirror" (Mirroring or Predicting)

- Deep learning is designed to pass the Turing test, but through a different way
  - "It can be far easier to train a system by showing it examples of desired input-output behavior than to program it manually by anticipating the desired response for all possible inputs." (Jordan and Mitchell 2015, p. 255)

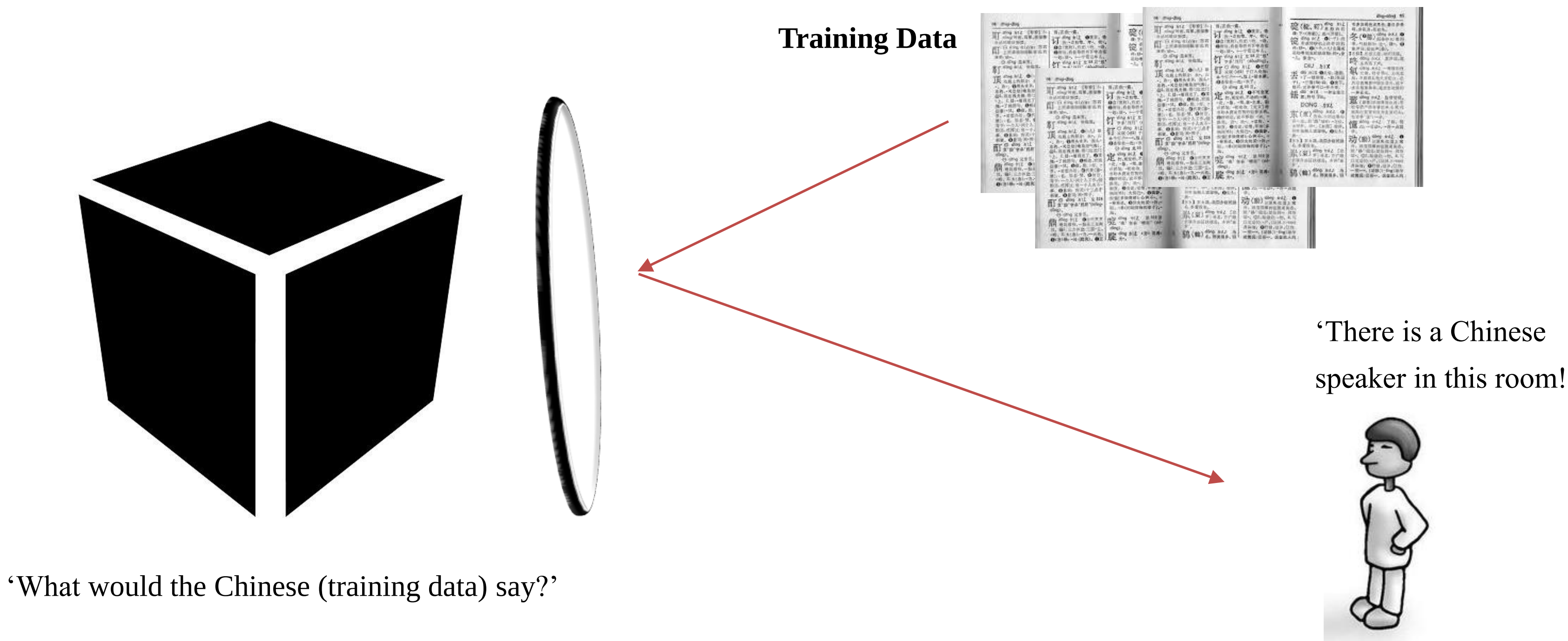


Jordan, M.I. and Mitchell, T.M., 2015. Machine Learning: Trends, Perspectives, and Prospects. *Science*, 349(6245), pp.255-260.

Predictive Bias in Machine Learning

Source of Predictive Bias

- There are two primary sources of predictive bias in machine learning
  - Training data: *"What data does the algorithm learn?*
  - Algorithm: *"How does the algorithm imitate the training data?"*



Predictive Bias from Training Data

- Machines may learn not only human language, but also absorb ingrained prejudices concealed within the patterns of language use in the training data.

COGNITIVE SCIENCE

Semantics derived automatically from language corpora contain human-like biases

Aylin Caliskan,[1*] Joanna J. Bryson,[1,2*] Arvind Narayanan[1*]

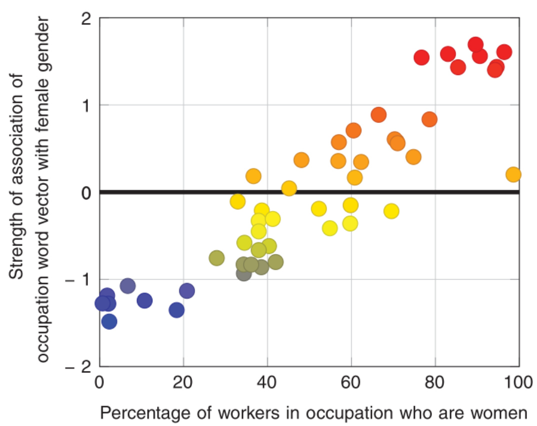


Fig. 1. Occupation-gender association. Pearson's correlation coefficient $\rho = 0.90$ with $P < 10^{-18}$.

White-sounding name
→ Bias-neutral algorithms link this name to a positive sentiment

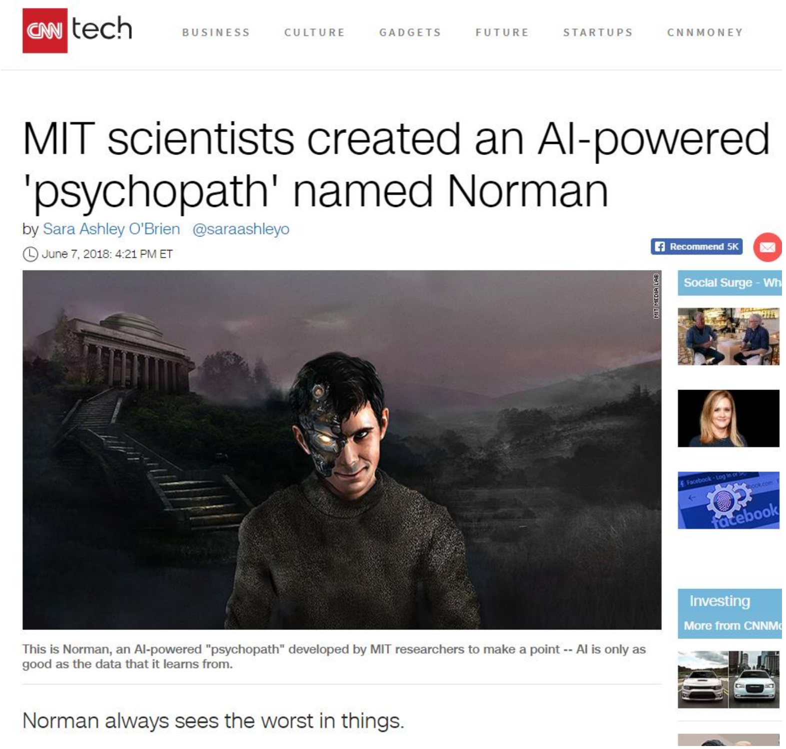
```
text_to_sentiment("My name is Emily")
2.2286179364745311

text_to_sentiment("My name is Heather")
1.3976291151079159

text_to_sentiment("My name is Yvette")
0.98463802132985556

text_to_sentiment("My name is Shaniqua")
-0.47048131775890656
```

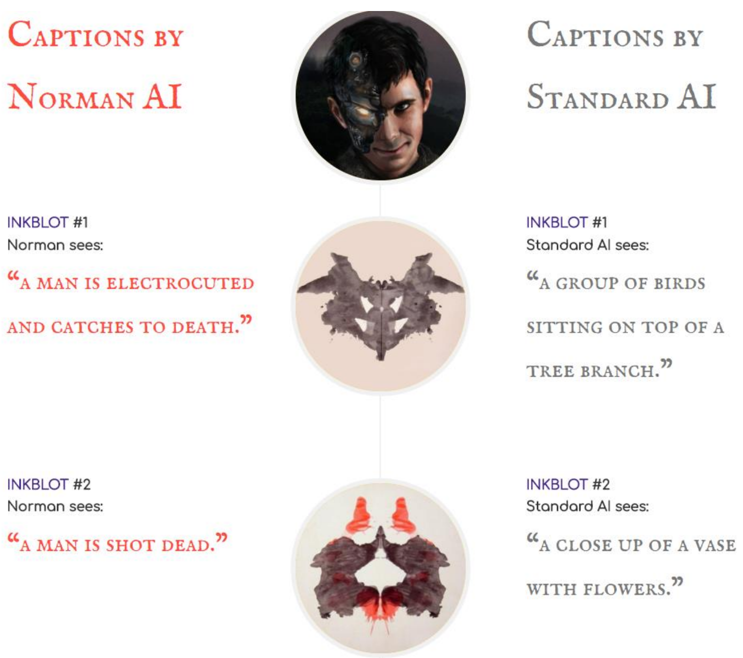
Black-sounding name
→ Bias-neutral algorithms link this name to a negative sentiment

Caliskan, A., Bryson, J.J. and Narayanan, A., 2017. Semantics Derived Automatically from Language Corpora Contain Human-Like Biases. *Science*, 356(6334), pp.183-186.
"How to make a racist AI without really trying," https://blog.conceptnet.io/posts/2017/how-to-make-a-racist-ai-without-really-trying/

Predictive Bias from Training Data

- MIT scientists demonstrate that "artificial intelligence is only as good as the data that is learns from" by creating a psychopath AI named Norman.



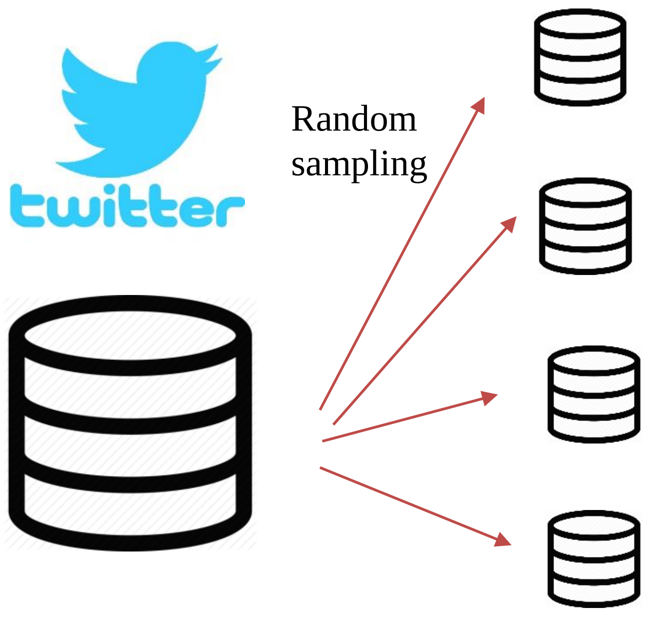


http://money.cnn.com/2018/06/07/technology/mit-media-lab-normal-ai/index.html

http://norman-ai.mit.edu

Predictive Bias from Training Data

- Training data can vary depending on unobserved or observed factors which researchers might not be recognize.
  - ➢ (Example) Twitter sample bias (González-Bailón et al. 2014)



González-Bailón et al. (2014) find that Twitter samples seem not random and different depending on (i) Search API vs. Streaming API and (ii) number of search keywords.

The authors suggest that the selecting search keywords seems to play more prominent role in creating biases in samples of large online social media than selecting APIs.

González-Bailón, S., Wang, N., Rivero, A., Borge-Holthoefer, J. and Moreno, Y., 2014. Assessing the Bias in Samples of Large Online Networks. *Social Networks*, 38, pp.16-27.

Predictive Bias from Training Data

- Predictive modeling may work well in some samples, but not work in some other samples.
  - ➢ (Example) Political orientations in Twitter (Cohen and Ruth 2013)

twitter

Predictive model of political orientation

Table 6: Performance results of training our SVM on one dataset and inferring on another, italicized are the averaged 10-fold cross-validation results

Dataset	Figures	Active	Modest
Figures	*91%*	72%	66%
Active	62%	*84%*	69%
Modest	54%	57%	*68%*

US politicians

Users with self-reported political orientation

Ordinary users

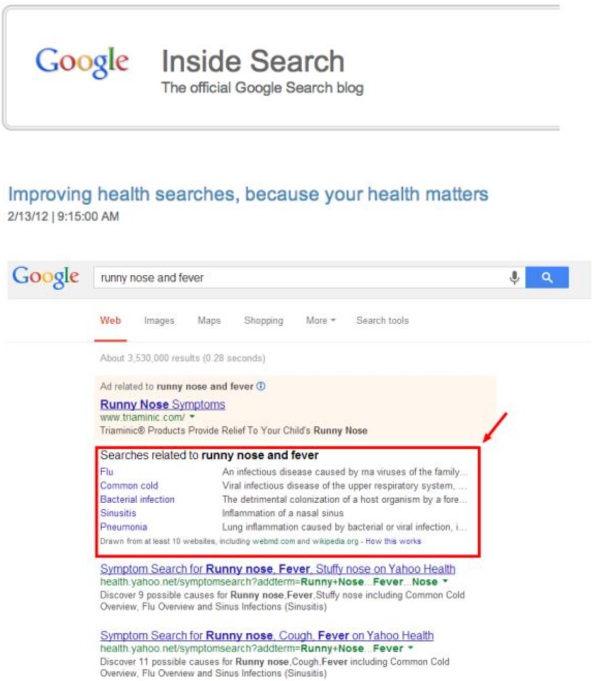
Cohen and Ruth (2013) find that the prediction tasks were largely successful when the samples were drawn from the population of politicians (*Figures*) or politically active users (*Active*). However, the same approach revealed much poorer outcomes for more normal, less politically vocal users. In addition, classifiers do not transfer across types of users.

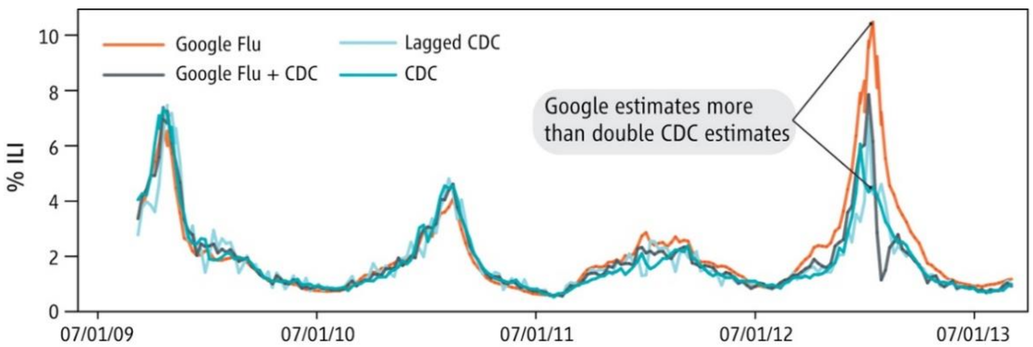
In predictive analytics, understanding the training data is critical in interpreting the results and assessing its performance (Boyd and Crawford 2012).

Cohen, R. and Ruths, D., 2013. Classifying Political Orientation on Twitter: It's Not Easy!. In *Seventh International AAAI Conference on Weblogs and Social Media*.
Boyd, D. and Crawford, K., 2012. Critical Questions For Big Data. *Information. Communication & Society*, 15(5), pp.662-679.

Predictive Bias from Algorithm

- Algorithm producing the data (and thus user utilization) has been modified by the service provider in accordance with their business model.
  - ➢ (Example) Google Flu and change in Google search engine (Lazer et al. 2014)



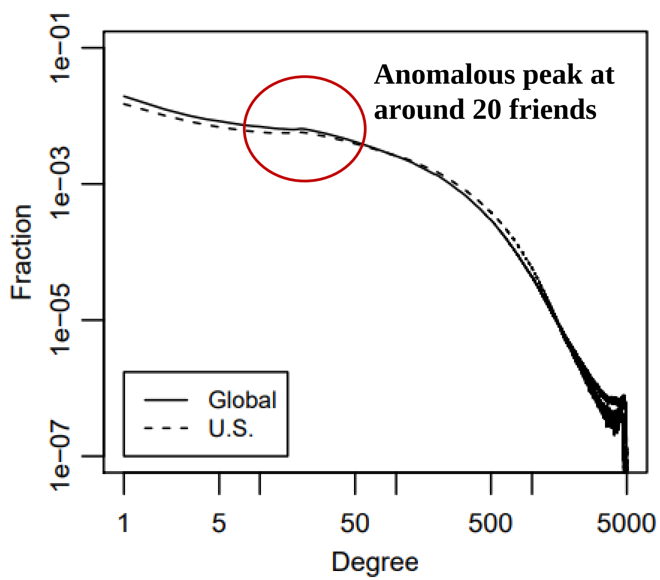


Google's changed policy at February 2012 may affect the performance of Google Flu algorithm which was developed before the policy change.

Lazer, D., Kennedy, R., King, G. and Vespignani, A., 2014. The Parable of Google Flu: Traps in Big Data Analysis. *Science*, 343(6176), pp.1203-1205.

Predictive Bias from Algorithm

- Algorithm designers can introduce patterns into data, even if not intended.
  - If researchers do not recognize the algorithmic confounding, they could arrive at wrong conclusions.
  - (Example) Number of Facebook friends (Ugander et al. 2011)



Someone might interpret this anomaly as the "sympathy group" of maximum 20 individuals with whom one can maintain meaningful relations and contact on a regular basis in Facebook.

However, this may be just a result of algorithm design. "This kink is due to forces within the Facebook product to encourage low friend count individuals in particular to gain more friends until they reach 20 friends." (Ugander et al. 2011, p. 3)

Ugander, J., Karrer, B., Backstrom, L. and Marlow, C., 2011. The Anatomy of the Facebook Social Graph. *arXiv preprint arXiv:1111.4503.*

Keep the Bias in Mind When Applying Machine Learning

Science

TECHNOLOGY AND THE ECONOMY

What can machine learning do? Workforce implications

Profound change is coming, but roles for humans remain

By Erik Brynjolfsson[1,2] *and* Tom Mitchell[3] | engine and electricity, which spawns a pleth-

(1) Learning a function that maps well-defined inputs to well-defined outputs

(2) Large data sets exist or can be created containing input-output pairs

(3) The task provides clear feedback with clearly definable goals and metrics

(4) No long chains of logic and reasoning that depend on diverse background knowledge or common sense

(5) No need for detailed explanation of how the decision was made

(6) A tolerance for error and no need for provably correct or optimal solutions

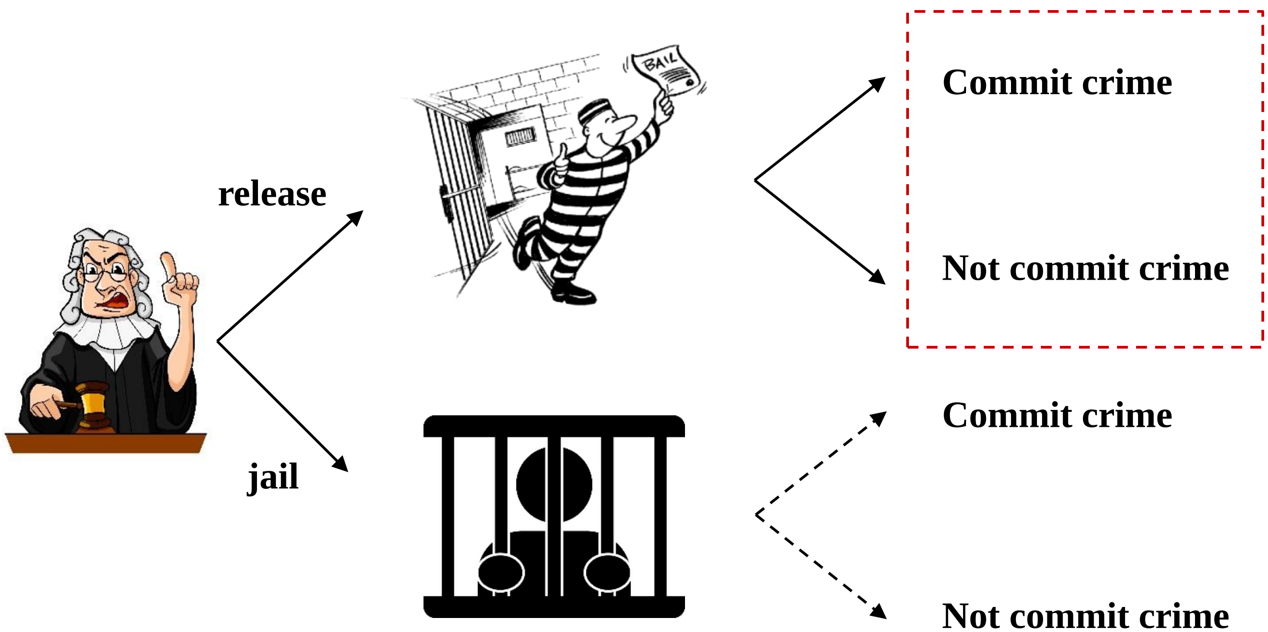
(7) The phenomenon or function being learned should not change rapidly over time

(8) No specialized dexterity, physical skills, or mobility required

Brynjolfsson, E. and Mitchell, T., 2017. What Can Machine Learning Do? Workforce Implications. *Science*, 358(6370), pp.1530-1534.

Assessing a Predictive Model is not Straightforward

- When comparing machine learning algorithms to existing decision-making systems, we can only observe the “selected” outcomes, which are used to train the algorithms (i.e., *selective labels*).
    - ➢ (Example) Bail decisions (jail or release) (Kleinberg et al. 2017)



We only have a label providing the correct answer for defendants judges chose to release.

What about defendants that judges chose not to release?

Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J. and Mullainathan, S., 2017. Human Decisions and Machine Predictions. *Quarterly Journal of Economics*, 133(1), pp.237-293.

Assessing a Predictive Model is not Straightforward

- The best way to solve this problem is to conduct a randomized controlled experiment or quasi-experiment, through which we could directly compare whether machine learning-based decision-making lead to better outcomes than those made on comparable cases using the current system.
  - ➢ (Example) Bail decisions (jail or release) (Kleinberg et al. 2017)

Lenient judges
(Release rate = 60%)

+

Algorithm prediction

→ Additional 10% of defendants to jail

Treatment groups

Stringent judges
(Release rate = 50%)

Control groups

Key identification assumption: defendants are as-good-as-randomly assigned to judges who differ in leniency

Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J. and Mullainathan, S., 2017. Human Decisions and Machine Predictions. *Quarterly Journal of Economics*, 133(1), pp.237-293.

Transparency and Algorithmic Fairness

Algorithmic Fairness

- Although designers do not intend to do, algorithms could become discriminatory, possibly due to training data or algorithm itself.
  - ➢ (Example) Online ad delivery: racial discrimination (Sweeney 2013)

Figure 1. Ads from a Google search of three different names beginning with first name "Latanya."

Figure 2. Ad from a search of three different names beginning with the first name "Jill."

Sweeney, L., 2013. Discrimination in Online Ad Delivery. *Communications of the ACM*, 56(5), pp.44-54.

Algorithmic Fairness

- Although designers do not intend to do, algorithms could become discriminatory, possibly due to training data or algorithm itself.
  - ➢ (Example) Online ad delivery: gender discrimination (Lambrecht and Tucker 2018)

Online ad on Facebook

Age Group	Male Impr.	Female Impr.
Age 18-24	746719	649590
Age 25-34	662996	495996
Age 35-44	412457	283596
Age 45-54	307701	224809
Age 55-64	209608	176454
Age 65+	192317	153470

Lambrecht and Tucker (2018) find that Science, Technology, Engineering and Math (STEM) ads are delivered to women less than men on Facebook.

(1) Bias from training data

The underlying data used to train an algorithm may be biased, reflecting a history of discrimination.

(2) Bias from algorithm

The algorithm's decision to display the STEM ad less often to women than to men was a reflection of the economics of ad delivery, embedded in ad bidding algorithms.

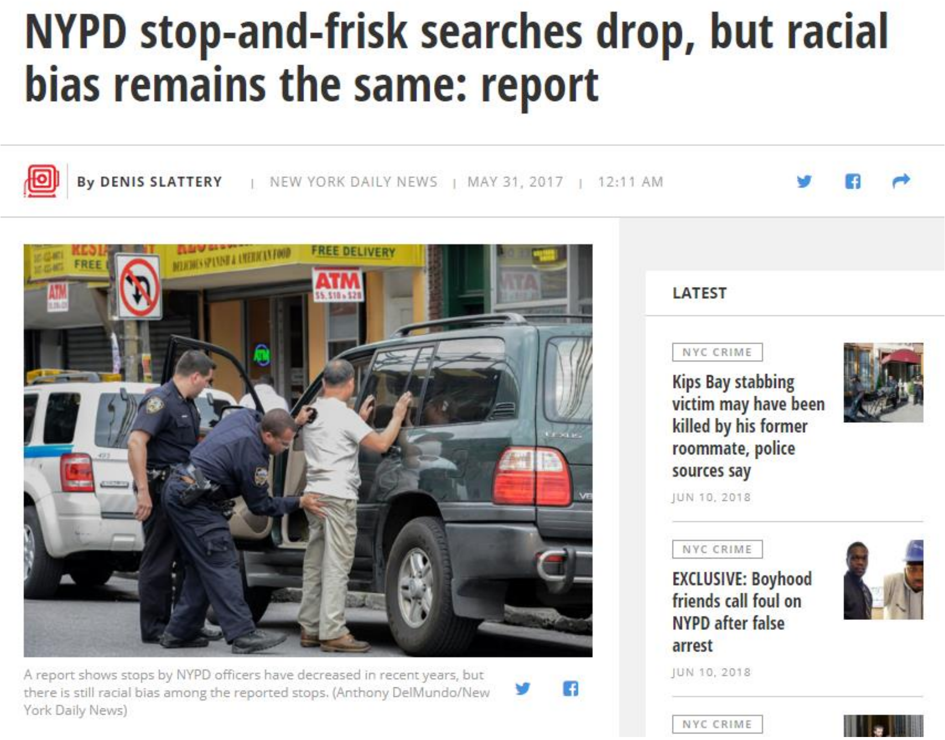
→ The authors conclude that this is the case.

Lambrecht, A. and Tucker, C.E., 2018. Algorithmic Bias? An Empirical Study into Apparent Gender-Based Discrimination in the Display of STEM Career Ads. *Working Paper*. Available at SSRN: https://ssrn.com/abstract=2852260

Predictive Decision-Making and Ethical Debate

- Example: Predictive policing
  - ➢ The movie "Minority Report" become a reality. (e.g., NYPD stop & frisk)

- Ethical Debate
  - ➢ Should a person be treated differently simply because they share attributes with others who have higher risks of crime?
  - ➢ The black box nature of machine learning even worsens this controversy.



NYPD stop-and-frisk searches drop, but racial bias remains the same: report

By DENIS SLATTERY | NEW YORK DAILY NEWS | MAY 31, 2017 | 12:11 AM

A report shows stops by NYPD officers have decreased in recent years, but there is still racial bias among the reported stops. (Anthony DelMundo/New York Daily News)

LATEST

NYC CRIME

Kips Bay stabbing victim may have been killed by his former roommate, police sources say

JUN 10, 2018

NYC CRIME

EXCLUSIVE: Boyhood friends call foul on NYPD after false arrest

JUN 10, 2018

NYC CRIME

http://www.nydailynews.com/new-york/nyc-crime/nypd-stop-and-frisk-tactics-drop-racial-bias-lingers-report-article-1.3208042

Conclusion

Properly Built Algorithm Can Solve the Problem

- Example: Criminal justice and racial discrimination
  - ➢ For the case of bail decisions, Kleinberg et al. (2017) argue that the algorithm can achieve the same crime rate as the judges buy by jailing 40.8% fewer minorities, including 38.8% fewer blacks and 44.6% fewer Hispanics.

DISPARATE IMPACT OF JAILING ADDITIONAL DEFENDANTS BY PREDICTED RISK

	Judges			Ensure algorithm		
	relative to most lenient quintile			matches judge Δ Crime		jails no more overall or black
	Δ Jail		Δ Crime	Δ Jail		Δ Crime
	Overall	Black		Overall	Black	
Second quintile	0.066	0.079	−0.099	0.027	0.037	−0.195
Third quintile	0.096	0.114	−0.137	0.042	0.054	−0.263
Fourth quintile	0.135	0.162	−0.206	0.068	0.085	−0.351
Fifth quintile	0.223	0.249	−0.307	0.112	0.137	−0.483

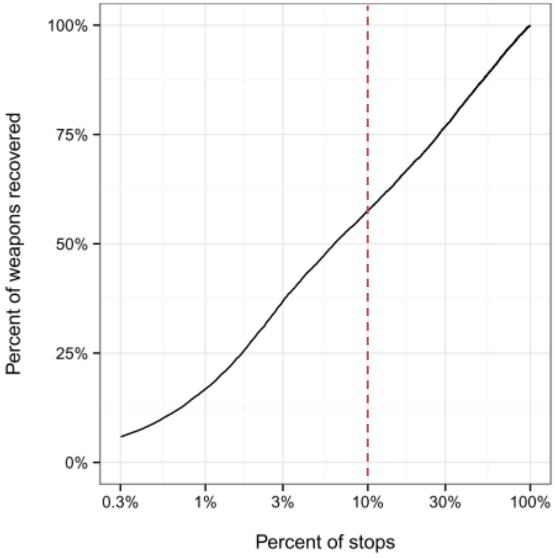
Increased percentage to jail, judged by humans

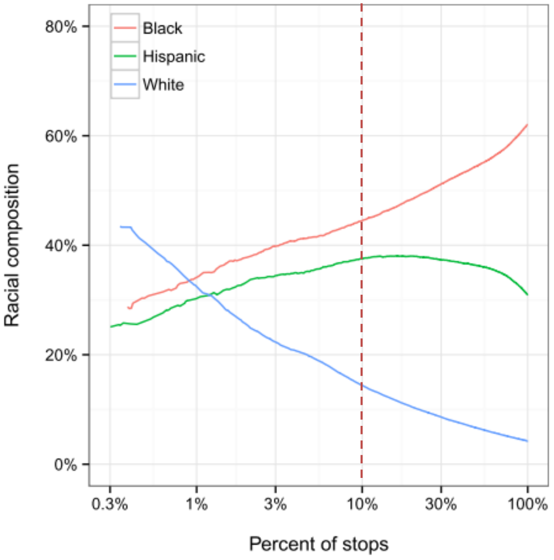
Increased percentage to jail, predicted by machines

Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J. and Mullainathan, S., 2017. Human Decisions and Machine Predictions. *Quarterly Journal of Economics*, 133(1), pp.237-293.

Properly Built Algorithm Can Solve the Problem

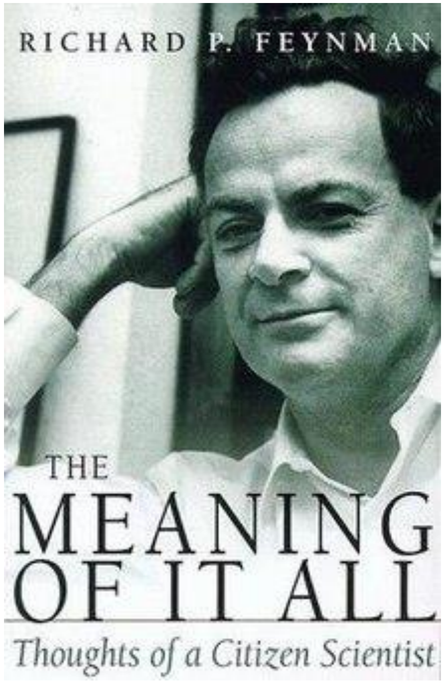
- Example: Criminal justice and racial discrimination
  - ➢ For the case of NYC stop and frisk, Goel et al. (2016) find that the best 10% of stops (based on predictive models) result in 58% of weapons recovered. Since low likelihood stops disproportionately involve black suspects, reducing the number of stops results in lowering the overall proportion of stopped suspects who are black.





Goel, S., Rao, J.M. and Shroff, R., 2016. Precinct or Prejudice? Understanding Racial Disparities in New York City's Stop-and-Frisk Policy. *The Annals of Applied Statistics*, 10(1), pp.365-394.

It Depends on How Researchers Deal With



Richard Feynman

"In a way it is a key to the gates of heaven, and the same key opens the gates of hell, and we do not have any instructions as to which is which gate.

Shall we throw away the key and never have a way to enter the gates of heaven? Or shall we struggle with the problem of which is the best way to use the key?

That is, of course, a very serious question, but I think that we cannot deny the value of the key to the gates of heaven."

Fairness, Accountability, and Transparency in Machine Learning

- Fairness, accountability, and transparency (FAT) in machine learning is a very active research area in computer science, as well as in information systems.
  - ➢ (Example) FAT ML workshop (https://www.fatml.org/)

Bringing together a growing community of researchers and practitioners concerned with fairness, accountability, and transparency in machine learning

The past few years have seen growing recognition that machine learning raises novel challenges for ensuring non-discrimination, due process, and understandability in decision-making. In particular, policymakers, regulators, and advocates have expressed fears about the potentially discriminatory impact of machine learning, with many calling for further technical research into the dangers of inadvertently encoding bias into automated decisions.

At the same time, there is increasing alarm that the complexity of machine learning may reduce the justification for consequential decisions to "the algorithm made me do it."

The annual event provides researchers with a venue to explore how to characterize and address these issues with computationally rigorous methods.

Fairness, Accountability, and Transparency in Machine Learning

- Fairness, accountability, and transparency (FAT) in machine learning is a very active research area in computer science, as well as in information systems.
  - ➢ (Example) INFORMS CIST (http://www.cistconf.org/)

Conference on Information Systems and Technology (CIST) 2018 Call for Papers

TAKE IS/ML for Business and Society: Transparent, Accountable, Kind, and Efficient IS/ML for Business and Society

The recent advancement of information technology has enabled artificial intelligence to rival and, in some cases, surpass human intelligence. Our daily life becomes increasingly empowered while threatened at the same time by artificial intelligence and machine learning; new economic and social paradigm also began to emerge. As we approach this critical junction, CIST 2018 calls for innovative ideas on the transformational power of information technology, artificial intelligence and machine learning on individual behavior, business strategies, economic relationships, and social changes. We also call for innovative ideas on new interplay between machine and human as well as new ways to address challenges and threats brought by the new machine age.

End of Document