Algorithmic Accountability and Transparency

Jeanna Matthews

Mujeres en Data Science

16 de octubre 2017

Clarkson
UNIVERSITY
defy convention™

Data&Society
datasociety.net

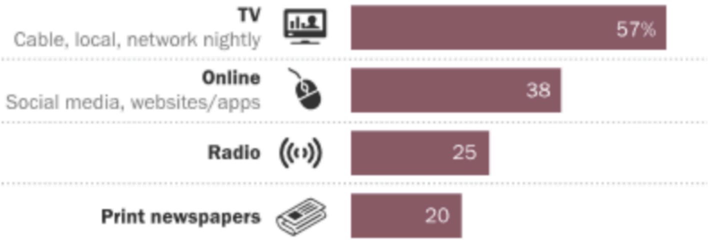
acm

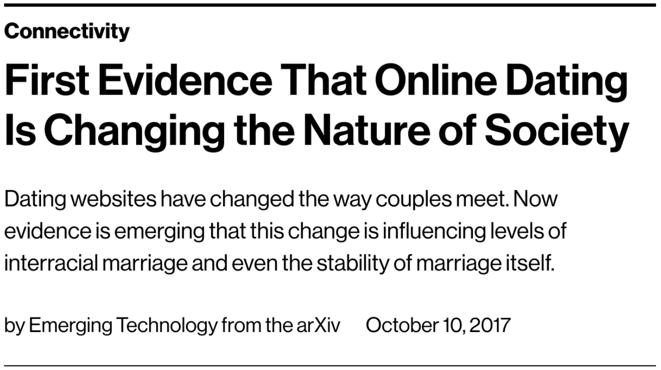
A bit about me

- Computer Science professor at Clarkson University
- Fellow at Data and Society
- Co-chair of the US-ACM Subcommittee on Algorithmic Accountability and Transparency

Algorithms and Platforms Reshaping Society

- Big-data trained algorithms increasingly used for big life decisions
  - Hiring, housing, policing, public resources, etc.
- Platforms like Facebook, Twitter, Uber making profound impacts on our personal and public relationships
  - How do we find a job? How do we get our news? How do we find a spouse?




Connectivity

First Evidence That Online Dating Is Changing the Nature of Society

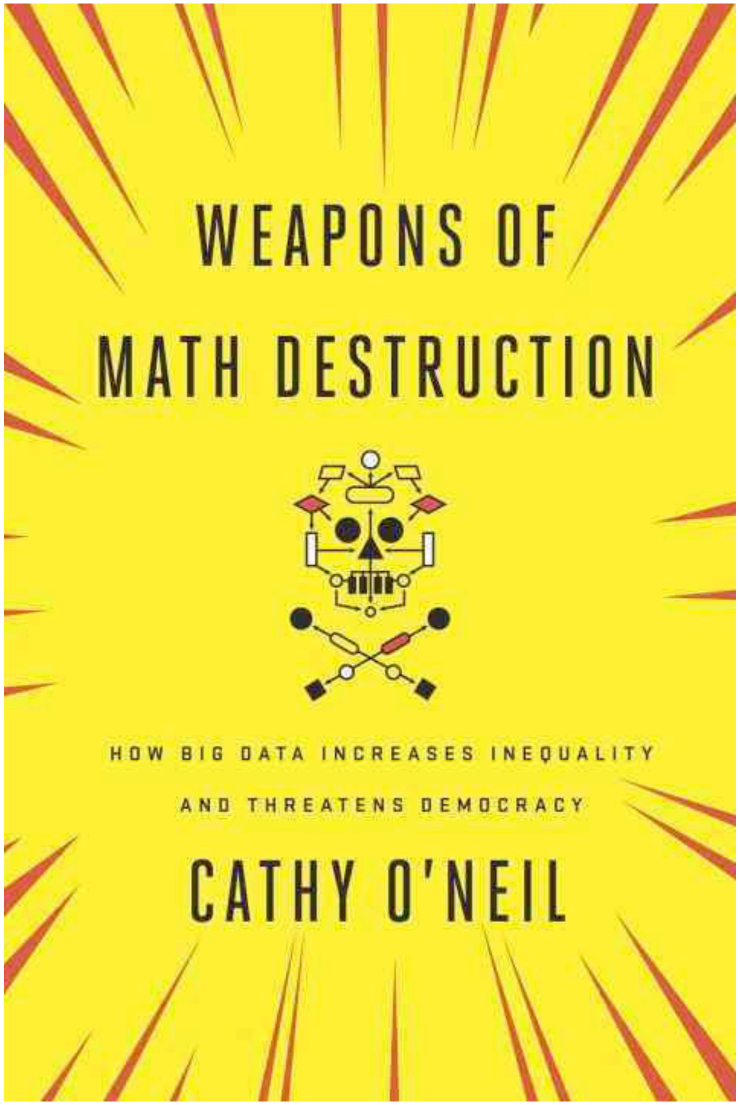
Dating websites have changed the way couples meet. Now evidence is emerging that this change is influencing levels of interracial marriage and even the stability of marriage itself.

by Emerging Technology from the arXiv October 10, 2017

Examples

- **Where you work:** How is the list of applicants for a job sorted? Did you hear about that job? How are your working hours managed? How is your performance rated at the job?
- **Where you live** Can you buy a house? Will you have access to credit? Will you be shown ads for appropriate houses?
- **Government services** Will there be more police surveillance in your neighborhood? If, you are arrested, will you go to jail? Access to healthcare services?
- **How are these decisions made?**
  - What right do we have to understand biases built in by programmers or more likely biases learned from historical data?

Weapons of Math Destruction by Cathy O’Neil



WEAPONS OF MATH DESTRUCTION

HOW BIG DATA INCREASES INEQUALITY AND THREATENS DEMOCRACY

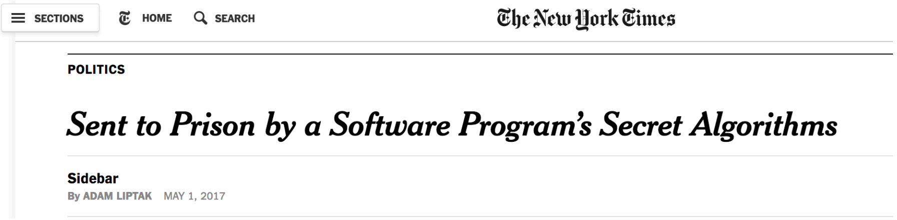
CATHY O'NEIL

What the goals? Who is the customer?

- Algorithm optimized for efficiency/reduced risk for decider
  - Protection for individual's impacted by decisions?
- Platform optimized for advertizing and engagement
  - Protection for society? Democracy?
  - Actors deliberately "gamifying" the system
- We need to actively examine these platforms for the benefit of individuals and society.

Proprietary algorithms

- Proprietary algorithms used in public decisions
- Example in the US of COMPAS software used to assign a numeric score to a person's likelihood of committing another crime
  - Loomis vs. Wisconsin case
  - Protecting intellectual property of the company above the rights of defendants to understand how their score is calculated
- Source code? Training data sets? Constantly evolving rule sets?



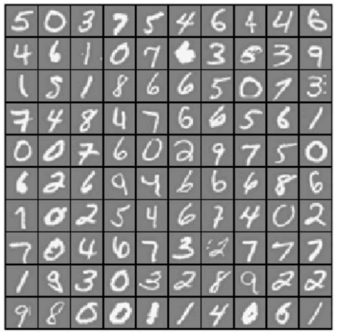
POLITICS

Sent to Prison by a Software Program's Secret Algorithms

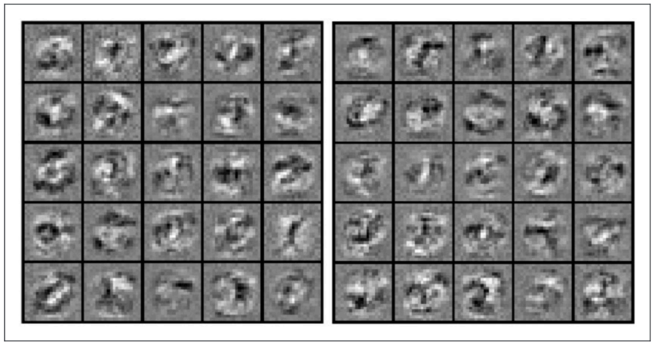
Sidebar

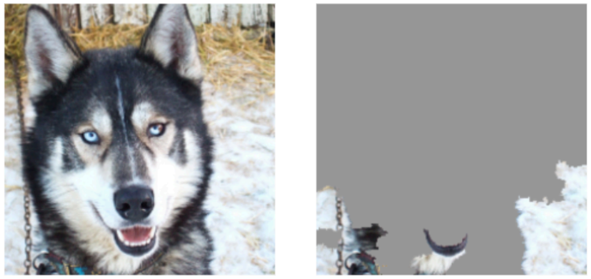
By ADAM LIPTAK MAY 1, 2017

Black Box Decision Making



- Some machine learning algorithms more able to export "explanations"
  - – Decision trees vs neural nets
- Impact of training data
  - Wolves vs Dogs
- Machine learning on personal data? Ability to predict?
  - ❒ Credit card charge for marriage counseling => raise interest rates, lower credit limit
  - ❒ Credit card charge? Facebook like? Colleague on Linked In? Happen to be near a demonstration site?





(a) Husky classified as wolf (b) Explanation

Figures from "How the Machine 'Thinks:' Understanding Opacity in Machine Learning Algorithms", Burrell and ""Why Should I Trust You?": Explaining the Predictions of Any Classifier", Ribeiro et al.

EACH OF US NOW LEAVES A TRAIL OF DIGITAL EXHAUST, AN INFINITE STREAM OF PHONE RECORDS, TEXTS, BROWSER HISTORIES, GPS DATA, AND OTHER INFORMATION, THAT WILL LIVE ON FOREVER.

Instead of "find my iPhone," some auto insurance companies are offering a service that may enable parents to "find my teenager." Progressive Insurance, for example, offers the Snapshot, a tracking device that reports on a car's location, acceleration, braking, and distance traveled. Owners who install the device can get a 10 to 15 percent discount on their policy. Privacy activists, however, fear the technology is ripe for abuse. PHOTO: JACK PARKER

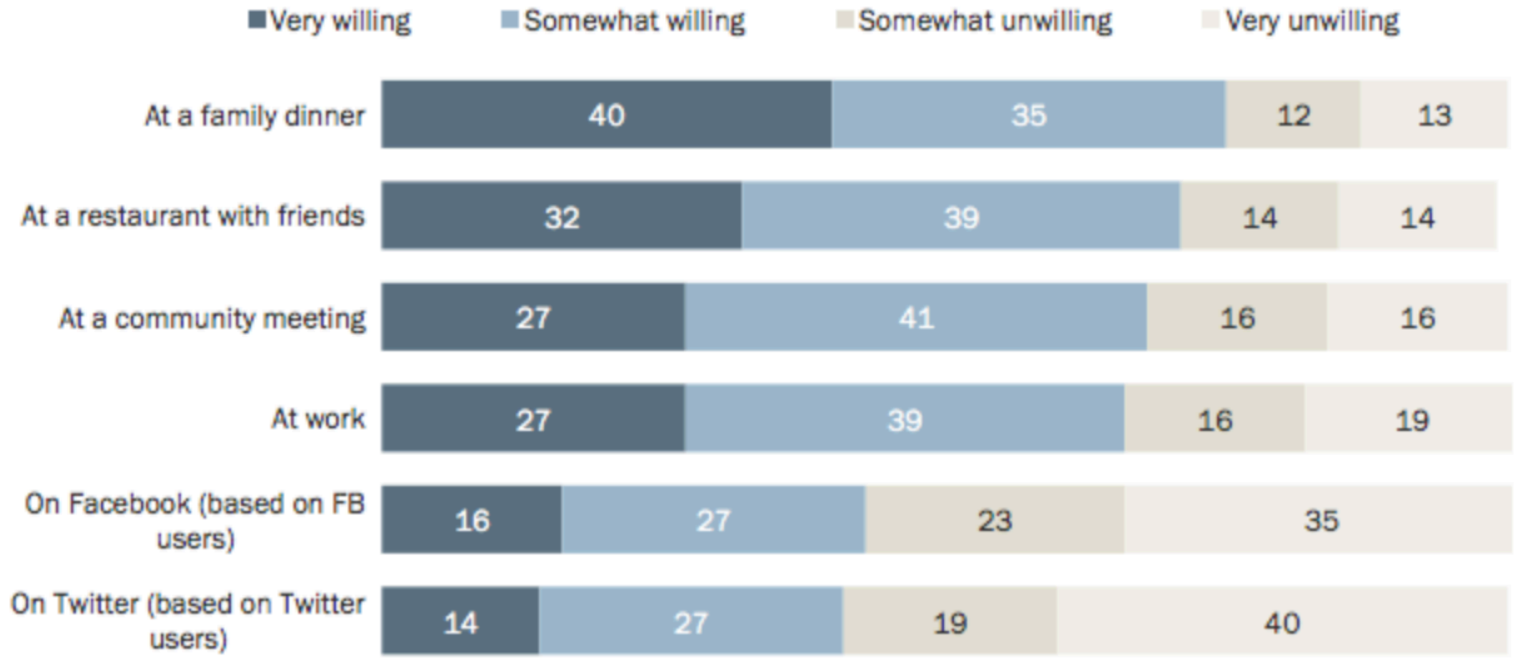
Think about your "digital exhaust"

- Emails, texts
- Social media
- Web browsing history, web site use and cross site correlations
- Cell phone location
- Purchase history, credit cards, wish lists, products viewed/ reviewed, frequent buyer cards,
- Cameras (yours, others, on street, accidental, aware/ unaware, facial recognition) , GPS tags in pictures
- Fitbits, microphones, Google glass,
- License plate readers, passport use, radio-frequency identification (RFID) readers, satellite imagery
- E-readers, streaming video use, MOOCs,

Inability to predict the cost of your action leads to chilling effect on civil discourse

If the topic of the government surveillance programs came up in these settings, how willing would you be to join in the conversation?

% of population

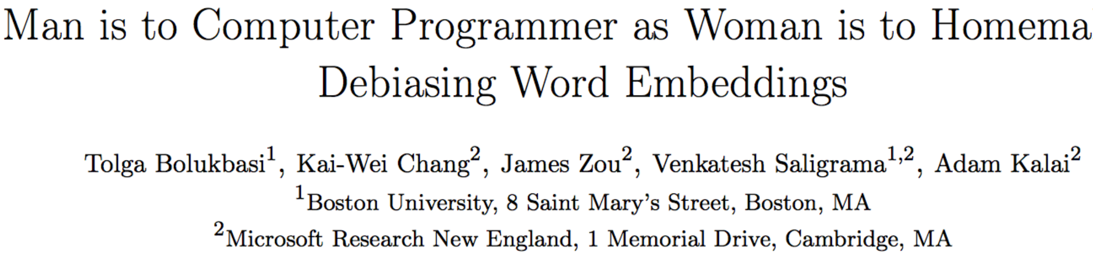


Source, Pew Research Center Internet Project Survey August 7-September 16, 2013. N=1,801 adults.

PEW RESEARCH CENTER

Man is to Computer Programmer as Woman is to Homemaker?

- Word embeddings trained on relatively high quality text like Google News articles (not deliberately biased text) exhibit strong stereotypes
- Word embeddings used in countless applications from web search to sorting resumes for jobs



Man is to Computer Programmer as Woman is to Homemaker?
Debiasing Word Embeddings

Tolga Bolukbasi[1], Kai-Wei Chang[2], James Zou[2], Venkatesh Saligrama[1,2], Adam Kalai[2]
[1]Boston University, 8 Saint Mary's Street, Boston, MA
[2]Microsoft Research New England, 1 Memorial Drive, Cambridge, MA

Unbiased decision made by computer

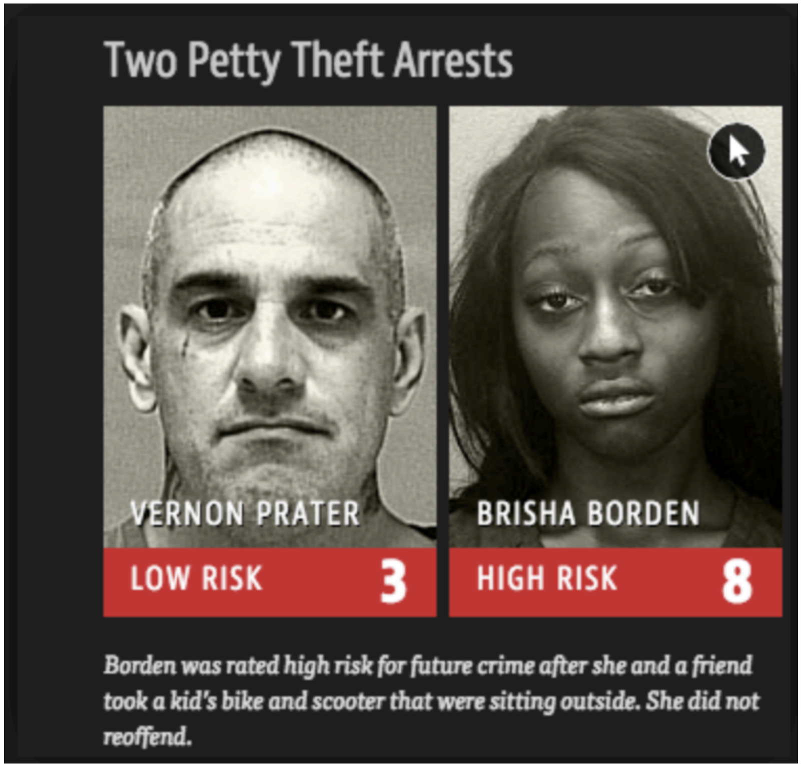
- Attempt to label decisions “unbiased decision made by computer”/logical/pure of human bias
- Can even deliberately ignore attributes/ features like race or gender or religion but so many other proxies for this direct information
    - Shopping habits -> gender?
    - Address -> race?

Proxies

- Some seemingly impartial descriptors, operate as referents for race
  - Zip code, ancestry, disease predisposition, linguistic characteristics, last name, criminal record, and socioeconomic status
- For gender? For religion? For disability status?

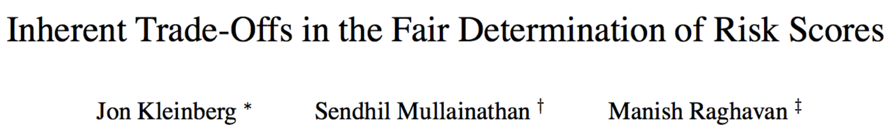
Competing definitions of fairness

- Actually many different definitions of fairness
- Individual fairness vs. group fairness
- Well-calibrated: number/ metric means the same thing regardless of the group
- Balance for the positive class and negative class
- Tradeoffs are unavoidable



Two Petty Theft Arrests

Borden was rated high risk for future crime after she and a friend took a kid's bike and scooter that were sitting outside. She did not reoffend.

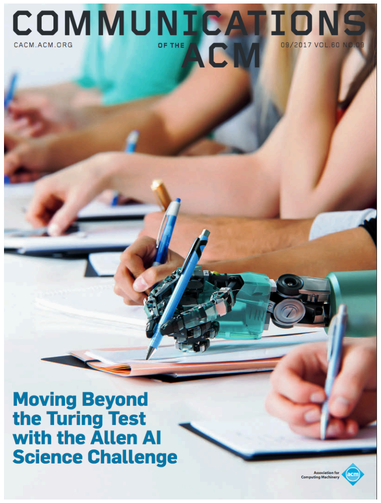


Inherent Trade-Offs in the Fair Determination of Risk Scores

Jon Kleinberg [*] Sendhil Mullainathan [†] Manish Raghavan [‡]

Algorithmic Transparency and Accountability Efforts

- US-ACM/EUACM Statement on Algorithmic Transparency and Accountability
  - 7 principles
  - Awareness, access and redress, accountability, explanation, data provenance, auditability, validation and testing
- CACM article “Toward Algorithmic Transparency and Accountability”


COMMUNICATIONS OF THE ACM
CACM.ACM.ORG
09/2017 VOL.60 NO.09
Moving Beyond the Turing Test with the Allen AI Science Challenge

https://www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_algorithms.pdf

Awareness

Owners, designers, builders, users, and other stakeholders of analytic systems should be aware of the possible biases involved in their design, implementation, and use and the potential harm that biases can cause to individuals and society.

Access and redress

Regulators should encourage the adoption of mechanisms that enable questioning and redress for individuals and groups that are adversely affected by algorithmically informed decisions.

Accountability

Institutions should be held responsible for decisions made by the algorithms that they use, even if it is not feasible to explain in detail how the algorithms produce their results.

Explanation

Systems and institutions that use algorithmic decision-making are encouraged to produce explanations regarding both the procedures followed by the algorithm and the specific decisions that are made. This is particularly important in public policy contexts.

Data Provenance

A description of the way in which the training data was collected should be maintained by the builders of the algorithms, accompanied by an exploration of the potential biases induced by the human or algorithmic data-gathering process. Public scrutiny of the data provides maximum opportunity for corrections. However, concerns over privacy, protecting trade secrets, or revelation of analytics that might allow malicious actors to game the system can justify restricting access to qualified and authorized individuals.

Auditability

Models, algorithms, data, and decisions should be recorded so that they can be audited in cases where harm is suspected.

Validation and Testing

Institutions should use rigorous methods to validate their models and document those methods and results. In particular, they should routinely perform tests to assess and determine whether the model generates discriminatory harm. Institutions are encouraged to make the results of such tests public.

Approaches to Accountability and Transparency

- Source code review
  - Intellectual property
  - Understandable by whom?
- Disclosure of training data
  - Tons and private
- Expert review
  - Who pays for the experts? Ability to enforce compliance
- Disclosure of input data
  - Types and content
- Allow black box testing/multiple queries
  - Scraping Audit/Sock puppet audit
  - Cooperation of platforms?

Other promising research

- New ways to export “explanations”
- Program verification/proofs of fairness for various definitions
- Tracking/Detecting points of policy change in decision making systems
- Quantitative Input Influence (QII) measures that capture the degree of influence of inputs on outputs of systems
- Quantify the costs of different definitions of fairness on accuracy/efficiency
- Debiasing of word embedding

Bottom Line

- Algorithms and platforms are not *only* suggesting and entertaining
- Increasingly used for big decisions about people's lives and fundamentally changing society
- To build the world we want, we need these algorithms and platforms to be accountable and transparent