DSci529:
Security and Privacy In Informatics

Post Mid-term Lecture
AI and Bias With Big Data

Prof. Clifford Neuman

Lecture 7
26 February 2021
Online

Course Outline

- Overview of Security and Privacy
- What data is out there and how is it used
- Technical means of protection
- Identification, Authentication, Audit
- Reasonable expectation of privacy
- Big Data – Technology and Privacy
- **AI and Bias**
- The Internet of Things and Security and Privacy
- Social Networks and the use of our Data
- Access to Data by Governments - Privacy in a Pandemic
- Privacy Regulation - GDPR, CCPA, CPRA
- Influence of Social Media – Free Speech – Disinformation
- CryptoCurrency - TOR - Privacy Preserving Technologies

Upcoming Presentations: Internet of Things – March 5th

- Pratheek Athreya
- Arzu Karaer
- Bolong Pan
- Danielle Sim
- Haipeng Yu
- Jinyu Zhao
- Pu Zhao
- Junbo Sheng
  - This group will have one hour 30 minutes to present.

Upcoming Presentations
Social Media – March 19th

- Addison Allred
- Yixiang Cao
- Lei Gao
- Brianna Hefferin
- Mingliao Xu
- Shengwang Zhang
- Zixin Zheng
- Hehan Xie
- Chengyuan Zhou
- Hehan Xie

- This group will have 100 minutes to present.

Upcomming Presentations Pandemic/Govt Data Use – March 26th

Pandemic (40 minutes)

- Yuemeng Gao
- Tanmay Ghai – Privacy Preserving Contact Tracing
- Yi Lin – Big Data in China related to the COVID Pandemic
- Gan Xin – Health QR Code in China

Other government use of data (50 min)

- Yi Jin – How US and China collect and use personal data
- Congrui Li
- Michelle Muldoon – Law Enforcement and Privacy w.r.t. Data Brokers
- Griffin Weinhold – Decentralized Search and Search Histories in Court
- Xihao Zhou – Use of Data by Governments
- Jinglun Chen – Use of location data
- Jiemin Tang – Security and Privacy regulation for food delivery services

Upcoming Presentations
Privacy&Security Regulation – April 2nd

- Jia Yu Lee
- Yansong Wang
- Kaifan Lu – Assessing China's Cybersecurity Law

- 30 minutes for this group to present

Upcoming Presentations Healthcare – April 2nd

- Vartan Batmazyan
- Phuong Ngo
- Sharad Sharma (DNA Databases)
- Ye Zheng - Fitness apps

- This group will have 40 minutes to present.

Upcoming Presentations – April 9th Free Expression - Disinformation

- Adriana Nana – Deep Fakes and Privacy
- Resherle Verna – Should Social Media company's have right of censorship

  – This group will have 20 minutes to present.

Upcoming Presentations Privacy and Finance – April 16th

- Jonathan De Leon – Privacy in Finance
- Sidong Wang – History and Technologies for Cryptocurrencies
- Saurabh Jain – Privacy of Credit Card/Payment card information
- Yifeng Shi -Financial value of data gathered through free services

- 40 minutes

Secure Communication – Privacy Preserving Technologies – April 16th

- Zihuan Ran – Privacy Preserving Database Technologies
- Aziza Saulebay – 5G and Data Privacy
- Carol Varkey – Messaging Application Privacy
- Francisco Ventura – Encryption Technologies and implications

- 40 minutes

Upcoming Presentations
Other Security Topics – April 23rd

- Yo-Shuan Liu – User experience and Multi-Factor Authentication
- Philana Williams – Security for Web App Development
- Haonan Xu – Privacy issues in Cloud Computing
- Pratishtha Singh – Card privacy Concerns in India

Privacy and Big Data

Required reading:

Big Data and the Future of Privacy
Epic.org

Will Democracy Survive Big Data and Artificial Intelligence?
Scientific American – 25 February 2017

"Muslim registries", Big Data and Human Rights
Amnesty International – 27 February 2017.

Bias in Big Data

- Confirmation bias
  - Relying on data to confirm a certain hypothesis
  - https://www.asia.aonhumancapital.com/home/insights-at-work/confirmation-bias-affect-data-driven-decisions
- Availability heuristic/availability bias
  - Relying on only data that is readily available or recent.
  - Assuming that all instances are equally prevalent.
- Selection bias
  - Sample not representative of the relevant population
  - https://catalogofbias.org/biases/selection-bias/
- Confounding variables
  - Relationship between variables is only true when combined with a third (overlooked) variable.

Examples of big data bias

"Predictive policing" in Chicago

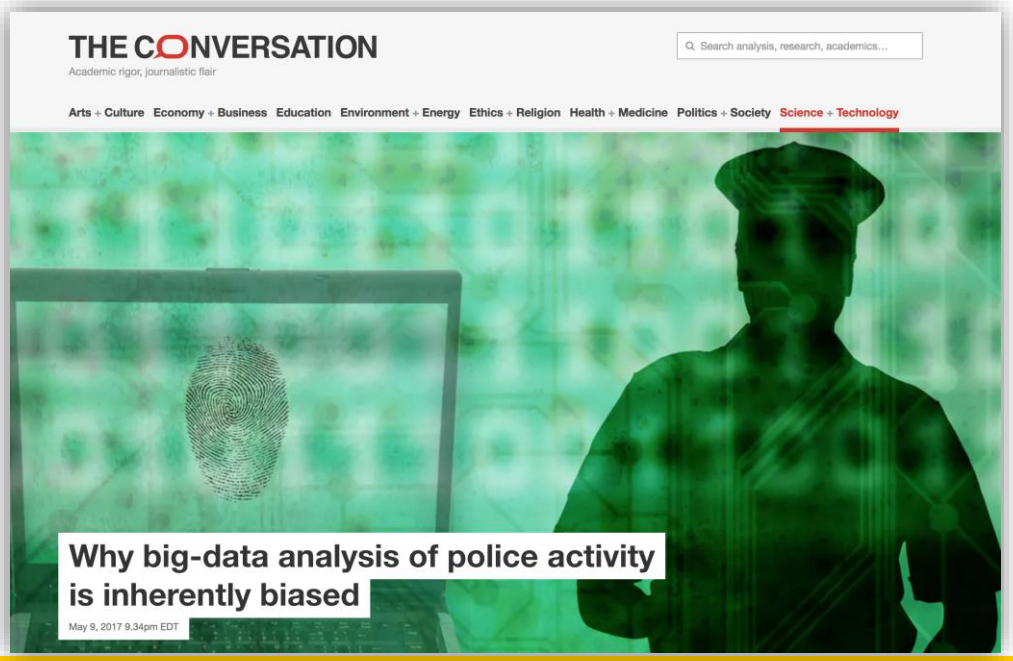
"The Chicago police will use data and computer analysis to identify neighborhoods that are more likely to experience violent crime, assigning additional police patrols in those areas. In addition, **the software will identify individual people who are expected to become – but have yet to be – victims or perpetrators of violent crimes**. Officers may even be assigned to visit those people to warn them against committing a violent crime."

Why big-data analysis of police activity is inherently biased, *The Conversation,* May 9, 2017

Examples of big data bias

Why big-data analysis of police activity is inherently biased, *The Conversation,* May 9, 2017

"Neighborhoods with lots of police calls aren't necessarily the same places the most crime is happening. They are, rather, where the most police attention is – though where that attention focuses can often be biased by gender and racial factors."



Can algorithms illegally discriminate

CNBC – and Whitehouse report

But when it comes to systems that help make such decisions, the methods applied may not always seem fair and just to some, according to a panel of social researchers who study the impact of big data on public and society.

The panel that included a mix of policy researchers, technologists, and journalists, discussed ways in which big data—while enhancing our ability to make evidence-based decisions—does so by inadvertently setting rules and processes that may be inherently biased and discriminatory.

The rules, in this case, are algorithms, a set of mathematical procedures coded to achieve a particular goal. Critics argue these algorithms may perpetuate biases and reinforce built-in assumptions.

Also

http://www.nextgov.com/big-data/2017/02/cfpb-wants-know-how-alternative-data-changes-credit-scores/135695/

Critics allege big data can be discriminatory, but is it really bias?

Pradip Sigdyal | @PSigdyal
Sunday, 8 May 2016 | 4:00 PM ET
CNBC

Getty I 187131740

Big data is increasingly viewed as a strategic asset that can transform organizations through its use of powerful predictive technologies.

But when it comes to systems that help make such decisions, the methods applied may not always seem fair and just to some, according to a panel of social researchers who study the impact of big data on public and society.

Can algorithms illegally discriminate

- **Hiring**
  - https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G
- **Housing**
  - https://www.technologyreview.com/2020/10/20/1009452/ai-has-exacerbated-racial-bias-in-housing-could-it-help-eliminate-it-instead/
- **Advertising (in the above areas)**
  - https://hbr.org/2019/11/how-targeted-ads-and-dynamic-pricing-can-perpetuate-bias
- **Criminal Justice**
  - https://www.technologyreview.com/2019/01/21/137783/algorithms-criminal-justice-ai/
- **Credit**
  - https://www.wired.com/story/the-apple-card-didnt-see-genderand-thats-the-problem/
- **Exam proctoring**
  - https://www.technologyreview.com/2020/08/07/1006132/software-algorithms-proctoring-online-tests-ai-ethics/

Current Event Discussion

- http://csclass.info/USC/INF529/s21-lec7-ce.html