Data Ethics: Understanding Big Data, Algorithmic Bias, and Research Ethics

Cara Marta Messina and Jeff Sternberg
Philosophy: PHIL 1102 & 1112
Katy Shorey
Fall 2019

Northeastern University
NULab for Texts, Maps, and Networks

Feel free to ask questions at any point during the presentation!

Discussion: China's Social Credit System

- What is China's Social Credit system? How does it work?
- How does China define prosocial behaviors through these technologies? Why would China want to do this?
- In what ways might America have similar or different technological infrastructures?

Northeastern University
NULab for Texts, Maps, and Networks

Feel free to ask questions at any point during the presentation!

Workshop Agenda

- Introduce "Big Data"
- Discuss data, privacy, and data categorization
- Activity: Adopt or Not?
- Discuss ethical implications of big data and more generally digital research

Slides, handouts, and data available at
http://bit.ly/diti-fall2019-shorey

Feel free to ask questions at any point during the presentation!

Workshop Goals

- Understand the ways that data is being used in society as well as how algorithms impact and shape our daily lives
- Understand the ways in which technology reflects cultural, social, and political biases
- Explore the ways in which privacy and security are being reshaped and redefined through big data, algorithms, and policy
- Explore the ways in which these questions and methods are influencing how humanists and social scientists do research

Feel free to ask questions at any point during the presentation!

Big Data and Surveillance

Feel free to ask questions at any point during the presentation!

What is 'Big Data'?

Big data has been called the 'new oil' by some.

Shoshana Zuboff argues that we now live in an era of 'surveillance capitalism.'

The four components of big data are: **volume**, **variety**, **velocity** and **veracity**

Northeastern University
NULab for Texts, Maps, and Networks

Feel free to ask questions at any point during the presentation!

Why should we care?

- Big data is characterized by its **scale**
- Big data **sources** include: digitized records, social media/internet activity, and sensors from the physical environment.
- Big data is often **privately owned**
  - Example: an insurance company purchasing social media activity from Facebook in order to make insurance sales decisions.

Feel free to ask questions at any point during the presentation!

Social Media Preferences

Social media sites collect, store, and sell information about you so you get better targeted ads and your newsfeed is tailored to your categories. **What are the targeted ads you see and why do you think you receive these ads?**

Some social media sites that do this:

- Facebook
- Instagram (owned by Facebook)
- Google
- YouTube (owned by Google)
- Twitter

Feel free to ask questions at any point during the presentation!

Google's File on You is 10 Times Bigger Than Facebook's — Here's How to View It

Google, Amazon, Apple, and Microsoft are all central players in "surveillance capitalism" and prey on our data.

THE ANTI-MEDIA

Example: If you have **location services** turned on for Google (if you use Google maps), Google can track your every move. Go to:

https://www.google.com/maps/timeline

Northeastern University
NULab for Texts, Maps, and Networks

Feel free to ask questions at any point during the presentation!

Ethical Implications

- Cambridge Analytica controversy
- Big data also raises questions of autonomy, anonymity, privacy, discrimination, and bias.
- Disparate impact
- Questions to consider:
  - How are we being represented online?
  - How is our data being used?
  - Who is using it and for what purposes?
  - How might it be used in the future?

Feel free to ask questions at any point during the presentation!

DIY Cybersecurity and Tightening your Privacy

Want to make your life more private? Follow this "DIY Guide to Feminist Cybersecurity"

https://hackblossom.org/cybersecurity/

Northeastern University
NULab for Texts, Maps, and Networks

Feel free to ask questions at any point during the presentation!

Algorithms and Bias

Northeastern University
NULab for Texts, Maps, and Networks

Feel free to ask questions at any point during the presentation!

Activity: Adopt or Not?

Small Group: Find a partner or two! You all work for an adoption agency and have to decide if someone can adopt a dog. On your handouts, please read the four previous adoption applications and decide if the new adoption applicant can adopt or not.

Do you think this new applicant should be permitted to adopt a dog? Why or why not?

Feel free to ask questions at any point during the presentation!

Class Discussion: Adopt or Not?

- Did you ACCEPT or REJECT their application? Why?
- What questions from the application did you weigh more? Why?
- What might be some implicit biases in this application and in your choices?

Northeastern University
NULab for Texts, Maps, and Networks

Feel free to ask questions at any point during the presentation!

Adopt or Not? Algorithm

Algorithms to "read" through applications exist.

- Where might you see these algorithms being used to make decisions?
- What biases may be ingrained in those applications?
- How might these results impact choices?

Feel free to ask questions at any point during the presentation!

Is it all gloom and doom?

Are the ethics around algorithms and big data all negative or controversial?

Feel free to ask questions at any point during the presentation!

Initiatives for Justice

Code for America is an organization that works with the mass amount of undigitized, unorganized government documents to help previously incarcerated people.

One project they have, titled "Clear My Record," attempts to parse through the mass data of governmental records to help clear criminal records, particularly for people who were arrested for marijuana use/distribution.

https://www.codeforamerica.org/programs/clear-my-record

Northeastern University
NULab for Texts, Maps, and Networks

Want to learn more about accountability and best practices when creating algorithms?

Visit https://www.fatml.org/, or Fairness, Accountability, and Transparency in Machine Learning

Feel free to ask questions at any point during the presentation!

Thank you!

If you have any questions, contact us at:

Cara Marta Messina
Digital Teaching Integration
Assistant Director
messina.c@husky.neu.edu

Jeff Sternberg
Digital Teaching Integration
Research Fellow
sternberg.je@husky.neu.edu

Slides, handouts, and data available at **http://bit.ly/diti-fall2019-shorey**

DTI office hours: **Tuesdays, 1–3PM in 401 Nightingale Hall**

Northeastern University
NULab for Texts, Maps, and Networks

Feel free to ask questions at any point during the presentation!