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

Milan Skobic and Talia Brenner

CRIM 3600 Research Methods

Stacie St. Louis

Fall 2020, September 28, 10:30-11:35am

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?
- In what ways might America have similar or different technological infrastructures?

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

Workshop Agenda

- Objectives
- Introduce 'Big Data' Concepts
- Discuss data, privacy, and algorithms
- Activity: Can you adopt this dog?
- Discuss ethical implications of big data and lessons for (digital) research

Slides, handouts, and data available at
http://bit.ly/diti_fall2020-st_louis1

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

Workshop Objectives

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

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

What is "Big Data"?

Companies, governments, and other groups collect vast amounts of data ("big data") from vast numbers of users and analyze these data quickly for particular purposes (advertising, surveillance, search results, etc).

The goal of collecting and processing these data is to predict individual user behavior based on patterns from the user as well as patterns from "similar" users (based on demographic information, behavioral patterns, etc).

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, or 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

Northeastern University
NULab for Texts, Maps, and Networks

Questions to consider

- How are we being represented online?

- How are our data being used?

- Who is using our data and for what purposes?

- How might our data be used in the future?

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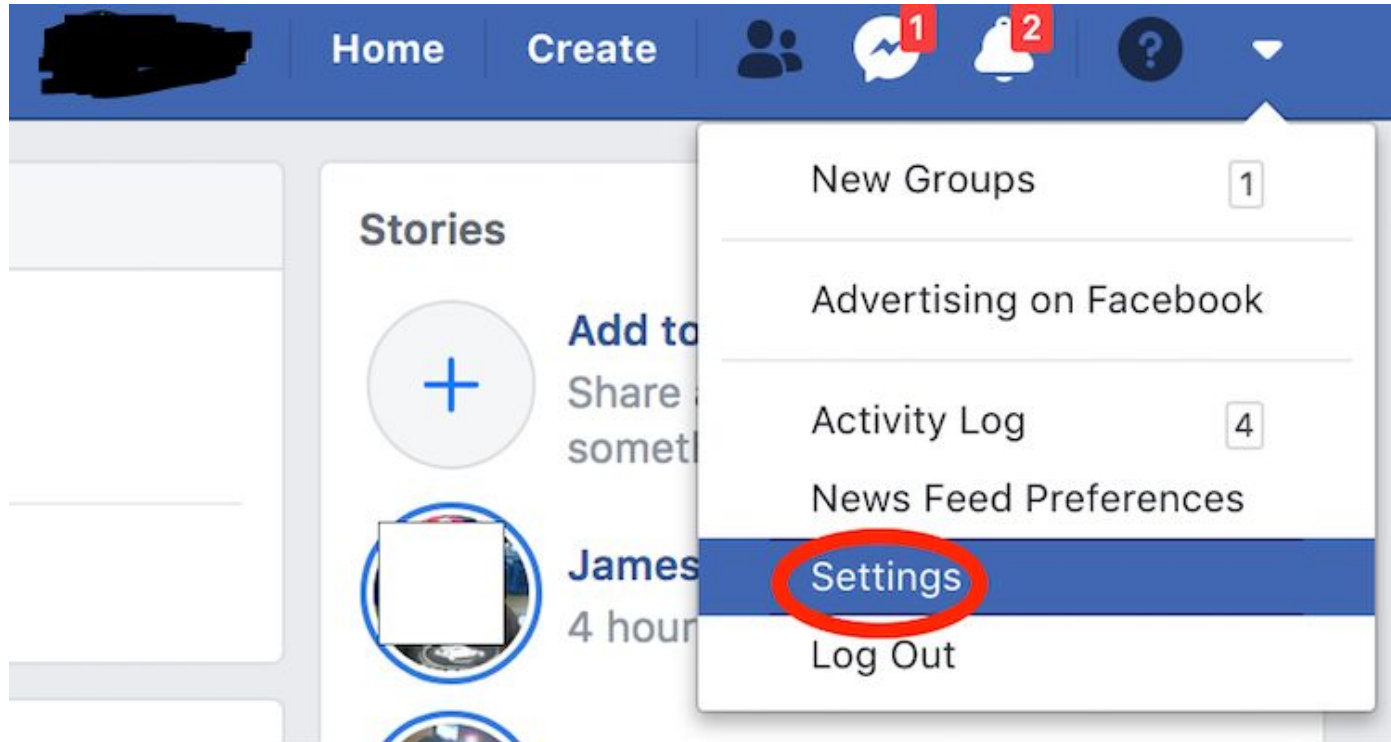
Facebook Preferences

Facebook collects, stores, and sells information about you so you get more targeted ads and your News Feed is tailored to your categories.
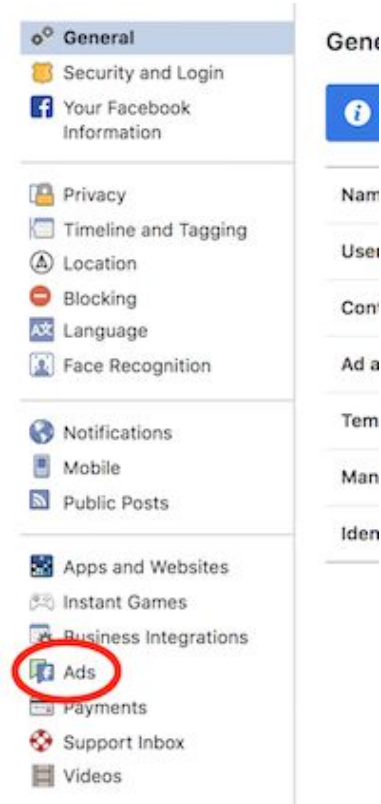
Other social media sites that do this:

- Instagram (owned by Facebook)
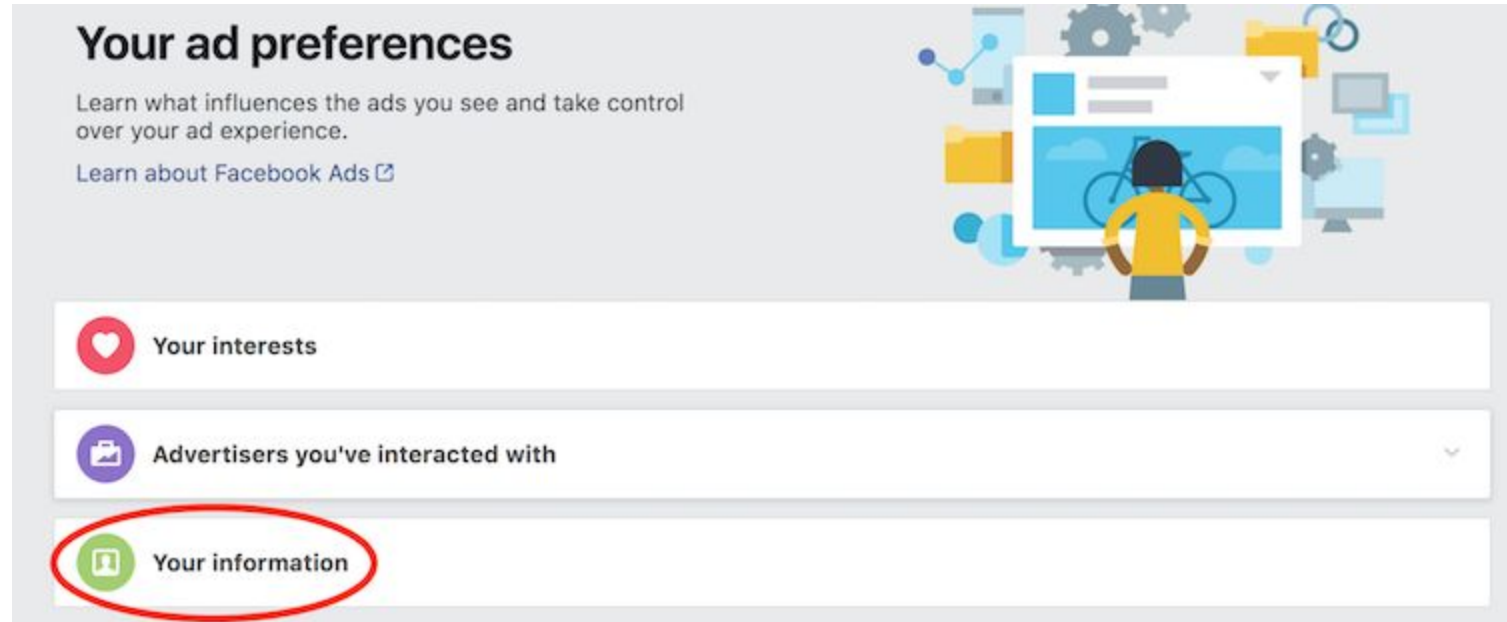- Google
- YouTube (owned by Google)
- Twitter
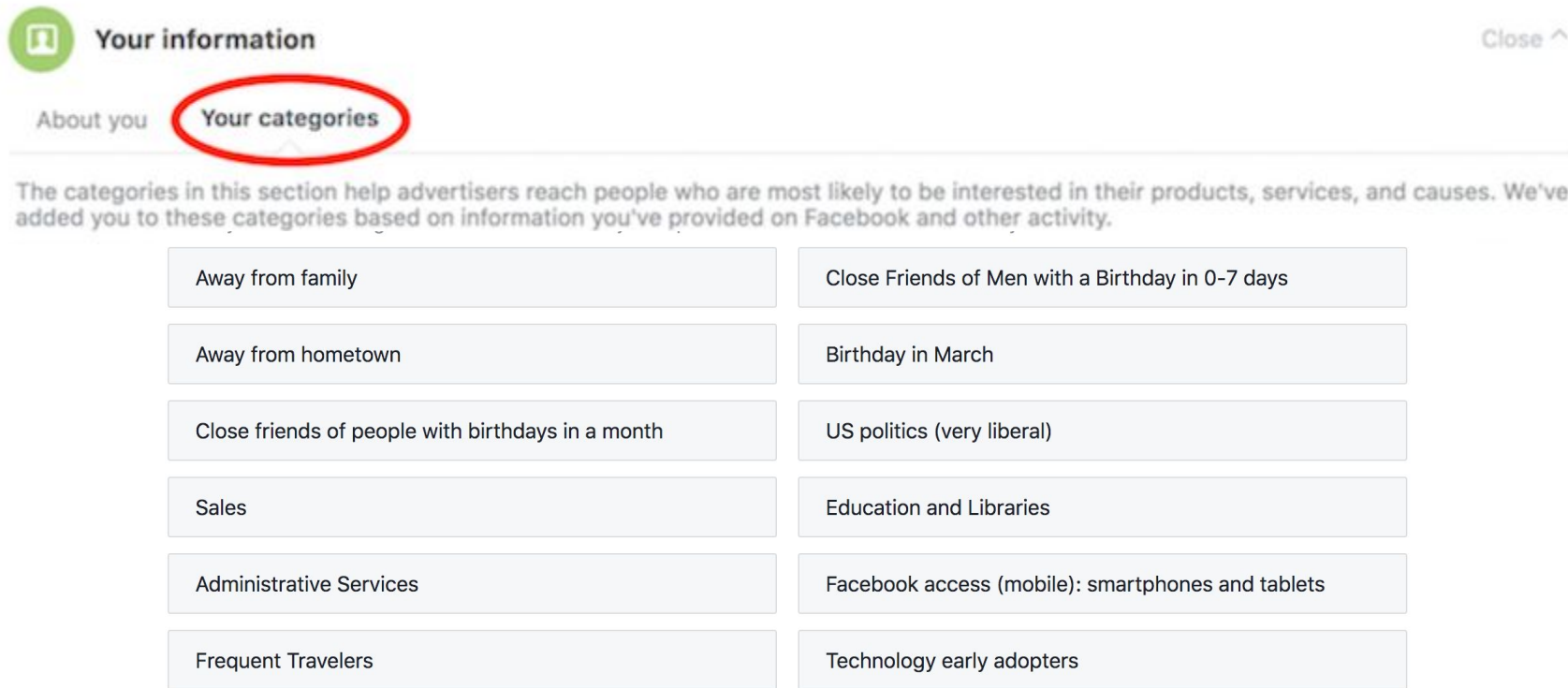
Northeastern University
NULab for Texts, Maps, and Networks

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

Settings > Ads > Your information > Categories

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

Settings > Ads > Your information > Categories

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

Settings > Ads > Your information > Categories

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

Settings > Ads > Your information > Categories

Your information — Close

About you **Your categories**

The categories in this section help advertisers reach people who are most likely to be interested in their products, services, and causes. We've added you to these categories based on information you've provided on Facebook and other activity.

Away from family	Close Friends of Men with a Birthday in 0-7 days
Away from hometown	Birthday in March
Close friends of people with birthdays in a month	US politics (very liberal)
Sales	Education and Libraries
Administrative Services	Facebook access (mobile): smartphones and tablets
Frequent Travelers	Technology early adopters

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 (like 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!

Downloading Your Data

Facebook: Settings > Your Facebook Information > Download your Information

Google:

https://support.google.com/accounts/answer/3024190?hl=en

Instagram app: Settings > Privacy and Security > Data download/Request Download

DIY Cybersecurity and Tightening your Privacy

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

https://hackblossom.org/cybersecurity/

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

"Big Data" Unbounded - Ethical Issues

- Controversies in the recent years:
  - Cambridge Analytica 2016 elections [controversy](#)
  - [Clearview AI](#): facial recognition "services" in 2020
  - General [use of facial recognition in policing](#) in recent years
  - Influence of algorithms in [racially differential health outcomes](#)
  - Use of algorithms in [grading in the UK](#) in 2020
  - And many, many more all across the world...
- "Big data" raises questions of power, autonomy, anonymity, privacy, discrimination, and bias.

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

Facial recognition in policing and beyond

- Case of Robert Williams, <u>wrongfully arrested</u> in 2019
- Most algorithms <u>have been found</u> to contain gender and racial bias when tested on accuracy of face recognition
- These issues were present <u>from the onset</u> of implementation of these technologies, but they still go introduced
- These biases are <u>reproduced</u> both in the programming of algorithms, and in collection of datasets from which algorithms are trained

Northeastern University
NULab for Texts, Maps, and Networks

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

# Algorithms and Bias - Activity

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

Algorithms

Do you rely on algorithms in your everyday life? Any examples?

An algorithm is a process of instructions provided, usually for computers to interpret and follow. There is usually an **input**, which is determined by the programmer; then there is a set of rules (the algorithm) that help lead to the **output**, or the results of the program following instructions.

Algorithms can be fairly simple, but they can also be much more complex.

What is Algorithm?

Input → Set of rules to obtain the expected output from the given input → Output

Algorithm

GG

Northeastern University
NULab for Texts, Maps, and Networks

Feel free to ask questions at any point during the presentation!
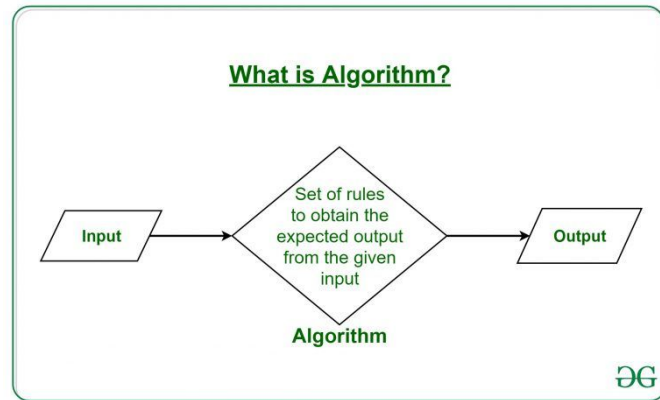
Activity: Can you adopt this dog?

Small Group: You will be assigned into small groups. You work for an animal shelter 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 allowed to adopt a dog? Why or why not?

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

Class Discussion: Can you adopt this dog?

- Would 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 form, the process, and in your choices?

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

Can you adopt this dog? Algorithm

Algorithms "read" through data such as these applications, and often help us make decisions. Here are some questions to think about when assessing algorithms:

- Where might you see these algorithms being used to make decisions? Why are they being used? What are they replacing or adding on to?
- What biases may be ingrained in the data collected for the algorithms? What biases may be ingrained in the actual process of using the algorithm?
- In what ways might the algorithm prevent or reinscribe human bias?

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

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

Northeastern University
NULab for Texts, Maps, and Networks

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

So what do 'big data' & algorithms have to do with research?

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

Questions Researchers Must Ask

- What **information** is being collected and from where? To whom do these data **belong**?
- How is the information being **collected**? Do **participants** know that it is collected, how it will be collected, and how it will be used?
- **How** will the data be analyzed? What **biases** and **ideologies** may be implicit in this analysis?
- Who will this research impact? Who will it **benefit**? Who will it potentially **harm**?

Northeastern University
NULab for Texts, Maps, and Networks

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

Discussion

- What are some benefits and what are some risks coming with the increased focus on "big data" in research and policy?
- Do you see any commonalities among the risks and how would you address them?
- Finally, "big data" is a vague term and refers to many different phenomena at once. What are some other terms you would use in order to think about these issues more precisely?

Northeastern University
NULab for Texts, Maps, and Networks

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

Thank you!

If you have any questions, contact DITI at nulab.info@gmail.com

Developed by Cara Messina and Garrett Morrow
Digital Integration Teaching Initiative
DITI Research Fellows

Taught by Milan Skobic
DITI Assistant Director
and Talia Brenner
NULab Research Assistant

Slides, handouts, and data available at **http://bit.ly/diti_fall2020-st_louis1**

Schedule an appointment with us! **https://calendly.com/diti-nu**

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