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

Cara Marta Messina and Garrett Morrow
SOCL 2321 Research Methods
Ineke Marshall
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?
- 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 data categorization
- Discuss ethical implications of big data and more generally digital research
- Activity: Adopt or Not?

Slides, handouts, and data available at **https://bit.ly/2H48Yc1**

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

Workshop Goals

- Understand the ways 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 the use of big data, algorithms, and policy
- Explore the ways in which these questions and methods are influencing how humanists and social scientists do research and practice their craft

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**

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 specific insurance sales decisions.

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

Facebook Preferences

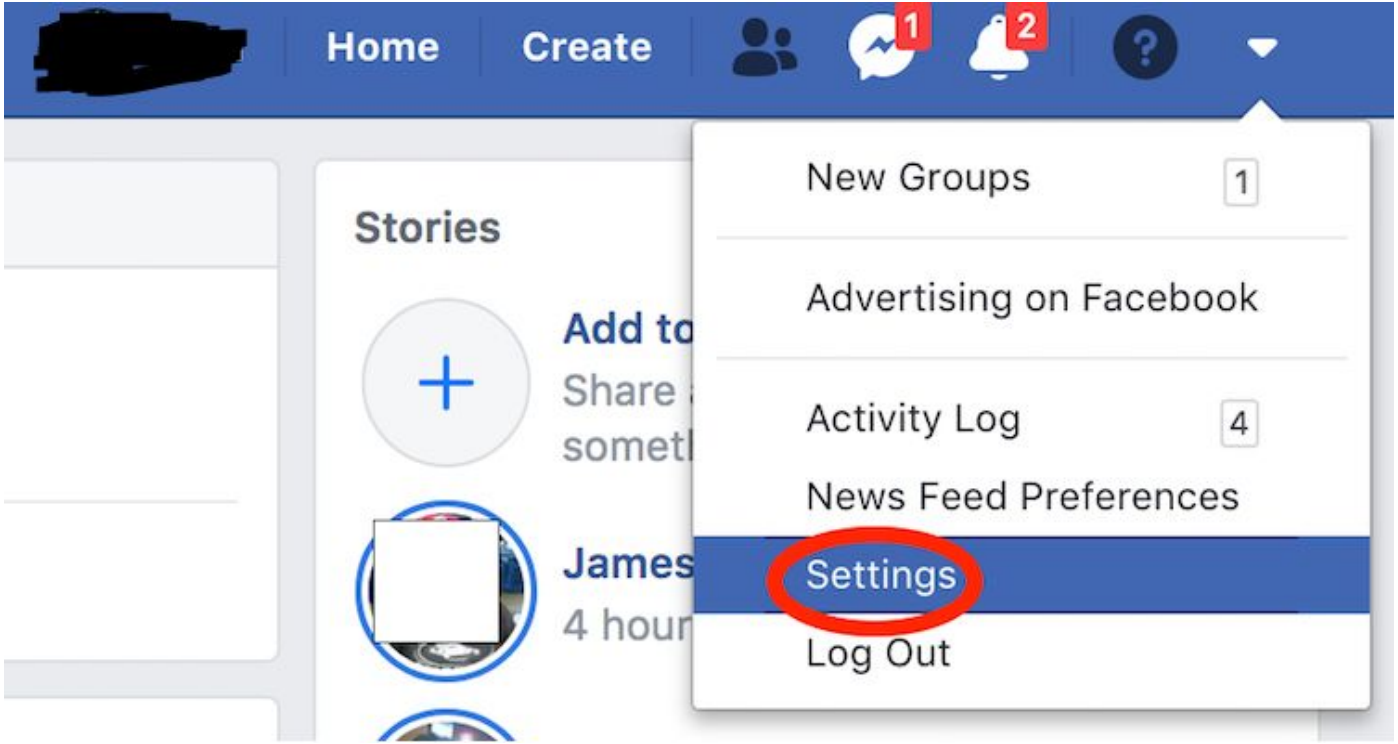
Facebook collects, stores, and sells information about you so you get better targeted ads and your newsfeed is tailored to your categories.

Other social media sites that do this:

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

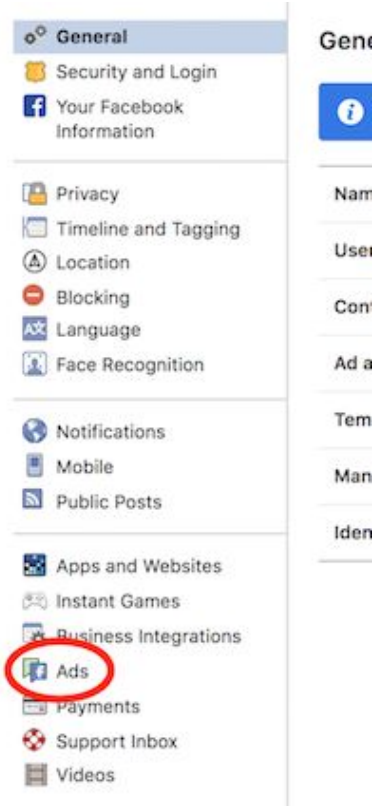
Feel free to ask questions at any point during the presentation!

Settings > Ads > Your information > Categories



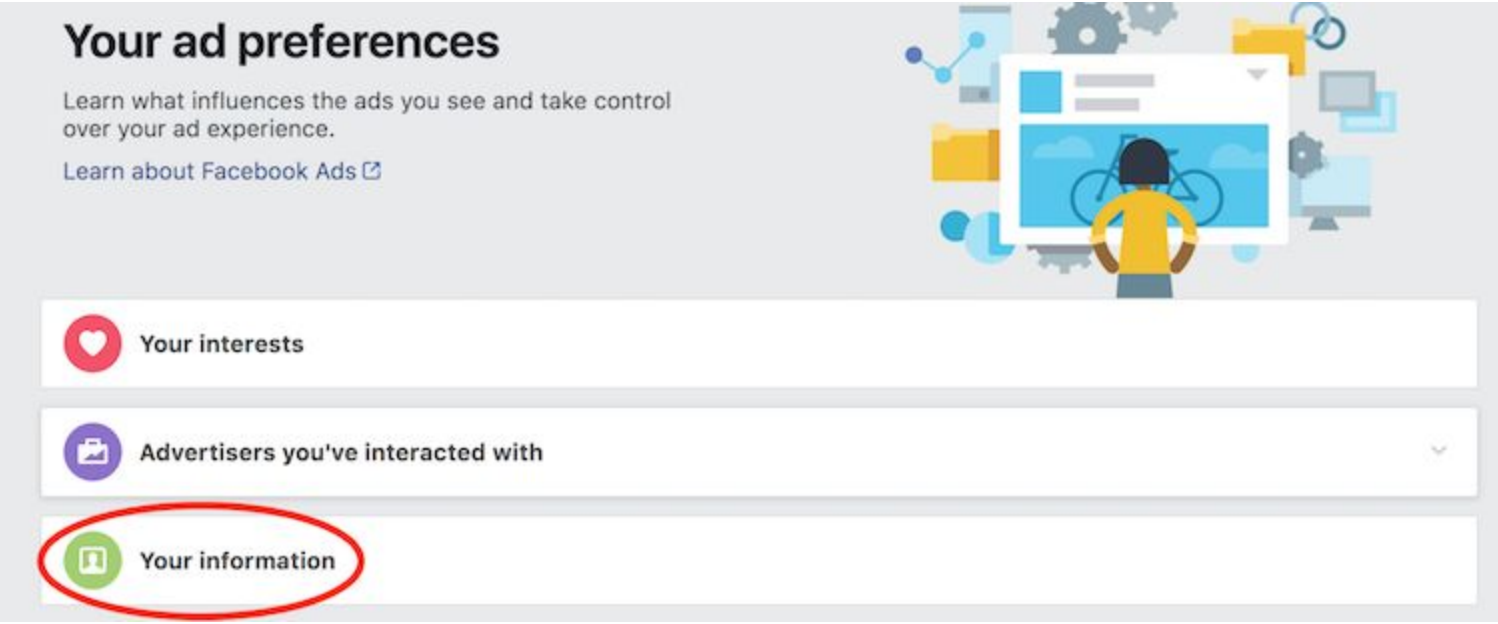
Feel free to ask questions at any point during the presentation!

Settings > Ads > Your information > Categories



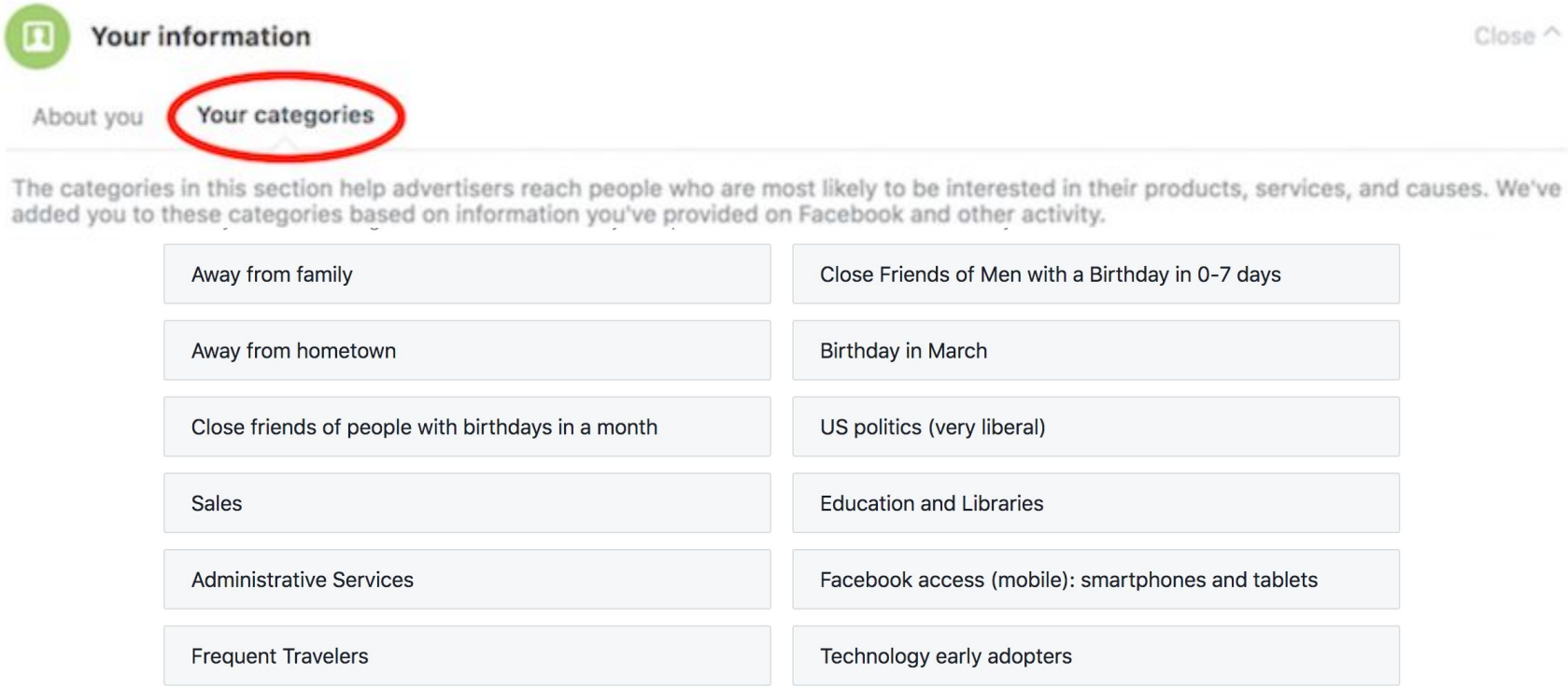
Feel free to ask questions at any point during the presentation!

Settings > Ads > Your information > Categories



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

Settings > Ads > Your information > Categories



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

AWARENESS | SCIENCE & TECH | AUG 3, 2019 AT 11:08 AM.

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

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: Settings > Security > Access Data/Download Data

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.
- 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/

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 can adopt a dog? Why or why not?

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

Class Discussion: Adopt or Not?

- Do 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?

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

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

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

Northeastern University
NULab for Texts, Maps, and Networks

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 does this data **belong**?
- How is it being **collected**? Do **participants** know that it is collected, how it will be collected, and how will it 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**?

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

Garrett Morrow
Digital Teaching Integration
Research Fellow
morrow.g@husky.neu.edu

Slides, handouts, and data available at **https://bit.ly/2H48Yc1**

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

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