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

Milan Skobic and Vaishali Kushwaha
Digital Integration Teaching Initiative (DITI)

ENGW 1111 Academic Writing
Maryam Monalisa Gharavi
Spring 2021

Northeastern University
NULab for Texts, Maps, and Networks

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

Discussion: Lyft and justice

- What made it so difficult for Lyft to cooperate in solving the case of Alison Turkos and preventing it from happening again?
- What made it so difficult for police and other state institutions to solve this case?
- Why was it so difficult for Alison Turkos to find and get adequate company and institutional response?
- What are some other ethical issues that you see in this case?

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

Ethical issues

- (Lack of) Accountability
- (Lack of) Transparency
- Intellectual property vs. oversight
- Imbalances of power
- Sanctity of contract and profit imperative in capitalism
- Appearance of neutrality and objectivity

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

Workshop Agenda

- Goals of the class visit
- Introduce ‘Big Data’ Concepts - as a reminder
- Discuss data, algorithms, and power, and the cases that help us think about these issues
- 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-spring2021-gharavi

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

Workshop Objectives

- Understand the ways data are being created and used in society as well as how they and shape our daily lives
- Explore the ways in which privacy, security and other aspects of social life are being reshaped and redefined through the use of big data, algorithms, and policy
- Understand the ways in which technology contributes to unequal power relations and reflect various societal 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).

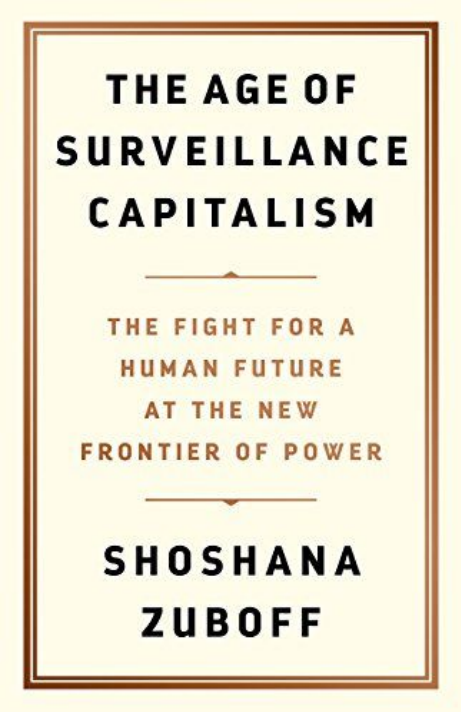
The goal of collecting and processing these data is to predict individual user behavior based on patterns from the user as well as patterns from “similar” users (based on demographic information, behavioral patterns, etc).

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

Surveillance Capitalism

Shoshana Zuboff defines “surveillance capitalism” as the commodification of human behavior. Our “data”--our demographic information, our everyday behaviors online and in person, and who we know--is collected and sold for analysis and advertising purposes. This is one of the main goals of big data.

Our information is one of the most valuable products in America and other countries with similar economic structures.



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

Why should we care?

- The **scale** of big data enables those who use, develop and control it to magnify their influence
- Big data is **omnipresent** - it's **sources** include: digitized records, internet activity, or sensors from the physical environment
- Big data is often **privately owned** and it is hard to ensure oversight over how it is developed, used and controlled

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

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!

Terms of service - what is their role?

- Why are we asked to consent to terms of service?
- Who writes and decides what goes into terms of service?
- How easy can each party enforce the application of these terms?

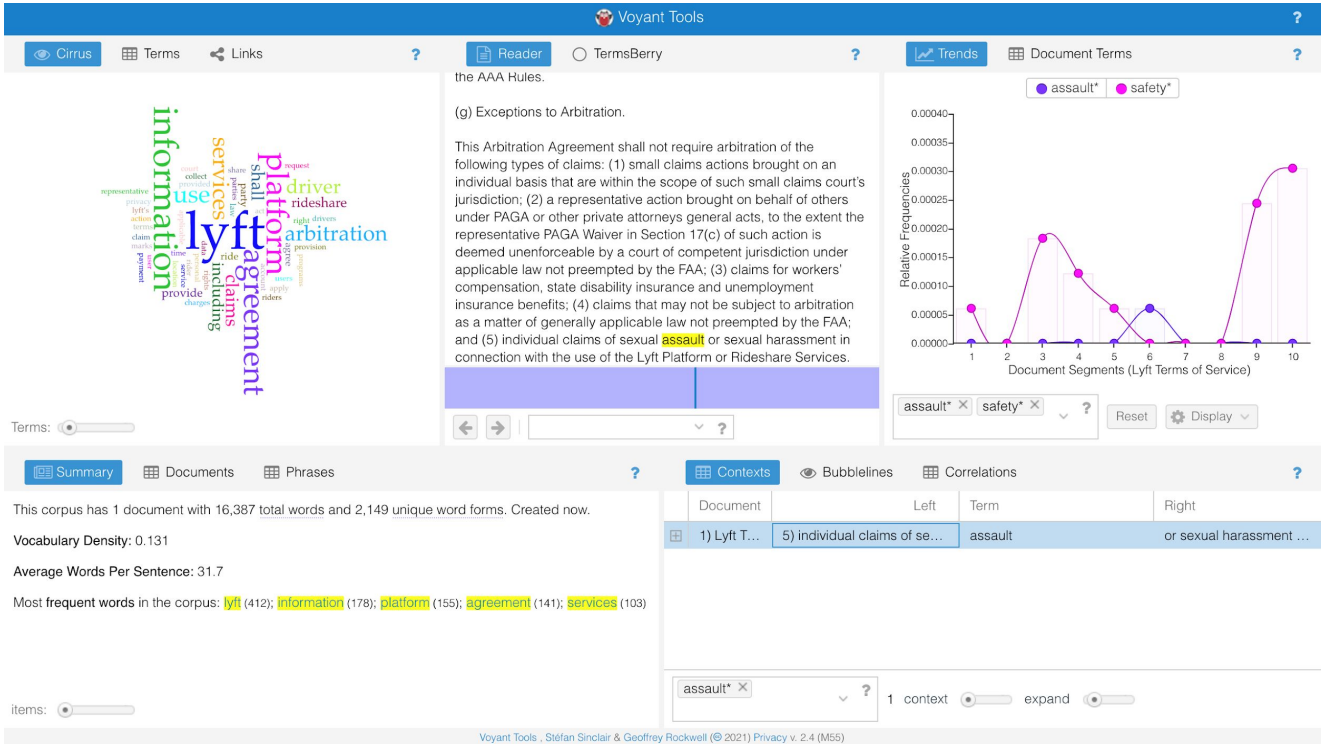
Feel free to ask questions at any point during the presentation!

Lyft's terms of service

- How would the approach to data ethics outlined by Dr. Brandeis Marshall be applied to issues with Lyft's Terms of Services
- How does this Terms of Service fare when brought into connection with what happened to Alison Turkos?
- What is accounted for in these terms of service, and what might be missing?

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

Terms of Service and Text Analysis



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

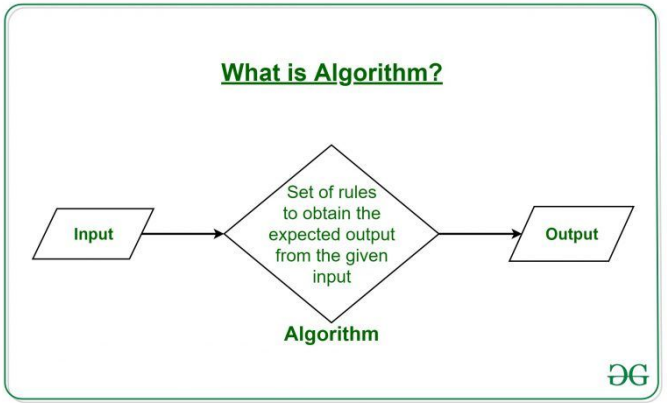
Feel free to ask questions at any point during the presentation!

Algorithms

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.



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

Machine learning

- Algorithms geared toward recognizing patterns in data and improving their recognition with minimal human input
- Considered a branch of artificial intelligence
- Machines are trained on large datasets - previously compiled by humans and also subject to various ethical issues previously mentioned

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...

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

Facial recognition in policing and beyond

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

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

Algorithms and Bias - Activity

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

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!

General questions when thinking about data

- 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**?

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?
- How come that the abundance of data in the case of Alison Turkos prevented rather than fostered finding the solution?
- What are the obstacles toward ensuring the ethical development and use of data?
- How could those obstacles be removed? More regulation, more bans, or something else?

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 and taught by Vaishali Kushwaha, Julianna Wessels and Milan Skobics
Digital Integration Teaching Initiative
NULab Research and teaching Fellows

Slides, handouts, and data available at **http://bit.ly/diti-spring2021-gharavi**

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

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