Digital Ethics: Understanding Big Data and Algorithmic Bias

Cara Marta Messina and Garrett Morrow
Advanced Writing in the Disciplines
Cecelia Musselman
Spring 2020

Northeastern University
NULab for Texts, Maps, and Networks

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

Class Discussion

Open Google's search engine and type in "Greatest authors of all time."

- What are some of the results? What do you notice about these results?
- How many authors on this list have you read? Do you agree with the list?
- Where do you think these results came from?
- What do these results suggest to you in terms of defining "greatest" and "authors"?

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

Class Discussion

Now try these results:

- Greatest women authors
- Greatest Black women authors
- Greatest Black authors
- Greatest white authors

“Black” leads to substantial results, while “white” does not. Why do you think this might be?

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

Technology is Not Neutral

Do white men really make better authors than white women, Black women, and Black men? According to the literary canon and Google, yes, they do! But should we just accept this?

Information systems like Google as well as data collection, data analysis, and algorithms are **not neutral**. They can reinforce and make explicit systemic, political, and cultural biases. They are affected by input data, the way that data is presented, how the data is interpreted by machines, and more. This means we also have the ability to challenge these biases, norms, and forms of discrimination.

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

Workshop Agenda

- Introduce ‘Big Data’ Concepts
- Discuss data, privacy, and data categorization
- Activity: Adopt or Not?
- Introduce your assignment

Slides, handouts, and data available at

http://bit.ly/diti-spring2020-musselman1

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

Workshop Objectives

- Understand the ways data is collected and analyzed 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.

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, including Andrew Yang.

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 **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.
- Big data can often reproduce results that may harm certain communities.

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

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

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/

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 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: Adopt or Not?

- 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!

Adopt or Not? Algorithm

Algorithms can "read" through data such as these applications, and 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 algorithms?
- In what ways might the algorithms prevent or reinscribe human biases?

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 digital ethics and algorithmic biases have to do with professionalization?

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

Important Questions

These are questions *all* professionals working in tech & tech-related industries must be thinking about:

- What **information** is being collected and from where? To whom does this data **belong**? **Who** is doing the collecting?
- 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** is analyzing the data?
- 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 Integration Teaching Initiative
Assistant Director
messina.c@husky.neu.edu

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

Slides, handouts, and data available at

http://bit.ly/diti-spring2020-musselman1

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

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