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

Sociology of the Family
Linda Blum
Summer II 2021

Northeastern University
NULab for Texts, Maps, and Networks

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

“We shape our tools, and thereafter our tools shape us.”

- Marshall McLuhan, the 20th century media theorist.

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

What do we mean by 'Data Ethics' and why are we talking about it?

What we're doing today:

- Exploring a few major **concepts** related to big data and data ethics
- Connecting the course **readings** to these larger ideas
- Having **discussions** that allow us to engage with ethical and practical questions pertaining to big data, algorithmic biases/neutrality, machine learning, and the professional application of these tools in the field

Slides available at https://bit.ly/diti-su2021-Blum

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

Big Data is here (and it's getting *bigger*)

Other than YouTube, what other algorithms come to mind that shape your tastes/preferences and social interactions?

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

“The models were rarely able to predict a student’s GPA, for example, and they were even worse at predicting whether a family would get evicted, experience unemployment, or face material hardship. And the models gave almost no insight into how resilient a child would become.”

from the 2020 Boston Globe Article, “Algorithms may never really figure us out, thank goodness”

But does the problem lie inherently in the **algorithm** or its **application**? What do you think? Who is to blame for this failure?

Another example, from Prof. Lazar at the C-19 Research Lab at Northeastern: https://www.networkscienceinstitute.org/covid-19

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

"Finding all that information about the mother, her three children and their three fathers in the county's maze of databases would have taken Byrne hours he did not have; **The algorithm, however, searched the files and rendered its score in seconds."** from the 2018 New York Magazine article by Dan Hurley

What are the industries/public policy areas that are most prone to errors in human judgment? How would they benefit from computational/algorithmic decision making?

Imagine you are a policymaker (e.g. a social worker, a vaccine-distribution manager, anyone whose decisions will impact the lives of many others). **How should you go about utilizing AI and machine learning, and big data, in general?**

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, Jeff Sternberg, Colleen Nugent, Milan Skobic, Yana Mommadova, & Claire Tratnyek
Digital Integration Teaching Initiative
DITI Research Fellows

Taught by
Yana Mommadova
DITI Fellow
Claire Tratnyek
DITI Assistant Director

Slides, handouts, and data available at **https://bit.ly/diti-su2021-Blum**
Schedule an appointment with us! **https://calendly.com/diti-nu**

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