

Empirical Methods of Data Science

PROFESSOR MICHELLE LEVINE

WEEK 6: ETHICS, PART 3

2/27/19

Defining ethics in the age of big data

Does our definition of ethics need to be modified?

Problem with consent and big data

IRB approval and big data

Last week

Current state of affairs

- Why and what do people share?
- Potential consequences of social media

“Surveillance Capitalism”

Shoshana Zuboff

- Harvard Business School faculty
- Background in sociology

“unilaterally claims human experience as free raw material for translation into behavioral data”

Some ethical issues in the news

Tracking

Data privacy

Using data for personal use

Conflicts of interest

Facebook's "consent"

Data Use Policy

- "may use the information we receive about you... for internal operations, including troubleshooting, data analysis, testing, research and service improvement."
- [Link to policy](#)

Facebook/Google vs Twitter

Twitter:

- Inherently thought of as being more public in nature
- Provides easy access to their data via their application programming interface (API)
 - streaming (real time) data and rest (historical) data
- But also has stricter privacy policies concerning data use
 - Twitter requires people to update their datasets if/when Tweets are deleted.
 - *Although this seems to protect the humans behind the data, what other issue does this cause?*
- [Link to policy](#)

Attitudes of social media users

Research using Social Media; Users' Views (Beninger et al., 2014)

Interviewed social media users to determine:

- User attitudes about the data being used for research
- How well they understand the research that may use their data

Found that people's views fell into 3 categories:

- 1) Skepticism
- 2) Acceptance
- 3) Ambivalence

Effect of research context

4 factors that influenced participants' views and expectations:

- 1) Mode of content of the post
- 2) Website being used
- 3) Expectations the user had when posting
- 4) Nature of the research and researchers' affiliation

Consent and anonymity are unnecessary

Responsibility lies with the user; they can choose where, what and how privately to post

The site owners should make it more clear how public posts are and who can access them

Informed consent is needed

Because it's morally and legally required

To promote trust between the researcher & the participant

To quote a username alongside a post

If a post is not recent, to confirm the user has not changed their opinion since

To publish photos or other imagery

If the post could be considered particularly sensitive or personal

To confirm if the user intended to post publicly

If the posts would be used to make a profit

For users to determine the quality & purpose of the research

Anonymity is needed

Especially if informed consent is not gained

To avoid harm, including judgements or ridicule

To preserve or protect your professional reputation

Emerging cs codes of ethics

ACM (Association for Computing Machinery)

- 100,000 members from 190 countries
- Updated code of ethics in 2018 after 26 years
- Now ethics incorporates emerging systems, discrimination and data privacy concerns.
- [Link to code](#)

IEEE (Institute of Electrical and Electronics Engineers)

- Also revised their code of ethics in 2018
- [Link to code](#)

Zook et al (2017)

Ten simple rules for responsible big data research

- “Social scientists now grapple with data structures and cloud computing, while computer scientists must contend with human subject protocols and institutional review boards (IRBs).”

- 1) Acknowledge that data are people and can do harm
- 2) Recognize that privacy is more than a binary value
- 3) Guard against the reidentification of your data
- 4) Practice ethical data sharing
- 5) Consider the strengths and limitations of your data; big does not automatically mean better

-
- 6) Debate the tough, ethical choices
 - 7) Develop a code of conduct for your organization, research community, or industry
 - 8) Design your data and systems for auditability
 - 9) Engage with broader consequences of data and analysis practices
 - 10) Know when to break these rules

Council for Big Data, Ethics, and Society

Started in 2014

Brings together social and cultural perspectives in big data in order to safe guard against mistakes in big data research

E.g. Mannheim, Young, & Rossman (2016)

3 related perspectives may help you examine the ethics behind a study:

- 1) Context: In what context is the research being conducted?
- 2) Expectation: What are user expectations surrounding social networking services and the use of their social media data for research purposes?
- 3) Value Analysis: Does the benefit of the research outweigh the potential risks or privacy violations to social media users?

Discuss debate from last week

Let's consider:

- Privacy expectations / trust
 - Does this play a role in whether running the experiment was ethical?
- Public vs private access of research results
 - Does this impact your view on whether it was ethical to run the study?
- View of data
 - Do you consider the data human subjects data or web documents?
- Consent through FB vs through university
 - Where does responsibility lie?