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

## What is "Big Data"?

"Big Data" is a term which refers to the information collected from large numbers of users or other entities and analyzed collectively. Often, the goal of big data is to predict individual user behavior from patterns developed based on demographically similar users.

- Users' information is made into a valuable product through "surveillance capitalism."
- Usage of big data continues to grow, especially by large companies.
- Because this data is often privately controlled, effective oversight can be very difficult or even impossible.

## Questions to Consider

- How are we being represented online?
- Where is data about our lives coming from, and how is it being collected?
- Who is using our data and for what purposes?
- How might our data be used in the future?
- How does "big data" impact our daily lives?

## Online Presence & Data Privacy

Big data impacts many facets of our daily lives. Audio and visual streaming services track user choices to redirect your experiences in real-time (providing "what to watch next" recommendations or endless music mixes). The healthcare industry uses it to keep records of your medical history and track the effectiveness of treatments across demographics. Big data is also used in:

- Shopping and marketing.
- Travel and transportation.
- Education and employment.
- News and information services.
- Public policy and safety.

Platforms which use big data often collect a greater amount of data than users typically realize. This data continues to become more granular (more specific and intimate) as services increase their usage of big data. For example, TikTok's user agreement allows it to collect biometric data such as faceprints and voiceprints to use for content and ad recommendations.

- Learn about your Google location history here: https://www.google.com/maps/timeline.

**Developed by:** Chris McNulty and Benjamin Grey
**Questions? Contact us!** nulab.info@gmail.com

- Developed by *The Markup*, the real-time privacy inspector called Blacklight allows you to reveal tracking technologies hidden in websites.
- Facebook, Google, Instagram and TikTok all have features which allow you to download a copy of your personal data.

**Ethics and Algorithmic Bias**

Algorithms (a set of processes or rules which a computer follows) are used to collect, parse, and analyze big data. These algorithms are not neutral, but are the result of programmatic choices. Categories used to collect data often render certain persons, identities, qualities, or characteristics invisible.

- They can reinforce systemic, political, and/or cultural biases.
- When this big data becomes the basis of policy decisions or resource allocation, the consequences can be significant.
- When data reflects biased realities, algorithms will continue to reproduce trained outcomes.

When training algorithms, we must ask what kinds of data are being counted, because "what gets counted counts."[1]

**Scholarly Biases and Archival Silences**

How can we challenge biases, assumptions, and norms within our academic research? Consider:

- Whose voices and expertise are valued and heard?
- What kinds of data are prioritized in scholarship, and how/how often are they used?
- Whose voices and experiences and bodies can we easily find in the historical record, and whose are missing?
- What other sources of information might help fill in gaps in the 'official' records found in archives and academic discourse?

When considering archives and the "historical record," ask yourself:

- What information gets saved, and what doesn't?
- Who makes the decisions about what can and cannot be included in "official" records?
- Whose voices, bodies, and experiences are missing from the historical record?
- How can our work be a response to, or disruption of these silences?

**Data Ethics Moving Forward**

How are data ethics relevant for you?

[1] Joni Seager quoted in D'Ignazio, C. and L. Klein. (2020) *Data Feminism*. Cambridge: MIT Press.

- You will likely be collecting, analyzing, and using data in some way.
- Consider from where, and by what methods your data was collected (were people aware of the process, what biases or choices were made in the collection process, who benefits or might be harmed by its collection?).
- Critically examine studies which rely on big data.
- Be mindful of infographics and data visualizations (what is obvious or obscured?).
- Verify non-traditional sources and cite them appropriately.