AI and Research Ethics – Theory & Practice

NoBIAS-Summer School 2021

Katharina Kinder-Kurlanda

Recap

- Building on lecture in NoBIAS Onboarding week
- Ethical decision making in the individual NoBIAS research projects
- Various ethical decision points arise in any research, often in connection with data used

Intro: Research ethics

The research must not cause harm or disadvantages to anyone
(physical or psychological)

- What should we do and why?
- What are the consequences for others / for society?

Informed consent

- Participation in studies must be voluntary
- Participation must be „informed", i.e. information about the goals and methods of the research must be supplied (who, what, how)
- Participants have a right to control their personal information, i.e. must be able to withdraw consent
- Personal data can only be recorded and processed with participants' consent

Privacy

- Anonymisation of reserach data to ensure that particpants' privacy is protected

Research ethics in the social age

The emergence of new technologies often leads to conceptual gaps in how we think about ethical problems, and how we address them

Privacy

- Presumption that because subjects make information available online, they do not have an expectation of privacy

- Assumes no harm could come to subjects if data is already "public"

But:
- *contextual* nature of sharing
- users may not be aware of how public content is
- old dichotomy of public/private does not always apply in today's networked-data society

"But the data is already public": on the ethics of research in Facebook

Michael Zimmer

Abstract In 2008, a group of researchers publicly released profile data collected from the Facebook accounts of an entire cohort of college students from a US university. While good-faith attempts were made to hide the identity of the institution and protect the privacy of the data subjects, the source of the data was quickly identified, placing the privacy of the students at risk. Using this incident as a case study, this paper articulates a set of ethical concerns that must be addressed before embarking on future research in social networking sites, including the

The dataset comprises machine-readable files of virtually all the information posted on approximately 1,700 [Facebook] profiles by an entire cohort of students at an anonymous, northeastern American university. Profiles were sampled at 1-year intervals, beginning in 2006. This first wave covers first-year profiles, and three additional waves of data will be added over time, one for each year of the cohort's college career.

Though friendships outside the cohort are not part of

Anonymization

- Anonymization difficult or impossible to achieve with e.g. social media data

- Very easy to find out many things about people from very little information

- Let's try this...

www.aau.at

- Often anonymization is very easily possible, but…
- A focus on privacy no longer necessarily helps to address important ethical concerns, or may even hide important issues

Consent

- Because something is shared online or available without a password, the author is not 'consenting' to it being harvested for research
  - Clicking OK to ToS is not 'informed consent'
- We must recognize that a user making something public online has specific assumptions and expectations about who can access data (and how, why, for what)

- It is no longer important to know who exactly you are, it is important to have information so that you can, for example, offer personalized services and advertising
- Without knowledge of the legal real world identity problematic practices of controlling and monitoring (or at least the attempts to do so) are applied
- It has become possible to draw various conclusions about the individual that could not be known before in the absence of identifiability (e.g. personalization)
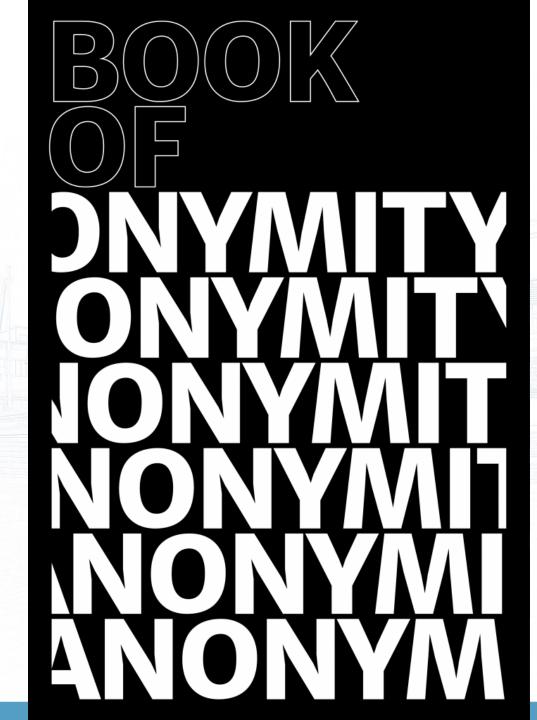
- Furthermore, it is possible to infer from the characteristics of some to those of others, whose consent to research is then no longer necessary, if only the sampling bias can be addressed
- A focus on increasingly sophisticated privacy preserving methods holds therefore not only little promise, but may even obscure the real problems
- Focus should rather be on the purposes for which different data practices are used

"Where, for example, anonymizing data, adopting pseudonyms, or granting or withholding consent makes no difference to outcomes for an individual, we had better be sure that the outcomes in question can be defended as morally and politically legitimate."

(Barocas, Solon & Helen Nissenbaum: Big Data's End Run around Anonymity and Consent" In: Book of Anonymity, edited by Anon Collective. Milky Way, Earth: punctum books (2021), p. 116-141.)

- Data-generating systems are designed and implemented to generate and hold very specific data that are not originally designed to be used as research data

- By using internet platform data, we also become complicit in such platforms' surveillance and in business practices that aim to generate value out of data

Response-ability

- Ethics of 'response-ability' (Haraway 2008)
- Responsibility toward the researcher's world
- A researcher's capacity and willingness to be moved, in both the affective and kinesthetic sense (Latour 2004)
- A practice of making oneself available to respond without knowing ahead of time which phenomena will call one's attention or what form the response should take

Martin, Aryn, Natasha Myers, and Ana Viseu. "The politics of care in technoscience." *Social Studies of Science* 45.5 (2015): 625-641.

Response-ability

- Not closing oneself off from research subjects
- No research at a distance

Research ethics as something that cannot be "outsourced"

AoIR guidelines

"guidelines rather than a code of practice so that ethical research can remain flexible, be responsive to diverse contexts, and be adaptable to continually changing technologies."

https://aoir.org/ethics/

Internet Research: Ethical Guidelines 3.0
Association of Internet Researchers

Unanimously approved by the AoIR membership October 6, 2019

aline shakti franzke (University of Duisburg-Essen), Co-Chair
Anja Bechmann (Aarhus University), Co-Chair,
Michael Zimmer (Marquette University), Co-Chair,
Charles M. Ess (University of Oslo), Co-Chair and Editor

The AoIR IRE 3.0 Ethics Working Group, including: David J. Brake, Ane Kathrine Gammelby, Nele Heise, Anne Hove Henriksen, Soraj Hongladarom, Anna Jobin, Katharina Kinder-Kurlanda, Sun Sun Lim, Elisabetta Locatelli, Annette Markham, Paul J. Reilly, Katrin Tiidenberg and Carsten Wilhelm.[1]

AI and Machine Learning: Internet Research Ethics Guidelines (IRE 3.0 6.1)

Anja Bechmann, Aarhus University & Bendert Zevenbergen, Princeton University

www.aau.at

Source of Training Data

The inferences and predictions of an AI system are closely connected to the source of the training data and here especially issues on systemic discrimination or biases are interesting to disclose and reflect upon as many previous studies have shown such effects (Barocas & Selbst, 2016; Bechmann & Bowker, 2019; boyd & Crawford, 2012; Crawford & Calo, 2016; Kroll et al., 2017; Sweeney, 2013). The use of AI systems to uncover or predict social phenomena can thus be tainted by biases in the training data set on certain demographics or proxies thereof, which may lead to unfair and unjust outcomes.

- What is the cultural and sociodemographic profile of the datasets used by the researcher to train the models?

40

- To what extent does the cultural and sociodemographic profile of the training data allow for generalizability of the resulting findings or predictors from the research study?
- Are there particular groups which may be advantaged or disadvantaged, in the context in which the researcher is deploying an AI-system? What is the potential damaging effect of uncertainty and error-rates to different groups?
- How has the demographic profile of the researcher(s) affected the composition of the training data?
- How does the training data as 'ground truth' affect different demographic profiles and proxies thereof?

Data Cleaning

Data cleaning is the process of detecting, correcting, replacing and even removing inaccurate and incomplete records from a database and structuring the data in a consistent way that

Group Work in Google Doc

GROUP 1 (A-K)

https://docs.google.com/document/d/1O42kTHPKqfDZQaTCqJVxoq-GC4Dnh5Opj55DN_RrpxA/edit?usp=sharing

GROUP 2 (L-Z)

https://docs.google.com/document/d/1As-7n35m_KcuZMinSjuHkNsVriWh96AGedPdNNQVOSU/edit?usp=sharing

References

- Ess, C. 2002. *Ethical decision-making and Internet research: Recommendations from the aoir ethics working committee.* Available: www.aoir.org/reports/ethics.pdf

- Markham, A., and Buchanan, E. 2012. *Ethical decision-making and Internet research 2.0: Recommendations from the aoir ethics working committee.* Available: www.aoir.org/reports/ethics2.pdf

- Barocas, S., & Nissenbaum, H. (2014). Big Data's End Run around Anonymity and Consent. In J. Lane, V. Stodden, S. Bender, & H. Nissenbaum (Eds.), Privacy, Big Data, and the Public Good: Frameworks for Engagement (pp. 44-75). Cambridge: Cambridge University Press.