

Exploring Ethical Trade-Offs in Social Media Research

Casey Fiesler, University of Colorado Boulder (Contact)

Stevie Chancellor, Georgia Institute of Technology

Katie Shilton, University of Maryland

Jessica Vitak, University of Maryland

Michael Zimmer, University of Wisconsin-Milwaukee

Abstract

Over the past decade, social media and user-generated content platforms have increasingly become rich research sites for the study of both computation and human behavior. This new source of pervasive human data has also sparked discussions within the research community about ethical challenges, and high profile examples have raised public awareness of ethical challenges as social media scholarship gains greater visibility. However, the research community lacks clear norms, and disagreement often comes down to how to identify and weigh potential benefits and harms. The goal of this workshop is to explore the most pressing ethical dilemmas within social media research, and how the ICWSM research community can best consider the ethical implications of our research and methods without compromising important work. Workshop participants will have the opportunity to shape a set of working guidelines to help researchers think through the ethics of social media research methods.

Motivation and Themes

With social media and other user-generated content platforms existing as part of everyday life, we are awash in digital traces of online behavior and communication. Technology companies use this information to conduct user research and to improve their products, marketers use it to create and target ads, and researchers use it to study not only social media use and online behavior but for a huge array of other research contexts. Research published at ICWSM is a perfect example of what we can learn from studying social media data. However, as conversations about research ethics propagate in both the research community and among the general public, it is important to weigh the benefits of this work against any potential harms to individual users, communities, or platforms. New emergent issues about methods, like transparency of machine learning algorithms and dataset bias and representation, also add to the conversation on how to conduct research in rigorous and equitable ways.

There is a large body of work about ethics in digital contexts and digital research methodologies. Many issues covered by this work are relevant to the themes of this workshop, such as informed consent (Hutton and Henderson 2015; Barocas and Nissenbaum 2014), verification of individuals online (O'Neill 2013), the role of review boards (Calo 2013; Vitak et al. 2017), appropriate and effective anonymization (Bruckman 2002; Barocas and Nissenbaum 2014), relationships with and dissemination of findings to participants (Beaulieu and Estallea 2012), the definition of public spaces (Bromseth 2002) and public data (Zimmer 2010) (including data mining), legal matters and Terms of Service (Fiesler, Lampe, and Bruckman 2016; Grimmelmann 2015), privacy (Horvitz and Mulligan 2015), algorithmic ethics (Tufekci 2015), and A/B testing and corporate research (Meyer 2015). Scholars have considered how to ethically conduct research in relevant application domains, like public health (Mikal, Hurst, and Conway 2016; Conway and O'Connor 2016).

Within the social media research community, for example, it is common practice to use large amounts of public online data (such as tweets) for analysis (Zimmer and Proferes 2014).

Regulatory guidance, particularly by bodies such as institutional review boards (IRBs) at U.S. universities, is uncertain (Vitak et al. 2017). U.S. regulations are also changing, with ramifications for online data research, and identifying circumstances under which data science is human subjects research may become even more complicated. Beyond complying with university regulations, what are the trade-offs of different approaches to research ethics in online data analysis, and what best practices should we consider? What might various ethical principles such as “avoiding harm” mean in the context of big data social media research?

In addition to academic interest, high-profile examples have also prompted public discourse about research conducted both by technology companies and by academic researchers. These include data manipulation in Facebook for experimental purposes (McNeal 2014), OKCupid’s manipulation of dating matches on their for-profit website (Robbins 2014), a researcher’s public release of a non-anonymized OKCupid dataset (Zimmer 2016) and research that used publicly available photographs to train a classifier to predict sexual orientation (Burdick 2017) or gender transition (Vincent 2017). Beyond direct impact on participants or the creators of data we use, should we also be considering secondary impacts of our work when it comes to how third parties might make use of our research?

Combined with increased public awareness on issues like privacy and algorithmic bias, research ethics as a topic is no longer limited to academic circles; Internet users, corporate researchers, and the public are considering their digital traces and who might be watching them. These questions are of increasing interest to the social computing research community, and yet there are no clear answers, with ethical practices varying from discipline to discipline and person to person (Vitak, Shilton, and Ashktorab 2016). Industry groups have begun consensus-based processes to set principles for ethical data science practice. This workshop will build towards a similar goal. By engaging members of the ICWSM community with different opinions about how to weigh research benefits and harms, we can outline guidelines that consider ethical trade-offs while maintaining the integrity of our research.

Workshop Structure

This workshop is structured to facilitate discussion of ethical challenges faced by researchers working with social media and similar sources. Over the course of one day, we will present and discuss case studies and hypotheticals, consider work done in industry and cognate research fields, conduct brainstorming sessions, and emerge with ideas for best practices. We also recognize that this is a space that is bound to present more questions than answers, and our goal is to encourage and facilitate more conversation about these issues outside the confines of the workshop. Though we anticipate adjusting the exact program based on the make-up of our participants, the basic structure of activities (spaced out across a roughly 8-hour workshop day, with a break for lunch) includes:

- A small number of brief presentations from participants focusing on case studies of actual ethical dilemmas faced or approaches adopted
- Group brainstorming about open questions and challenges for social media research ethics
- Small group activities around real or hypothetical scenarios (drawn from our participants, as well as from established case examples such as those from the Council for Big Data, Ethics, and Society), focused on identifying ethical pain points for researchers and brainstorming solutions

- Small group “task force” discussions about one particular ethical challenge (e.g., biased algorithms, use of public data, anonymization, sensitive communities), focused on the trade-offs of different approaches
- Group debate around and formulation of a set of best practices for social media researchers as well as data professionals
- Round table discussions of how to “bring back” the outcomes of the workshop to the rest of ICWSM, as well as home departments, labs, and other interested parties

Submissions

We will accept two types of submissions for participation in this workshop, longer position papers and short statements of interest. Submissions will be reviewed by the workshop organizers, with external reviewers tapped as necessary depending on submission numbers.

Position papers should be no more than 6 pages and should be submitted in AAAI format (see author guidelines for ICWSM). These papers if accepted will be included in the ICWSM workshop proceedings. These papers should be on the topic of ethics in social media research, including but not limited to:

- (1) studies or works-in-progress related to ethics;
- (2) description of a particular approach to ethics, supported by your or others’ work; or
- (3) case studies of particular ethical challenges faced.

Statements of interest should be no more than 2 pages and can be submitted in any format. These documents will not be included in the workshop proceedings, but if accepted will be shared on the workshop website. Topics can be similar to position papers or can simply explain the potential participants’ interest in the subject matter and why they wish to participate in the workshop.

Relevant topics for this workshop include any issues related to social media research and ethics, including but not limited to: informed consent, sensitive populations, ethical implications of data mining, algorithmic harm, definitions of public content and data, review boards, legal implications and obligations, privacy, A/B testing, algorithmic bias, secondary harms of research, relationships to study populations, and relationships between corporate and academic researchers. We would also welcome position papers that argue for the benefit of certain types of research outweighing potential harms. We invite proposals from researchers from both academia and industry, and would welcome a wide range of ethical approaches and disciplinary perspectives. We also invite papers focused on specific domains of research.

Related Workshops

Some of the organizers of this workshop have been involved in similar initiatives at previous conferences. Fiesler was the lead organizer of the “Ethics for Studying Sociotechnical Systems in a Big Data World” workshop at CSCW 2015 and “Ethics and Obligations for Studying Digital Communities” workshop at GROUP 2016. Fiesler and Chancellor were co-organizers of the ICWSM 2016 workshop “Challenges and Futures for Ethical Social Media Research.” Vitak has been an organizer for a series of “Networked Privacy” workshops at CHI and CSCW, and Zimmer is currently organizing a workshop on ethics for WebSci 2018.

Aside from the 2016 ethics workshop, previous workshops at ICWSM have also touched on related issues, such as the 2015 workshop on “Standards and Practices in Large Scale Social

Media Research” and the 2017 workshop on “Studying User Perceptions and Experiences with Algorithms.”

As with the ICWSM 2016 ethics workshop (which had 11 participants), this one focuses more on large-scale data and specifically on social media research, though for this workshop our focus is on hearing from diverse voices in the ICWSM community and in considering most closely how we make decisions based on both benefits and harms of research.

Organizers

Casey Fiesler is an assistant professor in the Department of Information Science at University of Colorado Boulder. She holds both a law degree and a PhD in human-centered computing, and her research focuses largely on forms of governance online, including social norms, law, and ethics. She has organized a series of research ethics workshops at conferences including CSCW, GROUP, and ICWSM. She is a member of the SIGCHI research ethics committee.

Stevie Chancellor is a PhD student in Human Centered Computing at Georgia Tech. Her research focuses on computational methods, like applied machine learning, to study deviant mental wellness communities online. Her community of interest currently are pro-eating disorder communities. She is also interested in research ethics for large-scale data analyses. She helped organize the 2016 ICWSM ethics workshop.

Katie Shilton is an associate professor in the College of Information Studies at the University of Maryland, College Park. Her research explores ethics and policy for the design of information collections, systems and technologies. Current projects include leading the multi-campus PERVADE: Pervasive Data Ethics research project; exploring privacy-sensitive search for email collections; analyzing ethical cultures in computer security research; and building tools to facilitate ethics discussions in mobile application development.

Jessica Vitak is an assistant professor in the College of Information Studies at the University of Maryland and Associate Director of the Human-Computer Interaction Lab (HCIL). She currently evaluates challenges to networked privacy, data ownership, and ethics in social computing research (NSF Awards #0916019 and #1704369). She has organized numerous workshops for CSCW and CHI on topics related to privacy and ethics.

Michael Zimmer is an Associate Professor in the School of Information Studies at the University of Wisconsin-Milwaukee (USA), where he also serves as Director of the Center for Information Policy Research. He is a privacy and internet ethics scholar, whose work focuses on digital privacy, the ethical dimensions of social media & internet technologies, and internet research ethics. Dr. Zimmer is a co-chair of Association of Internet Researchers (AoIR) Ethics Working Group, and a principle investigator in the PERVADE: Pervasive Data Ethics project. He is currently organizing an Ethics Workshop for Web Science 2018.

REFERENCES

- Barocas, S., and Helen Nissenbaum. 2014. “Big Data’s End Run around Anonymity and Consent.” In *Privacy, Big Data, and the Public Good: Frameworks for Engagement*, edited by J. Lane, V. Stodden, S. Bender, and Helen Nissenbaum, 44–75. Cambridge, UK.
- Beaulieu, A., and A. Estallea. 2012. “Rethinking Research Ethics for Mediated Settings.” *Information, Communication & Society* 15 (1): 23–42.
- Bromseth, J.C.H. 2002. “Public Places - Public Activities? Methodological Approaches and

- Ethical Dilemmas in Research on Computer-Mediated Communication Contexts.” In *Resesarching ICTs in Context*, edited by A. Morrison, 33–61. University of Oslo.
- Bruckman, Amy. 2002. “Studying the Amateur Artist: A Perspective on Disguising Data Collected in Human Subjects Research on the Internet.” *Ethics and Information Technology* 4 (3): 217–31.
- Burdick, Alan. 2017. “The A.I. ‘Gaydar’ Study and the Real Dangers of Big Data.” *The New Yorker*.
- Calo, Ryan. 2013. “Consumer Subject Review Boards: A Tough Experiment.” *Stanford Law Review Online*, 97–102.
- Conway, Mike, and Daniel O’Connor. 2016. “Social Media, Big Data, and Mental Health: Current Advances and Ethical Implications.” *Current Opinion in Psychology* 9: 77–82.
- Fiesler, Casey, Cliff Lampe, and Amy S. Bruckman. 2016. “Reality and Perception of Copyright Terms of Service for Online Content Creation.” In *Proceedings of the ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW)*, 1450–61.
- Grimmelmann, James. 2015. “The Law and Ethics of Experiments on Social Media Users.” *University of Colorado Technology Law Journal* 13: 219–72.
- Horvitz, Eric, and Deirdre Mulligan. 2015. “Data, Privacy, and the Greater Good.” *Science* 349 (6245): 253–55.
- Hutton, Luke, and Tristan Henderson. 2015. “‘I Didn’t Sign up for This!’: Informed Consent in Social Network Research.” In *Proceedings of the 9th International AAAI Conference on Web and Social Media (ICWSM)*.
- McNeal, Gregory. 2014. “Facebook Manipulated User News Feeds to Create Emotional Responses.” *Forbes*.
- Meyer, Michelle N. 2015. “Two Cheers for Corporate Experimentation: The A/B Illusion and the Virtues of Data-Driven Innovation.” *Colorado Technology Law Journal* 13: 273–332.
- Mikal, Jude, Samantha Hurst, and Mike Conway. 2016. “Ethical Issues in Using Twitter for Population-Level Depression Monitoring: A Qualitative Study.” *BMC Medical Ethics* 17 (1). *BMC Medical Ethics*: 22.
- O’Neill, Brian. 2013. “Who Cares? Practical Ethics and the Problem of Underage Users on Social Networking Sites.” *Ethics and Information Technology* 15 (4): 253–62.
- Robbins, Martin. 2014. “Does OKCupid Need Our Consent?” *The Guardian*, July 30.
- Tufekci, Zeynep. 2015. “Algorithmic Harms Beyond Facebook and Google: Emergent Challenges of Computational Agency.” *Colorado Technology Law Journal* 13: 203–18.
- Vincent, James. 2017. “Transgender YouTubers Had Their Videos Grabbed to Train Facial Recognition Software.” *The Verge*.
- Vitak, Jessica, Nicholas Proferes, Katie Shilton, and Zahra Ashktorab. 2017. “Ethics Regulation in Social Computing Research: Examining the Role of Institutional Review Boards.” *Journal of Empirical Research on Human Research Ethics* 12 (5): 372–82.
- Vitak, Jessica, Katie Shilton, and Z. Ashktorab. 2016. “Beyond the Belmont Principles: Ethical Challenges, Practices, and Beliefs in the Online Data Research Community.” In *Proceedings of the ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW)*, 941–53.
- Zimmer, Michael. 2010. “‘But the Data Is Already Public’: On the Ethics of Research in Facebook.” *Ethics and Information Technology* 12 (4): 313–25. doi:10.1007/s10676-010-9227-5.
- Zimmer, Michael. 2016. “OkCupid Study Reveals the Perils of Big-Data Science.” *WIRED*. May.
- Zimmer, Michael, and Nicholas Proferes. 2014. “A Topology of Twitter Research: Disciplines, Methods, and Ethics.” *Aslib Journal of Information Management* 66 (3): 250–61.