

Challenges and Futures for Ethical Social Media Research

Proposal for a Full Day ICWSM Workshop

Casey Fiesler, University of Colorado Boulder (Primary Contact)

Stevie Chancellor, Georgia Institute of Technology

Anna Lauren Hoffmann, University of California, Berkeley

Jessica Pater, Georgia Institute of Technology

Nicholas Proferes, University of Maryland

Abstract

Social media and user-generated content platforms have opened up new possibilities for communication and creativity while also providing huge amounts of information about people and their online behavior. The evolution of technology and research methods presents ongoing ethical challenges to studying people and their digital traces. These challenges are a continuing source of discussion within the research community and also a topic of public discourse as social media scholarship gains greater visibility. This workshop is aimed at exploring difficult and unanswered ethical questions, touching on issues such as consent, privacy, and responsibility, as well as developing a set of best practices for social media and big data research.

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. This unique access to large-scale data that can be used to better understand human behavior and interactions with technology is a huge boon to research, and yet also raises ethical implications and challenges.

There is a large body of work about ethics in digital contexts and their evolving research methodologies. Many issues covered by this work are relevant to the themes of this workshop, such as informed consent [1,10,11], verification of individuals online [15], the role of review boards [5–7], appropriate and effective anonymization [1,4,13], relationships with and dissemination of findings to participants [2], the definition of public spaces [3] and public data [21] (including data mining), legal matters and Terms of Service [8,9], algorithmic ethics [18], and A/B testing and corporate research [14].

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 [20]. Though this type of research is often not considered under the purview of university IRBs, under what circumstances might data science cross that line into human subjects research [9]? Is this a distinction that even continues to make sense in the context of current research? Regardless, what are our obligations as researchers to social media users? What are the tradeoffs of different approaches to research ethics in this field, and what best practices should we consider? What might various ethical codes—such as the ACM’s of “avoiding harm to others” and “being honest and trustworthy”—mean in the context of big data social media research?

In addition to academic interest, recent high-profile examples have also prompted public discourse about research conducted by technology companies. Two timely examples include data manipulation in Facebook for experimental purposes [12] and

OKCupid's manipulation of dating matches on their for-profit website [16,17]. Combined with increase public awareness on issues like privacy, 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 [19]. Part of the goal of this workshop is to encourage the ICWSM community to actively engage with these open questions.

Workshop Structure

This workshop is structured to facilitate discussion of ethical challenges faced by researchers working with social media and other big data sources. Over the course of one day, we will present and discuss case studies and hypotheticals, 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., 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

In addition to these activities, we will discuss how to best disseminate ideas among our research community as well as how to continue discussions both at ICWSM and ongoing after the workshop.

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 5 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, relationships to study populations, and relationships between corporate and academic researchers. We invite proposals from researchers from both academia and industry, and would welcome a wide range of ethical approaches and disciplinary perspectives.

Related Workshops

Some of the organizers of this workshop have been involved in similar initiatives at previous conferences. Fiesler was the primary organizer of the "Ethics for Studying Sociotechnical Systems in a Big Data World" workshop at CSCW 2015 (with 28 participants), and Anna Lauren Hoffmann co-organized a roundtable on "Teaching Ethics in Big Data and Social Media" at the 2015 Association of Internet Researchers conference. At ICWSM, the 2015 workshop on "Standards and Practices in Large-Scale Social Media Research" also touched on topics related to research ethics. Though the CSCW 2015 workshop is the most similar in format to this one, the proposed workshop will focus more on large-scale quantitative data and specifically on social media research. We hope that this is a continuing series of ethics-related workshops, with each tailored to a specific community of researchers.

Workshop Organizers

Casey Fiesler is an Assistant Professor of Information Science at the University of Colorado Boulder. Her research focuses on the intersection of social computing and law, and in particular the legal and ethical gray areas of online content creation. She holds a law degree from Vanderbilt University and a PhD in Human-Centered Computing from Georgia Tech.

Stevie Chancellor is a PhD student in Human Centered Computing at Georgia Tech. She is broadly interested in the intersection of deviant online communities and computation. Her current research focuses on creating automated methods to find and assess the well-being of mental illness communities, such as pro-eating disorder and self-injury.

Anna Lauren Hoffmann is a Postdoctoral Researcher and Instructor at the University of California, Berkeley. Her research is situated at the intersection of information, technology, ethics, and culture, with particular attention to the ways in which the design and use of information technology can promote or hinder the pursuit of social justice. She holds a BA in Studies in Cinema and Media Culture from the University of Minnesota and an MLIS and PhD in Information Studies from the University of Wisconsin-Milwaukee.

Jessica Pater is a research scientist at the Georgia Tech Research Institute and a PhD student in Human Centered Computing at Georgia Tech. Her research is focused on how everyday social computing uses impact the health of vulnerable population, most recently focusing on individuals with non-suicidal self-harm.

Nicholas Proferes is a Postdoctoral Scholar at the University of Maryland's College of Information Studies, His research explores the intersections between information flow, user cognition, algorithmic transparency, and values and ethics. He holds a PhD from the University of Wisconsin-Milwaukee's School of Information Studies.

REFERENCES

1. S. Barocas and Helen Nissenbaum. 2014. Big data's end run around anonymity and consent. In *Privacy, big data, and the public good: Frameworks for engagement*, J. Lane, V. Stodden, S. Bender and Helen Nissenbaum (eds.). Cambridge University Press, Cambridge, UK.
2. A. Beaulieu and A. Estallea. 2012. Rethinking research ethics for mediated settings. *Information, Communication & Society* 15, 23–42.
3. J.C.H. Bromseth. 2002. Public places - public activities? Methodological approaches and ethical dilemmas in research on computer-mediated communication contexts. In *Resesarching ICTs in Context*, A. Morrison (ed.). University of Oslo, 33–61.
4. Amy Bruckman. Studying the Amateur Artist: A Perspective on Disguising Data Collected in Human Subjects Research on the Internet. In *Ethics and Information Technology*.
5. E. Buchanan and Charles Ess. 2009. Internet research ethics and the Institutional Review Board: Current practices and issues. *Computers and Society* 39, 43–49.
6. Ryan Calo. 2013. Consumer Subject Review Boards: A Tough Experiment. *Stanford Law Review Online*, 97–102.
7. Kate Crawford and Jason Schultz. 2013. Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms. *Boston College Law Review* 55, 1.
8. Casey Fiesler, Cliff Lampe, and Amy S. Bruckman. 2016. Reality and Perception of Copyright Terms of Service for Online Content Creation. *Proceedings of the 2016 ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW)*.
9. James Grimmelman. 2012. The Law and Ethics of Experiments on Social Media Users. *Colorado Technology Law Journal* 13, 219–272.
10. James M. Hudson and Amy Bruckman. 2004. "Go Away": Participant Objections to Being Studied and the Ethics of Chatroom Research. *The Information Society* 20, 2, 127–139.
11. Luke Hutton and Tristan Henderson. 2015. "I didn't sign up for this!": Informed consent in social network research. *Proceedings of the 9th International AAAI Conference on Web and Social Media (ICWSM)*.

12. A.D.I. Kramer, J.E. Guillory, and Jeff T. Hancock. 2014. Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences* 111, 24, 8788–8790.
13. C. Markham and E. Buchanan. 2002. *Ethical decision-making and Internet research: Recommendations from the AOIR ethics working committee*. Idea Group Publishing.
14. Michelle N. Meyer. 2015. Two Cheers for Corporate Experimentation: The A/B Illusion and the Virtues of Data-Driven Innovation. *Colorado Technology Law Journal* 13, 273–332.
15. Brian O’Neill. 2013. Who Cares? Practical Ethics and the Problem of Underage Users on Social Networking Sites. *Ethics and Information Technology* 15, 4, 253–262.
16. Martin Robbins. 2014. Does OKCupid need our consent? *The Guardian*.
17. Christian Rudder. 2014. *Dataclysm: Who We Are (When We Think No One Is Looking)*. Crown Publishing Group, New York, NY.
18. Zeynep Tufekci. 2015. Algorithmic Harms Beyond Facebook and Google: Emergent Challenges of Computational Agency. *Colorado Technology Law Journal* 13, 203–218.
19. Jessica Vitak, Katie Shilton, and Z. Ashktorab. 2016. Beyond the Belmont Principles: Ethical Challenges, Practices, and Beliefs in the Online Data Research Community. *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW)*.
20. Michael Zimmer and Nicholas John Proferes. 2014. A Topology of Twitter Research: Disciplines, Methods, and Ethics. *Aslib Journal of Information Management* 66, 3, 250–261.
21. Michael Zimmer. 2010. “But the data is already public”: On the ethics of research in Facebook. *Ethics and Information Technology* 12, 4, 313–325.