Exploring Ethics and Obligations for Studying Digital Communities

Casey Fiesler
University of Colorado Boulder
casey.fiesler@colorado.edu

Jessica Pater
Georgia Institute of Technology
pater@gatech.edu

Pamela Wisniewski
University of Central Florida
pamwis@ucf.edu

Nazanin Andalibi
Drexel University
naz@drexel.edu

ABSTRACT

Many of the most prominent and unanswered ethical questions within HCI and social computing involve our ethical obligation to the communities that we study. Some of these questions fall under the purview of more traditional human subjects research ethics, but others hinge on when, for example, studies of *public* data trigger similar obligations. Basic rules to "do no harm" are complicated in digital communities by issues of consent and privacy, and ethics review boards are struggling to keep up even as research communities are similarly struggling to form appropriate norms. The goals of this workshop are to continue seeding conversations about research ethics within the SIGCHI community, to work towards norm setting, and in the meantime, to collectively help community members make good ethical decisions about research into sociotechnical systems and digital communities.

Keywords

ethics; hci; human subjects; online communities; policy; privacy; social computing; sociotechnical systems

1. INTRODUCTION

Research ethics in the context of social computing is an increasing topical point of discussion, particularly in light of recent public scrutiny. For instance, media accusations of unethical research by Facebook and their research collaborators sparked disagreement throughout our community about best practices for informed consent [18]. OKCupid's manipulation of dating matches sparked discussion about experimentation and the nature of harm to users [21]. Most recently, researchers released a large, non-anonymized scraped dataset of OKCupid profile data that provoked debate about the nature of public data, expectations of privacy, and obligations of researchers to follow Terms of Service [26].

In the face of evolving technology, research methods, and norms, online settings continue to challenge traditional notions of research ethics. Moreover, around many issues there are not currently clear rules or norms among researchers, who make highly contextualized decisions about ethics that often vary across

disciplines [23]. Ideas found in the Belmont Report or the ACM ethical codes of avoiding harm are no longer simple, particularly as the line between "human subject" and "public data" are blurred by the huge amount of user- generated content to which researchers have access. In this new environment, we should think critically about our obligations to the communities we study—whether those communities *know* that they are being studied or not.

2. BACKGROUND AND THEMES

Research ethics in digital contexts is a topic of increased interest to the SIGCHI community as it matures. This is partially because socio-technical systems provide a new context to investigate aspects of human behavior that have not always been accessible otherwise, and comes wit h new challenges, dilemmas, and opportunities. Recent workshops on the subject of research ethics for sociotechnical systems and social media [10,12] have also highlighted a number of salient issues for this community. Though much recent discussion around ethics has been focused on "big data," we focus specifically here on the *humans* involved in our research. For this workshop, our primary inquiry is framed as: *What are our ethical obligations to the communities that we study?* What if the participants of those communities are particularly vulnerable? What if we find out information that should be reported (such as threats of harm)? How can we appropriately protect their privacy? When and how should we disseminate our findings to these communities? What are our obligations to those who produce the *public data* that we study—e.g., non-consenting Twitter users? When does data science become human subjects research, and what does this mean for our ethical obligations?

Below, we present a number of ethical questions and challenges that may arise during various stages in the research process in order to identify key themes through would be of interest to the workshop organizers and participants.

2.1 Designing Research Protocols

At the *conception* of a new research project, how can we embed ethical principles into our study or technology designs? Past research has examined this challenge as it relates to the role of institutional review boards [8] and legal and contractual matters [11,13]. However, IRBs and lawyers may not be as knowledgeable about the context of our research as we are, thus missing some potential harm to participants or the institution. At times, researchers may leverage this lack of knowledge to their advantage, creating ethical challenges. For example, many IRBs do not require approval for research that does not explicitly

involve direct contact with human subjects. However, when researchers use automated approaches to, for instance, scrape public or semi-public data from users, humans are implicitly involved. Further, this type of automated scraping is often explicitly against a website's Terms of Service, making both the participants and the institution itself vulnerable.

When it comes to the design of corporate research, other ethical challenges can arise related to A/B testing [19] and algorithmic manipulation [22]. One example of these types of ethical conundrums is Facebook's "emotional contagion" study, where Facebook users were involved in a research study without their knowledge [18,19].

2.2 Obtaining Informed Consent

When obtaining *informed consent* [4,14,15] from participants, how do we ensure that this consent represents both *awareness* and *choice*? In the case of big data [4], the sheer volume of participants makes individualized consent difficult or impossible. In other contexts, such as chat rooms [14] and social networks [15], consent challenges are even more nuanced. How do we obtain consent while minimizing the burden on participants? When dealing with vulnerable populations [17] and minors [24], informed consent is increasingly necessary for preventing harm. Yet, some argue that these populations have limited capacity to provide such consent [17]. For example, parents are required to consent for minors to participate in human subjects research, but the assent of the minor is not required. Yet, the researchers' role as mandated reporter [24] forces us to breach the minor's confidentiality in the event that we detect that they are at imminent risk of harm.

2.3 Collecting and Analyzing Data

How do we ensure that the *data* we collect complies with ethical standards regardless of whet her we obtain it directly from participants or a third party? Does publicly available data within public spaces [6,25], for instance, give researchers the right to collect and aggregate this data for purposes other than the original intent of the poster? When aggregating otherwise disparate data, how do we ensure the appropriate anonymization [7,25] of this data?

When *analyzing* the data that has been collected, how do we ensure that our methods are *sound* and *robust*? Further, how can we address our *own* vulnerability as researchers, particularly when studying communities that are in distress [1,2] or present challenging content or behaviors [3,16,20]?

2.4 Disseminating Results

Finally, when *disseminating* our results, what is our responsibility to our participants? For example, have we appropriately anonymized [7,25,26] the data prior to publication? Even if the data were anonymized upon publication, could publically available data then be traced back to an individual through rudimentary means such as a Google search? At the *conclusion* of a research project, what has the researchers' relationship with the participants become after using highly engaging techniques such as ethnographic immersion [5]? What then are the norms related to participant privacy [4,9] and the dissemination of findings to participants within the community being studied [5]?

Questions like these are of increasing importance to our research community, and yet there are no clear answers. The recently appointed ACM SIGCHI ethics working group [27] has proposed, in part, that we have more community discussions about these thorny ethical issues so that we can begin to converge on best practices and norms. Therefore, we propose this workshop as one venue to start brainstorming this myriad of ethical issues so we can move toward this goal.

3. WORKSHOP STRUCTURE

The workshop will be structured to facilitate conversations around the aforementioned ethical challenges of human subjects research (defined broadly as research that involves *humans* or the *data* they produce) as it pertains to the field of collaborative and social computing. Throughout the course of the workshop, we will use actual case studies as well as hypotheticals to probe tension points and engage on a deeper level on the issues about which there are not clear answers.

We anticipate adjusting the program based on the interests and makeup of the participants in the workshop, but inspired by previous workshops on this topic [10,12], activities will include:

(1) Brief 2 minute introductions from all participants about one ethical challenge they have faced in their work and their respective reflections;

(2) Brief 5 or 10 minute presentations from a selected set of workshop participants, on specific topics of interest (such as the open questions noted above) or presenting case studies of actual ethical dilemmas faced;

(3) Group brainstorming to identify the most pressing questions and ethical challenges facing the community;

(4) Small "task force" style breakout groups to deep dive into specific issues or topics identified in ; and

(5) Group work around development of best practices and next steps for further engaging the broader community and disseminating the results of the workshop .

We will select participants based on the quality and depth of reflections presented in submissions. We will select presenters based on the potential to generate discussion and the extent to which their concerns are shared with all participants' submissions.

As part of (5), we also intend to come up with ways to engage the broader GROUP community during and after the conference—for example, by presenting provocative questions to attendees or getting feedback on ideas we propose at the workshop. Research ethics goes beyond IRB approval, and is something that should be important to every attendee of the conference. We feel that an important outcome of the workshop is to engage the entire community and not just workshop participants. With respect to deliverables, potential ideas include a collective blog where the organizers write the first post by summarizing the discussion points, and where participants as well as the broader community will be invited to contribute their ongoing reflections. Other ideas and specifics will be discussed and agreed upon together with participants.

Therefore, goals and planned outcomes for this workshop include: (1) documentation of the important open questions in the ethical space of studying digital communities; (2) documentation of brainstorming towards norm setting and current best practices; and (3) planning for engaging the GROUP community with these issues during the conference.

This workshop will be a full day, with a maximum of 30 participants including organizers (with no special equipment required beyond a projector).

4. SUBMISSIONS

In order to be considered for participation in the workshop, potential participants should present a short (1-2 page) statement of interest, which will be reviewed by the workshop organizers.

Submissions should also include a brief biographical sketch that includes current affiliation and research area.

Statements of interest can be structured in one of three ways: (1) an explanation of one's interest in research ethics (or a specific topic), ideally tied to one's own work; (2) a discussion of a specific topic in the area, e.g., one of the provocative open questions we have proposed; or (3) a case study discussion of an actual ethical dilemma faced in one's work.

All relevant topics related to ethics or obligations in studying digital communities, sociotechnical systems, or other technology-mediated groups are welcomed and encouraged. Examples include but are not limited to: informed consent, sensitive or vulnerable populations, algorithmic harm to users and communities, definitions of public content and data, the role of review boards, legal implications and obligations, reporting obligations, privacy, and relationships to study populations. We invite submissions from researchers from both academia and industry, and would welcome a wide range of disciplinary perspectives.

5. ORGANIZERS

Casey Fiesler is an Assistant Professor in the Department 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 has organized workshops on the topic of research ethics at multiple conferences, including CSCW and ICWSM, and is currently a member of the SIGCHI Ethics Committee. She holds a law degree from Vanderbilt University and a PhD in Human-Centered Computing from Georgia Tech.

Pamela Wisniewski is an Assistant Professor in the College of Engineering and Computer Science at the University of Central Florida. Her research interests are situated in Human-Computer Interaction and lie at the intersection of social computing and privacy. Her goal is to frame privacy as a means to not only protect end users, but more importantly, to enrich online social interactions that individuals share with others.

Jessica Pater is a research scientist at the Georgia Tech Research Institute and a PhD candidate in Human Centered Computing at Georgia Tech. Her research focuses on how everyday social computing uses impact the health of vulnerable populations, most recently focusing on individuals with non-suicidal self-harm. In this work, she has advocated for expanding conversations of what is ethical for researchers focused on working with this vulnerable population.

Nazanin Andalibi is a PhD candidate at the College of Computing and Informatics at Drexel University. Her research focuses on people's social media disclosure and social support practices in stigmatized contexts. She argues that it is important for the HCI community to consider researchers' occupational vulnerability in discussions, trainings, practices, and policies around research ethics.

6. REFERENCES

1. Nazanin Andalibi and Andrea Forte. 2016. Social Computing Researchers, Vulnerability, and Peer Support. *CHI 2016 Ethical Encounters in HCI Workshop*.
2. Nazanin Andalibi, Oliver L Haimson, Munmun De Choudhury, and Andrea Forte. 2016. Understanding Social Media Disclosures of Sexual Abuse Through the Lenses of Support Seeking and Anonymity. *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, ACM.
3. Nazanin Andalibi, Pinar Ozturk, and Andrea Forte. 2015. Depression-related Imagery on Instagram. *Proceedings of the ACM Conference Companion on Computer Supported Cooperative Work & Social Computing*, ACM, 231–234.
4. Solon 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.
5. Anne Beaulieu and Adolfo Estallea. 2012. Rethinking research ethics for mediated settings. *Information, Communication & Society* 15, 1: 23–42.
6. 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.
7. Amy Bruckman. Studying the Amateur Artist: A Perspective on Disguising Data Collected in Human Subjects Research on the Internet. In *Ethics and Information Technology*.
8. Elizabeth A. Buchanan and Charles Ess. 2009. Internet research ethics and the Institutional Review Board: Current practices and issues. *Computers and Society* 39: 43–49.
9. 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.
10. Casey Fiesler, Stevie Chancellor, Anna Lauren Hoffmann, Jessica Pater, and Nicholas John Proferes. 2016. Challenges and Futures for Ethical Social Media Research. *AAAI Conference on Web and Social Media (ICWSM): Workshop*.
11. Casey Fiesler, Cliff Lampe, and Amy S. Bruckman. 2016. Reality and Perception of Copyright Terms of Service for Online Content Creation. *CSCW '16: Proceedings of the 2016 ACM Conference on Computer-Supported Cooperative Work & Social Computing*.
12. Casey Fiesler, Alyson Young, Tamara Peyton, et al. 2015. Ethics for Studying Online Sociotechnical Systems in a Big Data World. *CSCW 2015 Companion*.
13. James Grimmelmann. 2015. The Law and Ethics of Experiments on Social Media Users. *University of Colorado Technology Law Journal* 13: 219–272.
14. 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. http://doi.org/10.1080/01972240490423030
15. 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)*.
16. Michelle Ibanez and Daniel D. Suthers. 2014. Detection of Domestic Human Trafficking Indicators and Movement Trends Using Content Available on Open Internet Sources. *Proceedings of the 47th Hawaii International Conference on Systems Science (HICSS)*.
17. Carol Levine, Ruth Faden, Christine Grady, Dale Hammerschmidt, Lisa Eckenwiler, and Jeremy Sugarman. 2010. The Limitations of "Vulnerability" as a Protection for Human Research Participants. *The American Journal of Bioethics* 4, 3: 44–49.
18. Gregory McNeal. 2014. Facebook Manipulated User News Feeds to Create Emotional Responses. *Forbes*.
19. 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.

20. Jessica A. Pater, Oliver L. Haimson, Nazanin Andalibi, and Elizabeth D. Mynatt. 2016. "Hunger Hurts but Starving Works": Characterizing the Presentation of Eating Disorders Online. *Proceedings of the ACM Conference on Computer-Supported Cooperative Work & Social Computing*.

21. Martin Robbins. 2014. Does OKCupid need our consent? *The Guardian*.

22. Zeynep Tufekci. 2015. Algorithmic Harms Beyond Facebook and Google: Emergent Challenges of Computational Agency. *Colorado Technology Law Journal* 13: 203–218.

23. 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*.

24. Pamela Wisniewski, Heng Xu, Mary Beth Rosson, Daniel F. Perkins, and John M. Carroll. 2016. Dear Diary: Teens Reflect on Their Weekly Online Risk Experiences. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, ACM, 3919–3930. http://doi.org/10.1145/2858036.2858317

25. Michael Zimmer. 2010. "But the data is already public": On the ethics of research in Facebook. *Ethics and Information Technology* 12, 4: 313–325. http://doi.org/10.1007/s10676-010-9227-5

26. Michael Zimmer. 2016. OkCupid Study Reveals the Perils of Big-Data Science. *WIRED*.

27. Facebook. *ACM SIGCHI Research Ethics*. https://www.facebook.com/groups/866225810172460/