Perspective

How scientists can take the lead in establishing ethical practices for social media research

Sherry Pagoto[1] and Camille Nebeker[2]

[1]Department of Allied Health Sciences, Institute for Collaboration on Health, Intervention, and Policy, University of Connecticut, Storrs, Connecticut, USA, and [2]Department of Family Medicine and Public Health, Center for Wireless and Population Health Systems, University of California, San Diego, San Diego, California, USA

Corresponding Author: Sherry Pagoto, PhD, UConn Center for mHealth and Social Media, UConn Institute for Collaboration in Health, Interventions, and Policy, Department of Allied Health Sciences, University of Connecticut, 2006 Hillside Rd, Unit 1248, Room 22, Storrs, CT 06268, USA (sherry.pagoto@uconn.edu)

ABSTRACT

Social media use has become ubiquitous in the United States, providing unprecedented opportunities for research. However, the rapidly evolving research landscape has far outpaced federal regulations for the protection of human subjects. Recent highly publicized scandals have raised legitimate concerns in the media about how social media data are being used. These circumstances combined with the absence of ethical standards puts even the best intentioned scientists at risk of possible research misconduct. The scientific community may need to lead the charge in insuring the ethical use of social media data in scientific research. We propose 6 steps the scientific community can take to lead this charge. We underscore the important role of funding agencies and universities to create the necessary ethics infrastructure to allow social media research to flourish in a way that is pro-technology, pro-science, and most importantly, pro-humanity.

Key words: ethics, Twitter, Facebook, social media

The data of nearly 90 million Facebook users were compromised when an academic researcher gave Cambridge Analytica access to data he obtained from Facebook.[1] This was not the first time that scientists misused social media data. In 2014, serious public concerns were raised when researchers from Facebook and Cornell conducted potentially sensitive experiments with Facebook users without informed consent.[2] Then, in 2016, Danish researchers published data from 70 000 OkCupid users, revealing private information including age, gender, and sexual preferences.[3] Each breach has implications for the public's trust of science and technology, which could eventually threaten future innovation. In this piece, we discuss the problems with the current regulatory environment, argue that scientists should take a more active role in developing ethical standards for social media research, and describe 6 ways that scientists can do so.

Research involving social media platforms, including Facebook, Twitter, and Instagram, is emerging rapidly in a fairly unregulated landscape, which may put even the most scrupulous scientists at risk. Technology companies are under no obligation to follow federal research ethics regulations unless they are receiving federal funding for the research. Some companies outsource ethics review when it is required, others develop internal review processes, and yet others have no review procedures in place.

Academia may, on the face, seem more regulated, and consequently its research more "ethical," given that it operates under federal regulations, but this would only be the case to the extent that federal regulations are current. In fact, federal regulations developed to protect human research participants along with the ethical principles described in the Belmont Report were devised well before the internet existed, let alone Facebook. Currently, federal guidelines do not specifically address the ethical use of social media data in research. The current U.S. administration has put a delay on the effective date of the most recent regulatory revisions; however, these do

not explicitly speak to many big data sources.[4] This lack of guidance and regulations leaves many scientists and institutional review boards (IRBs) to navigate the ethical, legal, and social implications on their own. A recent study revealed that scientists who sit on university institutional review boards (IRBs) are not confident that they can keep pace with rapidly changing technologies used in research.[5]

Like scientists, the public may not have a clear understanding of how social media data can be used in research. A recent study revealed that the majority of social media users surveyed did not have an accurate understanding on how their social media data can be used.[6] In that study, 60% of Twitter users surveyed were unaware that publicly available tweets can be used for research and 65% felt researchers should not be able to use public tweets without user permission. Twitter's privacy policy states that tweets may be used for research purposes; however, privacy policies are not written in a way that guarantees they will be read or understood.[7] Researchers recently found that privacy policies for mobile apps were on average 3500 words and written just over the 12th-grade level (the average reading level for adults is grade 7-8).[8] This is particularly challenging with platforms in which privacy settings can vary across different types of user activities and in which the line between public and private content is blurry given the degree of user control over audience size and membership (eg, posts in a private Facebook group that has 40 000 members and a moderator with complete control over who joins).

To the extent that social media platforms are used in human subjects research (ie, requiring IRB approval), the onus is on the scientist to ensure the participant understands the platform's privacy policy. When scientists study publicly available social media data, which do not require IRB approval, they should consider that users may not only be unaware their data are being used in research but, upon discovery, may feel just as angry and betrayed as they would if the data were private. To the extent users become aggravated and lose trust in the platform, social media companies' trust in academic researchers could erode. A strong negative public response from users about researcher activities could put pressure on platforms to block researchers from easily accessing data. Ultimately, researchers would greatly benefit from leading the way in building trust and open dialogue with industry and the public about the ethical use of social media data in research.

We propose 6 ways that the scientific community can lead the charge in setting ethical standards for social media research. First, the scientific community must collaborate across disciplines and sectors (eg, academia, industry) to establish conventions in a way that is responsive to rapidly changing technologies. Disciplines represented should reflect the full breadth of disciplines conducting social media research (eg, medicine, behavioral science, engineering, communications) as well as experts in research ethics, science policy, and law. A sustainable and public open-source site supported through a public-private partnership, as has been recently proposed to develop and maintain standards for health apps,[9] could be useful.

Second, "technology ethics boards" could be convened in universities and other research organizations to educate and advise scientists, research participants, IRBs, and the public. Boards could be comprised of individuals with expertise in the technology as well as those versed in the ethical, legal, and social implications of data use. Boards could be charged with devising institutional guidelines, producing informational materials for research participants and the community, developing a curriculum in technology ethics for students, and consulting with researchers on grant applications, study protocols, and manuscripts.

Third, scientists could develop coursework on tech ethics that could then be provided to students across any departments in the university where social media research is being conducted. Social media research does not reside in 1 academic silo, which means training in broad, cross-cutting issues like ethics may be overlooked or unavailable. Coursework of this type could be created for both undergraduate and graduate students so that both digital natives and future digital scientists develop an improved understanding of the ethical, legal and social implications of social media research. Example coursework in pioneering departments include digital anthropology, ethical and social implications of data, technology ethics, digital media ethics, and ethical issues and technology design, among others.[10] Universities that convene a technology ethics board, as per the second recommendation, could charge these boards with guiding the content of such coursework and quality assurance.

Fourth, IRBs must have access to training on the responsible conduct of social media research to develop the necessary expertise to review it. If this expertise is not available locally, professional associations like Public Responsibility in Medicine and Research may be a solution for connecting institutions with tech-ethics expertise. Likewise, the Connected and Open Research Ethics initiative is a new resource for researchers and IRBs facing the new challenges introduced by research using social media platforms.[11] Universities could also prioritize these skills in faculty hires across the wide range of departments in which this is relevant (eg, social science, computer science, engineering, law, public health) to ensure that IRBs have access to a critical mass of experts on campus.

Fifth, research funders should ensure that scientific review panels include expertise in technology ethics when grants describing social media research are being reviewed. Scientific review officers might assume that any researcher with experience using social media data understands the ethical implications, but this assumption could result in less-than-ideal practices proliferating. Reviewers that have some level of documented expertise on technology ethics should be required (eg, coursework, certifications). In reviewing the human subjects sections of grant proposals, reviewers should also be advised to comment specifically on how ethics for handling social media data was addressed in the application.

Sixth, the communication of both established and emerging ethical standards to the public and all relevant stakeholders is necessary for not only public education, but also transparency. Scientists can accomplish this by publishing in public facing outlets, by developing relationships with journalists who write about technology research in popular outlets (eg, Wired), or by working with citizen advocacy or public policy groups who address technology ethics. University communications offices typically provide training in science communication and can facilitate connections between scientists and journalists. While publications in public facing outlets are not considered "peer-reviewed publications," they can certainly be documented as productivity on curriculum vitae as well as cited as examples of scientific impact in promotion and tenure review, especially given that public outlets have far larger readerships than the typical academic journal.

These solutions are starting to take shape, albeit in an uncoordinated manner, and institutional incentives are lacking. In 2015, the Robert Wood Johnson Foundation funded scientists at the University of California San Diego to develop the Connected and Open Research Ethics initiative,[12] which has created a global "tech-ethics" learning community of more than 600 researchers, ethicists, technologists, and policy experts who are collectively developing ethical practices to guide digital health research. In 2017, the National

Science Foundation funded the PERVADE (Pervasive Data Ethics for Computational Research) team to conduct research to inform ethical standards for social media and other big data research.[13] Likewise, the Data for Good Exchange recently called for data scientists to develop an ethical oath similar to the Hippocratic Oath.[14] Researchers in the United Kingdom and Canada are also making progress and have developed guidelines to foster ethical social media research.[15,16] These moves are in the right direction but more pervasive and coordinated efforts are needed.

The next step is to coalesce efforts across scientific sectors and establish a community commons for standard setting and accountability. Stakeholders should include scientists and technologists, organizations that employ them, and funding agencies (eg, government, industry, nonprofit or private). Moreover, we need to involve policymakers to champion updates to federal regulations as well as journal editors who can elevate awareness for the need to address the ethical, legal, and social implications of this research prior to publication. Most importantly, we need to engage the public to ensure we have standards that respect privacy preferences and that adequately guide both informed consent and data management practices. These stakeholders must work together to establish the agenda and provide the resources for scientists to build an appropriate infrastructure that supports ethical and responsible practices for a changing technology-enabled research ecosystem.

The dizzying pace of technology has produced exciting innovations, but this cannot come at the expense of due diligence. Public trust is at stake as most recently evidenced by the trending hashtag #deleteFacebook in the wake of the Cambridge Analytica event. To the extent that data breaches continue to occur, public safety may be at risk in ways we are only beginning to understand. Scientists are in a unique position to lead the development of a responsive ethical infrastructure and inform stakeholders about how to innovate in a way that is pro-ethics, pro-tech, and pro-humanity.

AUTHOR CONTRIBUTIONS

Both authors contributed equally to this manuscript, including the concept, drafting, and editing.

DATA AND MATERIALS AVAILABILITY

None.

REFERENCES

1. Kozlowska A. The Cambridge Analytica scandal affected nearly 40 million more people than we thought. Quartz. 2018. Available at: https://qz.com/1245049/the-cambridge-analytica-scandal-affected-87-million-people-facebook-says/w. Accessed July 23, 2018.
2. Kramer AD, Guillory JE, Hancock JT. Experimental evidence of massive-scale emotional contagion through social networks. *Proc Natl Acad Sci U S A* 2014; 111 (24): 8788–90.
3. Hackett R. Researchers caused an uproar by publishing data from 70,000 OkCupid users. Fortune. 2016. Available at: http://fortune.com/2016/05/18/okcupid-data-research/. Accessed March 30, 2018.
4. Office for Human Research Protections. HHS and 15 Other Federal Departments and Agencies Announce an Interim Final Rule That Delays Both the Effective Date and General Compliance Date of the Revisions to the Federal Policy for the Protection of Human Subjects to July 19, 2018. January 17. Available at: https://www.hhs.gov/ohrp/interim-final-rule-common-rule.html. Accessed October 16, 2018.
5. Nebeker C, Harlow J, Espinoza Giacinto R, *et al*. Ethical and regulatory challenges of research using pervasive sensing and other emerging technologies: IRB perspectives. *AJOB Empir Bioeth* 2017; 8 (4): 266–76.
6. Casey F, Nicholas P. "Participant" perceptions of Twitter research ethics. *Soc Media Soc* 2018; 4 (1): 1–14.
7. Obar J, Oeldorf-Hirsch A, eds. The biggest lie on the internet: ignoring the privacy policies and terms of service policies of social networking services (June 2018). In: TPRC 44: The 44th Research Conference on Communication, Information and Internet Policy. Taylor and Francis; 2016.
8. Das G, Cheung C, Nebeker C, *et al*. Privacy policies for apps targeted toward youth: descriptive analysis of readability. *JMIR mHealth Uhealth* 2018; 6 (1): e3.
9. Bates DW, Landman A, Levine DM. Health apps and health policy: what is needed? *JAMA* 2018; 320 (19): 1975–6.
10. Fiesler C. Tech ethics curricula: a collection of syllabi. July 5, 2018. Available at: https://medium.com/@cfiesler/tech-ethics-curricula-a-collection-of-syllabi-3eedfb76be18. Accessed October 16, 2018.
11. Connected & Open Research Ethics (CORE). What is CORE? Available at: https://thecore.ucsd.edu/. Accessed July 23, 2018.
12. Torous J, Nebeker C. Navigating ethics in the digital age: introducing Connected and Open Research Ethics (CORE), a tool for researchers and institutional review boards. *J Med Internet Res* 2017; 19 (2): e38.
13. PERVADE Pervasive Data Ethics for Computational Research. The project. Available at: https://pervade.umd.edu/about/general/. Accessed March 30, 2018.
14. Simonite T. Should data scientists adhere to a hippocratic oath?: Wired. February 2, 2018. Available at: https://www.wired.com/story/should-data-scientists-adhere-to-a-hippocratic-oath/. Accessed March 30, 2018.
15. Townsend L, Wallace C. Social media research: a guide to ethics. Available at: https://www.gla.ac.uk/media/media_487729_en.pdf. Accessed July 23, 2018.
16. Warfield K. Selfies, sexting, and NSFW content: difficult images in digital research. Available at: https://www.mta.ca/uploadedFiles/Community/Research_and_creative/Research_Office/Research_Ethics_Board/Education_and_training/KWarfield_REB_Workshop1.pdf. Accessed July 23, 2018.