Ethical Research Protocols for Social Media Health Research

Adrian Benton
Center for Language and Speech Processing
Johns Hopkins University
adrian@cs.jhu.edu

Glen Coppersmith
Qntfy
glen@qntfy.com

Mark Dredze
Human Language Technology Center of Excellence
Johns Hopkins University
mdredze@jhu.edu

Abstract

Social media have transformed data-driven research in political science, the social sciences, health, and medicine. Since health research often touches on sensitive topics that relate to ethics of treatment and patient privacy, similar ethical considerations should be acknowledged when using social media data in health research. While much has been said regarding the ethical considerations of social media research, health research leads to an additional set of concerns. We provide practical suggestions in the form of guidelines for researchers working with social media data in health research. These guidelines can inform an IRB proposal for researchers new to social media health research.

1 Introduction

Widely available social media data – including Twitter, Facebook, discussion forums and other platforms – have emerged as grounds for data-driven research in several disciplines, such as political science (Tumasjan et al., 2011), public health (Paul and Dredze, 2011), economics (Bollen et al., 2011), and the social sciences in general (Schwartz et al., 2013). Researchers have access to massive corpora of online conversations about a range of topics as never before. What once required painstaking data collection or controlled experiments, can now be quickly collected and analyzed with computational tools. The impact of such data is especially significant in health and medicine, where advances in our understanding of disease transmission, medical decision making, human behavior and public perceptions of health topics could directly lead to saving lives and improving quality of life.

Health research often touches on sensitive topics that relate to ethics of treatment and patient privacy. Based on decades of research experience and public debate, the research community has developed an extensive set of guidelines surrounding ethical practices that guide modern research programs. These guidelines focus on human subjects research, which involves research with data from living individuals. The core principles of human subjects research were codified in the Belmont Report (National Commission, 1978), which serves as the essential reference for institutional review boards (IRBs) in the United States. IRB guidelines include a range of exemptions from full review for research protocols that consider certain types of data or populations. For example, research projects that rely on online data sources may be exempt since the data are publicly available. Historically, public data exemptions included previously compiled databases containing human subject data that have entered the public domain. The recent proposal to modernize the U.S. Common Rule for the Protection of Human Subjects acknowledges the widespread use of social media for health research, but does little to clarify the ethical obligations of social media health researchers, generally reducing oversight necessary for research placed under expedited review (National Research Council, 2014).

> A more participatory research model is emerging in social, behavioral, and biomedical research, one in which potential research subjects and communities express their views about the value and acceptability of research studies. This participatory model has emerged alongside a broader trend in American

Proceedings of the First Workshop on Ethics in Natural Language Processing, pages 94–102,
Valencia, Spain, April 4th, 2017.

> society, facilitated by the widespread use of social media, in which Americans are increasingly sharing identifiable personal information and expect to be involved in decisions about how to further share the personal information, including health-related information that they have voluntarily chosen to provide.

In general, it provides a more permissive definition of what qualifies as exempt research. It suggests exempting observational studies of publicly available data where appropriate measures are taken to secure sensitive data, and demonstrably benign behavioral intervention studies.

The intersection of these ethics traditions and social media research pose new challenges for the formulation of research protocols. These challenges are further complicated by the discipline of the researchers conducting these studies. Health research is typically conducted by researchers with training in medical topics, who have an understanding of human subjects research protocols and issues regarding IRBs. In contrast, social media research may be conducted by computer scientists and engineers, disciplines that are typically unaccustomed to these guidelines (Conway, 2014).

Although this dichotomy is not absolute, many researchers are still unclear on what measures are required by an IRB before analyzing social media data for health research. Conversations by the authors with colleagues have revealed a wide range of "standard practice" from IRBs at different institutions. In fact, the (excellent) anonymous reviews of this paper stated conflicting perceptions on this point. One claimed that online data did not necessarily qualify for an exemption if account handles were included, whereas another reviewer states that health research solely on public social media data did not constitute human subjects research.

The meeting of non-traditional health researchers, health topics, and non-traditional data sets has led to questions regarding ethical and privacy concerns of such research. This document is meant to serve as a guide for researchers who are unfamiliar with health-related human subjects research and want to craft a research proposal that complies with requirements of most IRBs or ethics committees.

How are we to apply the ethical principles of human subjects research to projects that analyze publicly available social media posts? What protections or restrictions apply to the billions of Twitter posts publicly available and accessible by anyone in the world? Are tweets that contain personal information – including information about the author or individuals known to the author – subject to the same exemptions from full IRB review that have traditionally been granted to public data sources? Are corpora that include public data from millions of individuals subject to the same informed consent requirements of traditional human subjects research? Should researchers produce annotations on top of these datasets and share them publicly with the research community? The answers to these and other questions influence the design of research protocols regarding social media data.

Ethical issues surrounding social media research have been discussed in numerous papers, a survey of which can be found in McKee (2013) and Conway (2014). Additionally, Mikal et al. (2016) used focus groups to understand the perceived ethics of using social media data for mental health research. Our goal in this paper is complementary to these ethics surveys: we want to provide practical guidance for researchers working with social media data in human subjects research. We, ourselves, are not ethicists; we are practitioners who have spent time considering practical suggestions in consultation with experts in ethics and privacy. These guidelines encapsulate our experience implementing privacy and ethical ideals and principles.

These guidelines are not meant as a fixed set of standards, rather they are a starting point for researchers who want to ensure compliance with ethical and privacy guidelines, and they can be included with an IRB application as a reflection of current best practices. We intend these to be a skeleton upon which formal research protocols can be developed, and precautions when working with these data. Readers will also note the wide range of suggestions we provide, which reflects the wide range of research and associated risk. Finally, we include software packages to support implementation of some of these guidelines.

For each guideline, we reference relevant discussions in the literature and give examples of how these guidelines have been applied. We hope that this serves as a first step towards a robust discus-

sion of ethical guidelines for health-related social media research.

2 Discussion

The start of each research study includes asking core questions about the benefits and risks of the proposed research. What is the potential good this particular application allows? What is the potential harm it may cause and how can the harm be mitigated? Is there another feasible route to the good with less potential harm?

Answers to these questions provide a framework within which we can decide which avenues of research should be pursued. Virtually all technology is dual-use: it can be used for good or ill. The existence of an ill use does not mean that the technology should not be developed, nor does the existence of a good mean that it should.

To focus our discussion on the pragmatic, we will use mental health research as a concrete use case. A research community has grown around using social media data to assess and understand mental health (Resnik et al., 2013; Schwartz et al., 2013; Preotiuc-Pietro et al., 2015; Coppersmith et al., 2015a; De Choudhury et al., 2016). Our discussion on the benefits and risks of such research is sharpened by the discrimination and stigma surrounding mental illness. The discrimination paired with potentially lethal outcomes put the risks and benefits of this type of research in stark relief – not sufficiently protecting users'/subjects' privacy, may exacerbate the challenge, discourage individuals from seeking treatment and erode public trust in researchers. Similarly, insufficient research results in a cost measured in human lives – in the United States, more than 40,000 die from suicide each year (Curtin et al., 2016). Mental health may be an extreme case for the gravity of these choices, but similar risk and benefits are present in many other health research domains. Clearly identifying the risks and the potential reward helps to inform the stance and guidelines one should adopt.

We found it helpful to enumerate facts and observations that inform each research protocol decision:

- We want to make a positive impact upon society, and one significant contribution we may provide is to better understand mental illness. Specifically, we want to learn information that will aid mental health diagnosis and help those challenged by mental illness. Thus, the driving force behind this research is to prevent suffering from mental illness.

- Intervention has great potential for good and for harm. Naturally, we would like to help those around us that are suffering, but that does not mean that we are properly equipped to do so. Interventions enacted at a time of emotional crisis amplify the risks and benefits. The approach we have taken in previous studies was to observe and understand mental illness, not to intervene. This is likely true for many computer and data science research endeavors, but that does not absolve the consideration of interventions. Ultimately, if the proposed research is successful it will inform the way that medicine is practiced, and thus will directly or indirectly have an effect on interventions.

- Machine learning algorithms do not learn perfectly predictive models. Errors and misclassifications will be made, and this should be accounted for by the researcher. Even less clearly error-prone systems, such as databases for sensitive patient data, are liable to being compromised.

- Social media platforms, like Twitter, are often public broadcast media. Nevertheless, much has been written about the perception that users do not necessarily treat social media as a purely public space (McKee, 2013). Mikal et al. (2016) found that many Twitter users in focus groups do have a skewed expectation of privacy, even in an explicitly public platform like Twitter, driven by *"users' (1) failure to understand data permanence, (2) failure to understand data reach, and (3) failure to understand the big data computational tools that can be used to analyze posts"*.

Our guidelines emerge from these tenets and our experience with mental health research on social media, where we try to strike a balance between enabling important research with the concerns of risk to the privacy of the target population. We encourage all researchers to frame their own research tenets first to establish guiding principles as to how research should proceed.

3 Guidelines

In contrast to others (Neuhaus and Webmoor, 2012; McKee, 2013; Conway, 2014) who have offered broad ethical frameworks and high-level guidance in social media health research, we offer specific suggestions grounded in our own experience conducting health research with social media. At the same time, the risk of a study varies depending on the type of health annotations collected and whether the research is purely observational or not. Therefore, we do not provide hard rules, but different options given the risk associated with the study.

Researchers familiar with human subjects research may ask how our guidelines differ from those recommended for all such research, regardless of connections with social media data. While the main points are general to human subjects research, we describe how these issues specifically arise in the context of social media research, and provide relevant examples. Additionally, social media raises some specific concerns and suggestions described below, such as (1) concern of inadvertently compromising user privacy by linking data, even when all the linked datasets are public, (2) using alternatives to traditionally obtained informed consent, (3) additional steps to de-identify social media data before analysis and dissemination, and (4) care when attributing presenting information in public forums. Furthermore, our intended audience are readers unfamiliar with human subjects research guidelines, as opposed to seasoned researchers in this area.

3.1 Institutional Review Board

In the United States, all federally-funded *human subject* research must be approved by a committee of at least five persons, with at least one member from outside of the institution (Edgar and Rothman, 1995). This committee is the Institutional Review Board (IRB), and in practice, many American institutions require all performed research to be sanctioned by the IRB. Ethics committees serve a similar role as IRBs in European Union member states (European Parliament and Council of the European Union, 2001). These committees have different regulations, but typically make similar approval judgments as IRBs (Edwards et al., 2007).

Human subjects are any living individual about whom an investigator conducting research obtains "(1) Data through intervention or interaction with the individual, or (2) Identifiable private information" (US Department of HHS, 2009). Collecting posts, examining networks, or in any way observing the activity of people means that social media health research qualifies as human subjects research (O'Connor, 2013) and requires the review of an IRB. The distinction between social media research that involves human subjects and research that does not is nebulous, as the inclusion of individuals in research alone is insufficient. For example, research that requires the annotation of corpora for training models involves human annotators. But since the research does not study the actions of those annotators, the research does not involve human subjects. By contrast, if the goal of the research was to study *how* humans annotate data, such as to learn about how humans interpret language, then the research may constitute human subjects research. When in doubt, researchers should consult their appropriate IRB contact.

IRB review provides a series of exemption categories that exempt research protocols from a full review by the IRB. Exemption category 4 in section 46.101 (b) concerns public datasets (US Department of HHS, 2009):

> *Research involving the collection or study of existing data, documents, records, pathological specimens, or diagnostic specimens, if these sources are publicly available or if the information is recorded by the investigator in such a manner that subjects cannot be identified, directly or through identifiers linked to the subjects.*

Since these projects pose a minimal risk to subjects, they require minimal review. Since most social media projects rely on publicly available data, and do not include interventions or interactions with the population, they may qualify for IRB exempt status (Hudson and Bruckman, 2004). Such research still requires an application to the IRB, but with a substantially expedited and simplified review process. This is an important point: research that involves human subjects, even if it falls under an exemption, must obtain an exemption from the IRB. Research that does not involve human subjects need not obtain any approval from the IRB.

3.2 Informed Consent

Obtain informed consent when possible.

A fundamental tenant of human subjects research is to obtain informed consent from study participants. Research that analyzes public corpora that include millions of individuals cannot feasibly obtain informed consent from each individual (O'Connor, 2013). Therefore, the vast majority of research that analyzes collected social media posts cannot obtain such consent. Still, we advocate for informed consent where possible due to the central role of consent in human subjects research guidelines. In cases where researchers solicit data from users, such as private Facebook or Twitter messages, informed consent may be required (Celli et al., 2013). Be explicit about how subject data will be used, and how it will be stored and protected. OurDataHelps[1], which solicits data donations for mental health research, provides such information.

Even if you have not explicitly dealt with consent while collecting public subject data, attaching a "statement of responsibility" and description of how the data were compiled and are to be used will give you, the researcher, a measure of accountability (Neuhaus and Webmoor, 2012; Vayena et al., 2013). This statement of responsibility would be posted publicly on the research group's website, and contains a description of the type of data that are collected, how they are being protected, and the types of analyses that will be conducted using it. Users could explicitly choose to opt-out their data from the research by providing their account handle. An IRB or ethics committee may not explicitly request such a statement[2], but it serves to ensure trust in subjects who typically have no say in how their online data are used.

3.3 User Interventions

Research that involves user interventions may not qualify for an IRB exemption.

Research that starts by analyzing public data may subsequently lead to interacting with users or modifying user experience. For example, research may start with identifying public Twitter messages on a given topic, and then generating an interaction with the user of the message. The well known study of Kramer et al. (2014) manipulated Facebook users' news feeds to vary the emotional content and monitor how the feed influenced users' emotional states. This study raised particularly strong ethical reservations since informed consent agreements were never obtained, and was followed by an "Editorial Expression of Concern". While we cannot make definitive judgements as to what studies can receive IRB exemptions, interacting with users often comes with testing specific interventions, which typically require a full IRB review. In these cases, it is the responsibility of the researchers to work with the IRB to minimize risks to study subjects, and such risk minimization may qualify for expedited IRB review (McKee, 2013). In short, researchers should be careful not to conflate exemptions for public datasets with blanket permission for all social media research.

3.4 Protections for Sensitive Data

Develop appropriate protections for sensitive data.

Even publicly available data may include sensitive data that requires protection. For example, users may post sensitive information (e.g. diagnoses, personal attributes) that, while public, are still considered sensitive by the user. Furthermore, algorithms may infer latent attributes of users from publicly posted information that can be considered sensitive. This is often the case in mental health research, where algorithms identify users who may be challenged by a mental illness even when this diagnosis isn't explicitly mentioned by the user. Additionally, domain experts may manually label users for different medical conditions based on their public statements. These annotations, either manually identified or automatically extracted, may be considered sensitive user information even when derived from public data.

Proper protections for these data should be developed before the data are created. These may include:

1. Restrict access to sensitive data. This may include placing such data on a protected server, restricting access using OS level permissions, and encrypting the drives. This is common practice for medical record data.

[1] https://ourdatahelps.org

[2] Although some IRBs do require such a statement and the ability for users to opt-out of the study. See the University of Rochester guidelines for social media research: https://www.rochester.edu/ohsp/documents/ohsp/pdf/policiesAndGuidance/Guideline_for_Research_Using_Social_Media.pdf

2. Separate annotations from user data. The raw user data can be kept in one location, and the sensitive annotations in another. The two data files are linked by an anonymous ID so as not to rely on publicly identifiable user handles.

The extent to which researchers should rely on these and other data protections depends on the nature of the data. Some minimal protections, such as OS level permissions, are easy to implement and may be appropriate for a wide range of data types. For example, the dataset of users who self-identified as having a mental condition as compiled in Coppersmith et al. (2015a) was protected in this way during the *3rd Annual Frederick Jelinek Summer Workshop*. More extreme measures, such as the use of air-gapped servers – computers that are physically removed from external networks – may be appropriate when data is particularly sensitive and the risk of harm is great. Certainly in cases where public data (e.g. social media) is linked to private data (e.g. electronic medical records) greater restrictions may be appropriate to control data access (Padrez et al., 2015).

3.5 User Attribution

> De-identify data and messages in public presentations to minimize risk to users.

While messages posted publicly may be freely accessible to anyone, users may not intend for their posts to have such a broad audience. For example, on Twitter many users engage in public conversations with other users knowing that their messages are public, but do not expect a large audience to read their posts. Public users may be aware that their tweets can be read by anyone, but posted messages may still be intended for their small group of followers (Hudson and Bruckman, 2004; Quercia et al., 2011; Neuhaus and Webmoor, 2012; O'Connor, 2013; Kandias et al., 2013). The result is that while technically and legally public messages may be viewable by anyone, the author's intention and care with which they wrote the message may not reflect this reality. Therefore, we suggest that messages be de-identified or presented without attribution in public talks and papers unless it is necessary and appropriate to do otherwise. This is especially true when the users discuss sensitive topics, or are identified as having a stigmatized condition.

In practice, we suggest:

1. Remove usernames and profile pictures from papers and presentations where the tweet includes potentially sensitive information (McKee, 2013).

2. Paraphrase the original message. In cases where the post is particularly sensitive, the true author may be identifiable through text searches over the relevant platform. In these cases, paraphrase or modify the wording of the original message to preserve its meaning but obscure the author.

3. Use synthetic examples. In many cases it may be appropriate to create new message content in public presentations that reflects the type of content studied without using a real example. Be sure to inform your audience when the examples are artificial.

Not all cases require obfuscation of message authorship; in many situations it may be perfectly acceptable to show screen shots or verbatim quotes of real content with full attribution. When making these determinations, you should consider if your inclusion of content with attribution may bring unwanted attention to the user, demonstrate behavior the user may not want to highlight, or pose a non-negligible risk to the user. For example, showing an example of an un-anonymized tweet from someone with schizophrenia, or another stigmatized condition, can be much more damaging to them than posting a tweet from someone who smokes tobacco. While the content may be publicly available, you do not necessarily need to draw attention to it.

3.6 User De-identification in Analysis

> Remove the identity of a user or other sensitive personal information if it is not needed in your analysis.

It is good practice to remove usernames and other identifying fields when the inclusion of such information poses risk to the user. For example, in the 2015 CLPsych shared task, tweets were de-identified by removing references to usernames, URLs, and most metadata fields (Coppersmith et al., 2015b). Carefully removing such information can be a delicate process, so we encourage the use of existing software for this task: `https://github.com/qntfy/deidentify_twitter`. This tool is clearly

not a panacea for social media health researchers, and depending on the sensitivity of the data, more time-consuming de-identification measures will need to be taken. For example, before analyzing a collection of breast cancer message board posts, Benton et al. (2011) trained a model to de-identify several fields: named entities such as person names, locations, as well as phone numbers and addresses. When analyzing text data, perfect anonymization may be impossible to achieve, since a Google search can often retrieve the identity of a user given a single message they post.

3.7 Sharing Data

> Ensure that other researchers will respect ethical and privacy concerns.

We strongly encourage researchers to share datasets and annotations they have created so that others can replicate research findings and develop new uses for existing datasets. In many cases, there may be no risk to users in sharing data and such data should be freely shared. However, where there may be risk to users, data should not be shared blindly without concern for how it will be used.

First, if protective protocols of the kind described above were established for the data, new researchers who will use the data should agree to the same protocols. This agreement was implemented in the MIMIC-III hospital admissions database, by Johnson et al. (2016). Researchers are required to present a certificate of human subjects training before receiving access to a de-identified dataset of hospital admissions. Additionally, the new research team may need to obtain their own IRB approval before receiving a copy of the data.

Second, do not share sensitive or identifiable information if it is not required for the research. For example, if sensitive annotations were created for users, you may instead share an anonymized version of the corpus where features such as, for example, individual posts they made, are not shared. Otherwise, the original user handle may be recovered using a search for the message text. For NLP-centric projects where models are trained to predict sensitive annotations from text, this means that either opaque feature vectors should be shared (disallowing others from preprocessing the data differently), or the messages be replaced with de-identified tokens, allowing other researchers to use token frequency statistics as features, but not, for example, gazetteers or pre-trained word vectors as features in their models.

It is also important to refer to the social media platform terms of service before sharing datasets. For example, section F.2 of Twitter's Developer Policy restricts sharing to no more than 50,000 tweets and user information objects per downloader per day.[3]

3.8 Data Linkage Across Sites

> Be cautious about linking data across sites, even when all data are public.

While users may share data publicly on multiple platforms, they may not intend for combinations of data across platforms to be public (McKee, 2013). For example, a user may create a public persona on Twitter, and a less identifiable account on a mental health discussion forum. The discussions they have on this health forum should not be inadvertently linked to their Twitter account by an overzealous researcher, since it may "out" their condition to the Twitter community.

There have been several cases of identifying users in anonymized data based on linking data across sources. Douriez et al. (2016) describe how the New York City Taxi Dataset can be de-anonymized by collecting taxi location information from four popular intersections. Narayanan and Shmatikov (2008) showed that the identify of users in the anonymized Netflix challenge data can be revealed by mining the Internet Movie Database.

Combinations of public data can create new sensitivities and must be carefully evaluated on a case-by-case basis. In some cases, users may explicitly link accounts across platforms, such as including in a Twitter profile a link to a LinkedIn page or blog (Burger et al., 2011). Other times users may not make these links explicit, intentionally try to hide the connections, or the connections are inferred by the researcher, e.g. by similarity in user handles. These factors should be considered when conducting research that links users across multiple platforms. It goes without saying that linking public posts to private, sensitive fields (electronic health records) should be handled with the utmost care (Padrez et al., 2015).

[3] https://dev.twitter.com/overview/terms/agreement-and-policy

4 Conclusion

We have provided a series of ethical recommendations for health research using social media. These recommendations can serve as a guide for developing new research protocols, and researchers can decide on specific practices based on the issues raised in this paper. We hope that researchers new to the field find these guidelines useful to familiarize themselves with ethical issues.

References

Adrian Benton, Lyle Ungar, Shawndra Hill, Sean Hennessy, Jun Mao, Annie Chung, Charles E. Leonard, and John H. Holmes. 2011. Identifying potential adverse effects using the web: A new approach to medical hypothesis generation. *Journal of biomedical informatics*, 44(6):989–996.

Johan Bollen, Huina Mao, and Xiaojun Zeng. 2011. Twitter mood predicts the stock market. *Journal of computational science*, 2(1):1–8.

John D. Burger, John Henderson, George Kim, and Guido Zarrella. 2011. Discriminating gender on twitter. In *Empirical Methods in Natural Language Processing (EMNLP)*, pages 1301–1309.

Fabio Celli, Fabio Pianesi, David Stillwell, and Michal Kosinski. 2013. Workshop on computational personality recognition (shared task). In *Workshop on Computational Personality Recognition*.

Mike Conway. 2014. Ethical issues in using Twitter for public health surveillance and research: developing a taxonomy of ethical concepts from the research literature. *Journal of Medical Internet Research*, 16(12):e290.

Glen Coppersmith, Mark Dredze, Craig Harman, and Kristy Hollingshead. 2015a. From ADHD to SAD: analyzing the language of mental health on Twitter through self-reported diagnoses. In *NAACL Workshop on Computational Linguistics and Clinical Psychology*.

Glen Coppersmith, Mark Dredze, Craig Harman, Kristy Hollingshead, and Margaret Mitchell. 2015b. CLPsych 2015 shared task: Depression and ptsd on Twitter. In *Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, pages 31–39.

Sally C. Curtin, Margaret Warner, and Holly Hedegaard. 2016. Increase in suicide in the United States, 1999-2014. *NCHS data brief*, 241:1–8.

Munmun De Choudhury, Emre Kiciman, Mark Dredze, Glen Coppersmith, and Mrinal Kumar. 2016. Discovering shifts to suicidal ideation from mental health content in social media. In *Conference on Human Factors in Computing Systems (CHI)*, pages 2098–2110.

Marie Douriez, Harish Doraiswamy, Juliana Freire, and Cláudio T. Silva. 2016. Anonymizing nyc taxi data: Does it matter? In *IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, pages 140–148.

Harold Edgar and David J. Rothman. 1995. The institutional review board and beyond: Future challenges to the ethics of human experimentation. *The Milbank Quarterly*, 73(4):489–506.

Sarah J. L. Edwards, Tracey Stone, and Teresa Swift. 2007. Differences between research ethics committees. *International journal of technology assessment in health care*, 23(01):17–23.

European Parliament and Council of the European Union. 2001. Directive 2001/20/EC.

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.

Alistair E. W. Johnson, Tom J. Pollard, Lu Shen, Li-wei H. Lehman, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G. Mark. 2016. MIMIC-III, a freely accessible critical care database. *Scientific data*, 3.

Miltiadis Kandias, Konstantina Galbogini, Lilian Mitrou, and Dimitris Gritzalis. 2013. Insiders trapped in the mirror reveal themselves in social media. In *International Conference on Network and System Security*, pages 220–235.

Adam D. I. Kramer, Jamie E. Guillory, and Jeffrey T. Hancock. 2014. Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, 111(24):8788–8790.

Rebecca McKee. 2013. Ethical issues in using social media for health and health care research. *Health Policy*, 110(2):298–301.

Jude Mikal, Samantha Hurst, and Mike Conway. 2016. Ethical issues in using Twitter for population-level depression monitoring: a qualitative study. *BMC medical ethics*, 17(1):1.

Arvind Narayanan and Vitaly Shmatikov. 2008. Robust de-anonymization of large sparse datasets. In *IEEE Symposium on Security and Privacy*, pages 111–125.

National Commission. 1978. *The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research-the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research*. US Government Printing Office.

National Research Council. 2014. *Proposed revisions to the common rule for the protection of human subjects in the behavioral and social sciences*. National Academies Press.

Fabian Neuhaus and Timothy Webmoor. 2012. Agile ethics for massified research and visualization. *Information, Communication & Society*, 15(1):43–65.

Dan O'Connor. 2013. The apomediated world: regulating research when social media has changed research. *Journal of Law, Medicine, and Ethics*, 41(2):470–483.

Kevin A. Padrez, Lyle Ungar, H. Andrew Schwartz, Robert J. Smith, Shawndra Hill, Tadas Antanavicius, Dana M. Brown, Patrick Crutchley, David A. Asch, and Raina M. Merchant. 2015. Linking social media and medical record data: a study of adults presenting to an academic, urban emergency department. *Quality and Safety in Health Care*.

Michael J. Paul and Mark Dredze. 2011. You are what you tweet: Analyzing twitter for public health. In *International Conference on Weblogs and Social Media (ICWSM)*, pages 265–272.

Daniel Preotiuc-Pietro, Johannes Eichstaedt, Gregory Park, Maarten Sap, Laura Smith, Victoria Tobolsky, H. Andrew Schwartz, and Lyle Ungar. 2015. The role of personality, age and gender in tweeting about mental illnesses. In *Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*.

Daniele Quercia, Michal Kosinski, David Stillwell, and Jon Crowcroft. 2011. Our Twitter profiles, our selves: Predicting personality with Twitter. In *IEEE Third International Conference on Privacy, Security, Risk and Trust (PASSAT)*, pages 180–185.

Philip Resnik, Anderson Garron, and Rebecca Resnik. 2013. Using topic modeling to improve prediction of neuroticism and depression. In *Empirical Methods in Natural Language Processing (EMNLP)*, pages 1348–1353.

H. Andrew Schwartz, Johannes C. Eichstaedt, Margaret L. Kern, Lukasz Dziurzynski, Richard E. Lucas, Megha Agrawal, Gregory J. Park, Shrinidhi K. Lakshmikanth, Sneha Jha, Martin E. P. Seligman, and Lyle H. Ungar. 2013. Characterizing geographic variation in well-being using tweets. In *International Conference on Weblogs and Social Media (ICWSM)*.

Andranik Tumasjan, Timm O. Sprenger, Philipp G. Sandner, and Isabell M. Welpe. 2011. Election forecasts with twitter: How 140 characters reflect the political landscape. *Social science computer review*, 29(4):402–418.

US Department of HHS. 2009. Code of federal regulations. title 45. *Public Welfare CFR*, 46.

Effy Vayena, Anna Mastroianni, and Jeffrey Kahn. 2013. Caught in the web: informed consent for online health research. *Sci Transl Med*, 5(173):1–3.