Social Media Big Data as a “Sensor” of Mental Well-being: Ethical Insights and Challenges

Stevie Chancellor  
School of Interactive Computing  
Georgia Institute of Technology  
Atlanta, GA 30308  
schancellor3@gatech.edu

Vincent M. B. Silenzio  
Department of Psychiatry, Public Health Sciences, and Family Medicine  
University of Rochester  
Rochester, NY 14642  
Vincent_Silenzio@URMC.Rochester.edu

Eric D. Caine  
Department of Psychiatry  
University of Rochester  
Rochester, NY 14642  
Eric_Caine@URMC.Rochester.edu

Munmun De Choudhury  
School of Interactive Computing  
Georgia Institute of Technology  
Atlanta, GA 30308, USA  
munmund@gatech.edu

Abstract  
Social media has emerged as a rich sensor to infer psychological and mental health states, and discussions have emerged in the social computing community and beyond around the ethical challenges of this research. This position paper examines several ethical challenges that the authors have seen from their own work in the field in the past few years. Presented as a collaboration between social computing big data and psychiatry researchers, we discuss open questions around ethics that need to be addressed in order to continue pushing the boundaries of the greater social good that social media big data may bring to us.

Author Keywords  
big data; social computing; quantitative methods; ethics; mental health

ACM Classification Keywords  
H.5.m [Information interfaces and presentation (e.g., HCI)]: Miscellaneous

Introduction  
An emergent body of work in CSCW and HCI venues shows that social media platforms are a powerful source of information about an individual's health and well-being [10, 15, 5, 22]. In particular, research demonstrates how computational techniques may be developed to leverage individuals' linguistic, behavioral, and interactions cues to inform
one's psychological state, including forecasting future risk to stressful experiences [7]. Engagement in online mental health communities has also been linked to improved health and wellness [16] — these communities offer support, treatment advice, and assistance [19, 21]. At the same time, social media communities can also facilitate dangerous and harmful behaviors when individuals gather to support disordered thoughts and behavior patterns [4].

As the social computing and health communities are able to identify those in great distress and where and how they act around others, we will be faced with an increasing number of questions and ethical dilemmas as both responsible researchers and as law-abiding and concerned citizens. These issues also apply to researchers interested in studying at-risk, marginalized or hidden communities and also touch on the broader issues we face when collecting, analyzing, and interpreting big data to examine individuals.

This position paper will discuss the ethical challenges surrounding the intersection of big data, machine learning and inference techniques, and mental health research on social media platforms. First, we will discuss some of the background information on mental health and big data, ethical insights and challenges, and some preliminary solutions to these issues. These suggestions are by no means the end of the conversation, and we hope that through this paper and the workshop broadly, we can spur thoughtful consideration of issues that arise when social media big data and mental health research collide.

**Background**

*Inference of Psychological and Mental Health States*

Social computing research has shown that content and conversational patterns can be used to infer psychological states, well-being, and social support status. In an early work, De Choudhury et al. [8] analyzed how new mothers’ risk to postpartum depression may be detected from content posted to Facebook and Twitter. Other research includes utilizing social media content and interactions to identify conditions and symptoms related to cyberbullying and teenage distress [11], substance use [20, 18], and aspects of a person's mental state [10, 15, 5, 1]. Together, this body of work shows that the rich repository of social media data may allow for the discovery, tracking, and perhaps forecasting of mental health risk attributes longitudinally.

*Mental Health Online Communities*

Research has also examined communities of those suffering from mental illnesses. One community that has been researched is the pro-eating disorder community, a group of people that glorify eating disorders and promote disordered eating and exercise habits [12, 14]. In some extreme cases, individuals in these communities have been found to demonstrate pro self-mutilation and pro-suicide sentiments [13, 17]. To this end, research has examined whether social media big data may reveal insights that indicate one’s propensity to these risky behaviors, and if so, being able to quantify to what extent online communities can exacerbate or enhance recovery from such behavior and lifestyle choices [9, 18].

Together, the potential of this line of research is immense, as discussed in our prior work [6]. They can help create provisions for early awareness to thoughts that could damage one’s mental well-being, including therapeutic arrangements for psychotherapy and coping. Contrary to existing post-hoc approaches identifying the behavioral and cognitive markers of mental illness risk, these computational approaches open up opportunities of unobtrusive data collection to facilitate appropriate interventions and also minimize the hindsight bias sometimes induced by retrospective analyses of mental health concerns.
Ethical Concerns
In the light of this, the ethical concerns we present in this paper fall into two areas: (1) The first involves tracking and surveillance. These issues arise as methodological or data analysis questions while we conduct our research. (2) The second area touches on the implications of these kinds of research examinations. How much weight do we give to the inferences about individuals, derived from population-level social media big data? The second relates to whether it is ethical to be intrusive in the lives of people we encounter while studying social media.

Issues of Tracking and Surveillance
Public Data and Consent. In social computing research, public data is often the norm as a source of data. Do we have an obligation to exclude those who are underage from our analysis? What happens when the inferences we can make are more than the sum of their parts; for instance, inferring one's intrinsic psychological state based on otherwise seemingly benign and mundane data shared on social media about their day-to-day happenings?

When seeking help for mental illness, most individuals autonomously choose to seek treatment from a caregiver or clinician. Any research done on these populations must be approved by a review board to protect the privacy and rights of the participants. In the case of social media data, however, the data is public and easily accessible (e.g., Twitter and Instagram have large public data). Do we have an obligation to obtain consent from someone if we are attempting to infer if they have a mental illness? Are our obligations different if there are no interventions? Relatedly, is it ethical to interact with persons on social media in public settings for research without their consent?

Removed Content. Examining content that has been deleted or removed has emerged as an area of research interest; for mental health, our research [2] has shown that removed content often reveals cues of increased vulnerability and might be used toward triggering just-in-time interventions to mitigate mental health risk.

However, there are ethical issues concerning use of data that is no longer “visible” or accessible by third parties. Is it ethical to examine data that was once public but is no longer? Does this fall under the purview of review boards since the content is no longer public-facing? Are researchers obligated to ensure that their datasets are free of deleted data? Can we share this data that at one point was public but is no longer to promote transparency/reproducibility?

Interventions
Imminent Harm. When studying mental health communities, users will often voice dangerous ideas. In our research of pro-eating disorder communities, we saw that people use tags to describe their eating disorders and also use tags like #killme, #suicide, and #dying that indicate heightened levels of mental illness severity[3]. In these communities, it is not uncommon to read postings that suggest people are considering suicide. If a person were saying they would die by suicide in a public place (like a street corner), there is a much stronger ethical argument to immediately intervene. Do researchers have an obligation to interpose themselves when they see such declarations, or when they can infer potential self-injurious behaviors on a public social media platform?

Caregiver and Physician Responsibilities. Assuming the availability of both observed and predicted data, what responsibilities are implied for mental health caregivers at the individual level? For example, what responsibilities might parents or guardians bear regarding the data that is, or could be, available about their children? Another example would involve the responsibilities borne by physicians and other health care professionals.
Mental health information that is actively gathered would include direct and purposeful compilation regarding a specific individual. Examples of this might include direct visualization of a specific individual's data or examining predictive modeling results about a person. Passively collected information would include both observed or predicted data compiled for another or an \textit{a priori} nonspecific purpose, but which can be used to provide observed or predicted information about specific or identifiable individuals.

Information that has been gained either actively or passively imply different responsibilities. For example, a physician who discovers expressions of suicidal ideation by examining their patient's social media postings may be bound by the \textit{duty to treat}. Failure to act on such knowledge would be considered unethical; what constitutes an ethical 'act' of treatment would depend on other knowledge about the patient that is available to the physician, and what a usual and reasonable professional would do in that particular situation. As the level of expertise in the mental health area increases, the ethical duty likewise increases to act in accordance with what are considered best practices in the field. A psychologist or a psychiatrist would thus be expected to meet the standard of care based upon current scientific understandings of mental health concerns.

Information gathered passively, on the other hand, may not necessarily imply such a strong ethical responsibility or duty to treat, at least not until this information has been analyzed with regard to a specific, identified individual. Once this occurs, an ethical responsibility to act exists clearly for those observers who have existing relationships with the individuals being observed. But what of those who do not have such a relationship? In this case, the researcher or observer is more akin to a bystander. As such, the concept of \textit{duty to rescue} is useful to consider. This concept is more broad than the duty to treat, in that it also addresses to the duty of a bystander to come to the rescue of another party in peril. Unlike the concept of duty to treat, the duty to rescue has far more varied interpretations and standards from place to place and from culture to culture.

\textbf{Content Ban and Censorship.} One intervention style in use now is banning or censoring content perceived by the platform to be inappropriate for other users.

Whether content moderation is effective the spread of these ideas is up for debate. Some of our previous research looked at community impacts of banning several pro-eating disorder tags on Instagram. We found that users changed their tags to avoid those that were banned (from #thinspo to #thynsppoo). These new tags and their associated communities had higher participation, more support of pro-eating disorder ideas, and expressed more toxic, self-harm, and vulnerable content [4]. In this case, content moderation has been ineffective at decelerating pro-eating disorder behavior on Instagram.

Nevertheless banning content related to mental health touches on several ethical issues, including speech, censorship, and its overall effectiveness. Does banning content even work at stopping the spread of this information? Should we censor people on social media platforms? How do we apply expectations of free speech to socially contentious issues like mental illness?

\textbf{Conclusion}

In this workshop paper, we have touched on many ethical issues that relate to the intersections of mental health and social media research. We hope that these questions spur further conversations and awareness of these issues.
REFERENCES


