Algorithms, Oppression, and Mental Illness on Social Media

Jessica L. Feuston
Northwestern University
Evanston, IL 60208, USA
jes.feuston@u.northwestern.edu

ABSTRACT

My research explores how individuals with mental illness express themselves online and off. Through digital ethnography, including interviews with Instagram users and manual collection of public content on Instagram, I have holistically examined the experience of mental illness as expressed through social media. This user-centric approach reveals and addresses the limitations of computational techniques, which my dissertation work will address by combining qualitative methods with generative algorithms to explore new 'ways of seeing' mental illness. I will create new tools enabling users to generate representations from their own posts in support of creating new representations of mental illness, advancing algorithmic fairness, and confronting technological forms of oppression online.

CCS CONCEPTS • Human-centered computing~Human computer interaction (HCI)

KEYWORDS

Mental illness; social media; digital ethnography; generative algorithms; anti-oppressive design.

Table 1: Research Questions

RQ1: How do individuals express their subjective experience of living with mental illness online?

RQ2: What role do researchers, designers, and technological platforms play in the oppression and further marginalization of individuals with mental illness?

RQ3: How might we engage with individuals with mental illness to design anti-oppressive technology and create safe and supportive spaces for sharing about mental illness online?

1 INTRODUCTION

Mental illness is a complex and varied component of everyday life for a significant number of individuals. However, stigma against mental illness is prevalent in contemporary society and impacts how people with mental illness document and share their personal experiences. The increased societal use of Instagram and other social media platforms has created a rich space for enabling the sharing of content related to mental illness. For example, many individuals use Instagram to share personal experiences with mental illness, raise awareness, and seek community support and belonging [13]. Instagram presents scholars opportunities to understand the practices, attitudes, and everyday experiences of living with mental illness.

Prior work has addressed a broad range of topics, including self-disclosure [1], content moderation [6], and content classification and behavioral prediction [7]. Oftentimes, this research adopts a clinical stance toward mental illness, positioning those with mental illness as deviant or otherwise divergent from societal norms, and perpetuating power dynamics traditionally grounded in the patient-physician relationship. Findings from existing research are often used to inform technological solutions aimed at restricting certain expressive practices (i.e., platform-level content moderation practices that remove posts or user accounts) or predicting future user behaviors. Though these solutions are 'good' for some, they may be unintentionally oppressive and harmful for others.

In my research, I incorporate work examining the ethics of studying marginalized communities [2]. Despite the potential of social media and online communities for empowerment and activism [4], research suggests that marginalized communities face additional oppression and marginalization online [11]. In my dissertation work, I examine and address the ways in which social media platforms restrict expression and act as oppressive, controlling forces on individuals with mental illness.

2 RESEARCH APPROACH AND METHODS

My approach involves foregrounding the online experience of living with mental illness. I draw upon a variety of disciplines, including cultural studies, narrative research, and visual sociology, to critically analyze data collected during digital ethnography (e.g., interviews and Instagram content). Through qualitative examination of these data, I address the theoretical and cultural underpinnings implicated in how scholars of human-computer interaction study and design for mental illness. By calling upon social justice and design [10] and the anti-oppressive design framework [12], I also approach research and design for good "as that which strives to end one or more forms of oppression" [12], and attend to the ways in which marginalized communities negotiate socially oppressive forces online.

3 WORK TO DATE

My work has examined how individuals with mental illnesses [9], as well as individuals from other marginalized groups (e.g., individuals with invisible disabilities [8]), share and express aspects of their subjective experience online and off. The following studies address **RQ1** and **RQ2**.

3.1 Study 1: Social 'Recovery' after Brain Injury

Traumatic brain injury (TBI) can lead to chronic health conditions that persist across the lifespan and impact all facets of life. Recovery in this context implies a return to pre-injury levels of functioning, many of which are assessed through clinically-defined outcome measures reliant on life before injury as a baseline. Through interviews with survivors of TBI, as well as caretakers and other social contacts, I found that a comparative pre-injury baseline neglects the complex, dynamic process of *social re-emergence* [8]. Further, certain features of social media, such as Facebook's 'Memories'[1], while good-intentioned, propagate comparative practices between pre- and post-injury realities and identities that may be harmful or undesirable for survivors.

3.2 Study 2: Everyday Experiences with Mental Illness

Drawing on digital ethnography, I examined how individuals express mental illness on Instagram. I found that mental illness arises in the context of everyday life and that individuals must negotiate how they share their experience online due to the ways that platform features and other users exert control over content and modes of expression. These findings re-center the user by prioritizing their interactional experience online over population-level approaches that generalize findings at the expense of the individual. Technological designs (or 'solutions') to online expressions of mental illness require this nuanced understanding of how individuals express experiences with mental illness, react to the ways others interact with their content, and perceive the impact of certain social and platform interactions.

3.3 Study 3: Computational Limitations to Mental Illness

Drawing on the *coded gaze* [5], this study addresses theoretical assumptions underlying current computational practices in data collection and analysis [9]. My findings demonstrate that different aspects of mental illness (e.g., relapse, recovery) are difficult to distinguish for human analysts and algorithms. The advantage of human analysts, however, lies in examining how meaning emerges over time and is dependent on multiple interpretations of content. For example, posts do not necessarily reflect an objective record of reality or truth. A post depicting self-injury may indicate a user's reflection on a past experience, rather than a current relationship to self-harm. Automated content removal may be ineffective and dismissive of individual experience in these situations.

---

[1] Previously called 'On This Day'.

4 FUTURE DISSERTATION WORK

In future work, I will combine qualitative methods with generative algorithms to explore new 'ways of seeing' mental illness [3]. Studies 4 and 5 explore **RQ3**.

4.1 Study 4: New 'Ways of Seeing' Mental Illness

By using generative machine learning techniques, the coded gaze can be transformed into a mechanism for data exploration, rather than data prediction or classification. In this study, I will use generative text and image algorithms to create new Instagram posts from content (e.g., thousands of posts) previously collected during digital ethnography. Combining this approach with qualitative analysis, I will examine how these generated artifacts can reveal decisions that may be taken for granted or hidden in computational analysis and inform new ways of understanding mental illness.

4.2 Study 5: Personal Representations of Mental Illness

In this study, I will support individuals in creating new representations of themselves and their experiences with mental illness by inputting their own social media content into the same algorithms used in Study 4. Rather than using these novel self-representations as health metrics, this study facilitates the creation of new forms of self-expression to extend work on algorithmic transparency and to provide an opportunity to explore meaning-making and communicating the lived experience of mental illness to others.

5 CURRENT AND EXPECTED CONTRIBUTIONS

My work makes two primary contributions. First, my dissertation research contributes to new ways of representing mental illness for both researchers and users. Artifacts generated by individuals with mental illness can provide new insights that inform future work, technology design, and community participation. Second, my work contributes to algorithmic fairness and transparency by exploring how generative algorithms can reveal biases and decisions in computational techniques. By aligning notions of 'good' with anti-oppressive design, my work invites conversations with researchers and designers regarding how technology platforms can better provide safe and supportive spaces for those with mental illness.

ACKNOWLEDGMENTS

REFERENCES

[1]   Nazanin Andalibi, Pinar Ozturk, and Andrea Forte. 2017. Sensitive Self-disclosures, Responses, and Social Support on Instagram: The Case of #Depression. CSCW (2017).

[2]   Karin Barron. 1999. Ethics in qualitative social research on marginalized groups. *Scandinavian Journal of Disability Research* 1, 1: 38–49.

[3]   John Berger. 2008. *Ways of Seeing*.

[4]   Lindsay Blackwell, Jean Hardy, Tawfiq Ammari, Tiffany Veinot, Cliff Lampe, and Sarita Schoenebeck. 2016. LGBT Parents and Social Media: Advocacy, Privacy, and Disclosure During Shifting Social Movements. CHI (2016).

[5]   Joy Adowaa Buolamwini. 2017. Gender shades: intersectional phenotypic and demographic evaluation of face datasets and gender classifiers.

[6]   Stevie Chancellor, Jessica Annette Pater, Trustin Clear, Eric Gilbert, and Munmun De Choudhury. 2016. #Thyghgapp: Instagram Content Moderation and Lexical Variation in Pro-Eating Disorder Communities. CSCW (2016).

[7]   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. CHI (2016).

[8]   Jessica L. Feuston, Charlotte G. Marshall-Fricker, and Anne Marie Piper. 2017. The Social Lives of Individuals with Traumatic Brain Injury. CHI (2017).

[9]   Jessica L. Feuston and Anne Marie Piper. 2018. Beyond the Coded Gaze: Analyzing Expression of Mental Health and Illness on Instagram. CSCW (2018).

[10] Sarah Fox, Jill Dimond, Lilly Irani, Tad Hirsch, Michael Muller, and Shaowen Bardzell. 2017. Social Justice and Design: Power and Oppression in Collaborative Systems. CSCW Companion (2017).

[11] Jillian C. York and Karen Gullo. 2018. Offline/Online Project Highlights How the Oppression Marginalized Communities Face in the Real World Follows Them Online. *Electronic Frontier Foundation*.

[12] Thomas Smyth and Jill Dimond. 2014. Anti-oppressive Design. *interactions* 21, 6: 68–71.

[13] Renwen Zhang, Jordan Eschler, and Madhu Reddy. 2018. Online Support Groups for Depression in China: Culturally Shaped Interactions and Motivations. CSCW (2018).