Reclaiming the Narrative: Digital Recovery, AI & Mitigating Harm in Social Media Workshop at ICWSM 2024

Rezvaneh Rezapour,[1] **Layla Bouzoubaa,** [1] **Steven R. Wilson** [2]

[1] Drexel University, PA, USA
[2] University of Michigan-Flint, MI, USA
{shadi.rezapour, layla.bouzoubaa}@drexel.edu, steverw@umich.edu

Abstract

This workshop explored the role of digital narratives on social media and their use or impact on recovery and healing processes. By integrating expertise from psychology, health, social sciences, and artificial intelligence, we aimed to discuss how narratives related to mental health, substance use disorder, and sexual abuse shape public discourse and personal experiences. Social media serves as a medium that can facilitate destigmatization and foster community, particularly in areas of mental health and sexual abuse, yet it also poses risks such as cyberbullying, the spread of misinformation, and exposure to harmful content. This workshop was structured around three key themes: **Digital Recovery Narratives**, **Harm Reduction & Healing in the Digital Age**, and **Shame & Intersectional Stigma**. We brought together experts to offer a comprehensive perspective on leveraging these narratives for a deeper understanding of individual and collective recovery paths, thereby transforming our approach to digital interaction and support. During a collaborative working session, workshop attendees contributed to a shared vision for the future of research in digital recovery narratives, and we include the outcome of this brainstorming work in this report.

Introduction

Social media has emerged as a platform for sharing and discussing deeply personal and often sensitive narratives (Yoo et al. 2023; Scott et al. 2023; Bouzoubaa, Young, and Rezapour 2023). It has transformed into an important space where topics that were once considered private or stigmatized are now openly discussed. This includes issues such as substance use disorder (SUD) (Raza et al. 2023; Bouzoubaa et al. 2024; Bouzoubaa and Rezapour 2024), various aspects of pregnancy and maternal health (Antoniak, Mimno, and Levy 2019), mental health challenges including suicide (Li et al. 2018; Shing et al. 2018), and personal experiences of sexual abuse (Andalibi et al. 2016).

The impact of social media in this context is multifaceted, serving as both a powerful support tool and a potential catalyst for harm. Among those with stigmatizing identities, these platforms have become a powerful tool for destigmatization and support and community-building. In areas like sexual abuse recovery, increased openness in digital spaces is shifting public awareness and empowering survivors (Madden and Alt 2021). These online discussions also enhance understanding and foster connection among those with similar experiences or challenges (McAuliffe et al. 2022). However, the same platforms that offer support can also pose significant risks. Within recovery communities, exposure to triggering content related to sensitive topics, invalidation of experiences, or the spread of harmful stereotypes can occur (Bouzoubaa, Aghakhani, and Rezapour 2024). The way narratives are framed is a crucial concern, as the presentation and interpretation of stories can shape public understanding and lead to potential biases (Mendelsohn, Budak, and Jurgens 2021; Huang and Wang 2022). Misinformation is another growing issue, where unverified or false information can spread rapidly, often with harmful consequences (Tran et al. 2020; Ganti et al. 2023). Moreover, the lack of effective moderation and anonymity on these platforms amplify the risk of harm, such as exposure to negative feedback, harassment, or triggering content in sensitive discussions (Chandrasekharan et al. 2022; Kornfield et al. 2018). This duality underscores the need for careful and nuanced approaches to digital spaces in the context of recovery and support. Bringing together computational communities with theory/domain experts is key to developing holistic strategies that effectively tackle the challenges and harness the possibilities offered by social media narratives.

Our workshop was designed to promote the examination of digital narratives on social media and their use or impact on recovery and healing processes. Putting emphasis on the necessity for interdisciplinary approaches, we brought perspectives and expertise from psychology, health, social sciences, and artificial intelligence together to understand and explain the nuances of recovery and healing as reflected in digital narratives. This collaborative endeavor sought not only to deepen our understanding of these processes but also to redefine our engagement with them, thereby making an important contribution to the field of digital narrative analysis. To have a more holistic understanding of these phenomena, this workshop concentrated on three interconnected themes:

(1) Digital Recovery Narratives: Digital platforms weave recovery stories, shaping individual journeys and public perceptions. This theme explored how social media platforms

amplify recovery narratives, dissected how language shapes stigma and fosters empathy, and investigated the interplay between platforms, identities, and societal impact on recovery journeys.

(2) Harm Reduction & Healing in the Digital Age: This theme delved into the dual role of digital media in causing and alleviating harm, examined the use of AI and digital tools for harm reduction and healing, and discussed the complexities and ethical concerns involved.

(3) Shame & Intersectional Stigma: Individuals recovering from traditionally stigmatized afflictions often report stigma as a major barrier to receiving the support they need to overcome it. We explored the crucial issues of shame and stigma, particularly in their intersectional manifestations, and how these factors affect recovery narratives and the efficacy of digital interventions in these spaces.

We invited researchers, practitioners, policymakers, and anyone interested in the intersection of digital media, recovery narratives, and technological interventions to contribute to our workshop.

Workshop Format

We hosted a half-day hybrid workshop featuring a selection of keynote presentations, invited talks, panel discussions, and an interactive session, fostering deep discussion and collaborative learning. One unique feature of our workshop was a panel discussion and an interactive session titled "The Future of Research in Digital Recovery." This session brought together various communities in a collaborative environment, where participants shared diverse perspectives on digital recovery and explored topics such as fostering connections and awareness on social media.

Keynote

Dr. Rohini Srihari, a professor in the Department of Computer Science and Engineering at the State University of New York, Buffalo, delivered the keynote at our workshop. With her research spanning natural language processing (NLP), AI, machine learning, and information retrieval, her presentation, titled "Mitigating the Spread of Misinformation Through Conversational AI," focused on advanced strategies to mitigate the accelerating spread of misinformation on social media (Sayantan, Souvik, and Rohini 2023; Das and Srihari 2024). Dr. Srihari pointed out the pressing need for scalable solutions that surpass the limitations of current moderation practices, such as merely flagging inappropriate content (Sougata and Rohini 2023). A key focus of Dr. Srihari's talk was the development of persuasive chatbots capable of assessing user intent and engaging in meaningful, prolonged conversations designed to reduce the dissemination of misinformation (Madani, Saha, and Srihari 2024). By incorporating argumentation theory, these chatbots can interact with users more effectively and influence their thinking. Dr. Srihari highlighted two notable case studies: one addressing the detection and mitigation of toxic speech and the other focusing on the early identification of phishing schemes. These examples illustrate how conversational AI can be a powerful tool in tackling a range of online threats, contributing to a more secure and informed digital landscape. For more information on her research, please see: https://www.acsu.buffalo.edu/~rohini/

Invited Talks and Contributions

To have a more in-depth discussion on online narratives and harm, we invited three researchers to share their research on online narratives, digital recovery, and harm reduction.

Dr. Sunny Rai from the University of Pennsylvania delivered an insightful presentation on a comparative analysis between Hollywood and Bollywood films (Rai et al. 2024). Dr. Rai's presentation titled "Shame, Social Norms, and Cultural Alignment of LLMs" focused on understanding how social norms differ across cultures and how this knowledge can be applied to develop culturally sensitive NLP systems. Dr. Rai introduced a novel, culture-agnostic approach to discovering social norms by analyzing moral emotions such as shame and pride in narratives. This approach was applied to identify normative expectations and extract corresponding social norms from a dataset of 5,400 Bollywood and Hollywood films, including over 10,000 social norms. The dataset was then validated by native speakers, which confirmed the dataset's ability to highlight cultural differences in social norms. Dr. Rai emphasized that these variations align with the broader cultural dichotomy between Bollywood and Hollywood, as the latter tends to emphasize shame associated with deviations from social roles and pride in family honor, while the former focuses on shaming poverty and incompetence and expressing pride in ethical behavior.

Shravika Mittal from Georgia Institute of Technology presented a thought-provoking talk titled "Understanding Framings of Stigma in Online Communities" (Mittal et al. 2024). Mittal's research focused on how discussions of Violence Against Women (VAW) in publicly accessible platforms, such as online news media, can shape public perceptions and influence organizations' approaches to this issue. The findings highlighted the potential harm caused by language that reinforces stigma around VAW, which can lead to unethical portrayals of survivors and the trivialization of the violence itself. The study presented examined the presence of stigmatized framings in news media and explored how these framings vary depending on media attributes like regionality, political leaning, veracity, and the latent communities within news sources.

Dr. Dylan Thomas Doyle from the University of Colorado Boulder presented a talk titled: "AI & the Afterlife: How the LLM Craze is Impacting Our Research Narratives Around Grief and Loss" (Brubaker et al. 2024; Doyle, Brahm, and Brubaker 2024). Dr. Doyle's talk focused on exploring future work and engaging in a more provocative, meta-level discussion on the evolving role of AI in shaping how we approach grief, loss, and legacy. Dr. Doyle emphasized that AI technologies are going to significantly alter existing practices and create new ones in areas such as end-of-life planning, remembrance, and legacy management. These changes are expected to have deep legal, economic, emotional, and religious implications. Dr. Doyle stressed that this pivotal moment of technological advancement offers a unique opportunity for the Human-Computer Interaction

(HCI) community to guide the conversation on this critical issue through approaches that are sensitive to values and centered on community needs. The talk highlighted the need for interdisciplinary collaboration, bringing together experts from fields such as HCI, AI, law, economics, and religious studies. Dr. Doyle's presentation underscored the importance of proactively shaping these emerging practices to ensure they are aligned with human values and societal needs.

The Future of Research in Digital Recovery

A central activity of our workshop was a collaborative mind-mapping exercise designed to explore the intersection of digital tools, particularly AI, and recovery processes. This interactive session aimed to foster a collective vision for the future of research in digital recovery narratives.

We used Mural [1], a digital collaboration tool, to facilitate the mental mapping exercise. The activity was structured around seven pre-determined themes: 1) Ethical considerations for GenAI use in online forums; 2) Implications of AI in recovery contexts; 3) Interpersonal relationships & connections; 4) GenAI for healing; 5) Specific narrative considerations; 6) Synthetic data; 7) Mental health. Participants were randomly assigned to these themes via color-coded sticky notes distributed at the beginning of the workshop, ensuring diverse perspectives within each thematic group. Online participants were moved to breakout rooms to work on the task in teams. The mind mapping activity spanned approximately 45 minutes, with 30 minutes dedicated to the collaborative expansion of the mind map and 15 minutes for team presentations. During the expansion phase, members of each team, which were organized by randomly assigned colors, collaborated to develop sub-themes within their designated area. In the presentation phase, each team had around 3 minutes to present their vision and key insights to the larger group.

The collaborative effort resulted in a significantly expanded mind map (shown in Figure 1), with "Ethical considerations" and "GenAI for Healing" emerging as the most extensively developed themes. This exercise broadened participants' perspectives on the potential applications of digital platforms and AI tools in supporting healing processes. Key discussions and insights that emerged included content moderation and trigger warnings, bereavement support, fostering connections, privacy risks, and SUD remission support. A notable discussion focused on using AI to support individuals in active remission from SUD by moderating potentially triggering content during internet browsing. The mind map activity proved to be a catalyst for identifying several potential research directions in the field of digital recovery narratives, highlighting the complex interplay between AI technologies, ethical considerations, and the deeply personal nature of recovery processes. Participants left the workshop with a better understanding of how digital platforms and AI tools can be leveraged to support healing, while also thinking about the potential risks and ethical considerations.

[1] https://shorturl.at/iaObH

Workshop Outcomes and Future Direction

This workshop resulted in significant outcomes that have set a clear direction for future research and collaboration in the field of digital recovery. Our workshop underscored the need for interdisciplinary approaches to address the complex interplay between technology and recovery. Moving forward, the insights gained from this workshop will serve as a foundation for continued exploration of how digital tools can be effectively and ethically integrated into recovery support, with a focus on addressing privacy risks, supporting individuals in remission, and enhancing content moderation. This workshop has laid the groundwork for ongoing research and collaboration aimed at harnessing the power of narratives, AI, and digital platforms to create more supportive and effective recovery environments.

Workshop Organization

The workshop was organized by:

- Rezvaneh (Shadi) Rezapour, an Assistant Professor in the Information Science Department at Drexel University, with research interests in NLP, computational social science, and NLP for Social Good.
- Layla Bouzoubaa, a PhD candidate in the Information Science Department at Drexel University, focused on NLP, harm reduction, and computational social science.
- Steven Wilson, an Assistant Professor in the College of Innovation & Technology at the University of Michigan-Flint who studies how NLP can be used for computational social science.
- Elizabeth D. Nesoff, an Assistant Professor, in the Biostatistics, Epidemiology & Informatics Department at the University of Pennsylvania Perelman School of Medicine, USA. She works on Disease Prevention and Health Promotion, Opioid Epidemic, and Social Determinants of Health.

Full information about the organizers and program committee can be found on the workshop's website at https://sites.google.com/view/reclaiming-the-narrative

References

Andalibi, N.; Haimson, O. L.; De Choudhury, M.; and Forte, A. 2016. Understanding social media disclosures of sexual abuse through the lenses of support seeking and anonymity. In *Proceedings of the 2016 CHI conference on human factors in computing systems*, 3906–3918.

Antoniak, M.; Mimno, D.; and Levy, K. 2019. Narrative paths and negotiation of power in birth stories. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW): 1–27.

Bouzoubaa, L.; Aghakhani, E.; and Rezapour, S. 2024. Words Matter: Reducing Stigma in Online Conversations about Substance Use with Large Language Models. *arXiv preprint arXiv:2408.07873*.

Bouzoubaa, L.; Aghakhani, E.; Song, M.; Trinh, Q.; and Rezapour, S. 2024. Decoding the Narratives: Analyzing Personal Drug Experiences Shared on Reddit. In Ku, L.-W.; Martins, A.; and Srikumar, V., eds., *Findings of the As-*

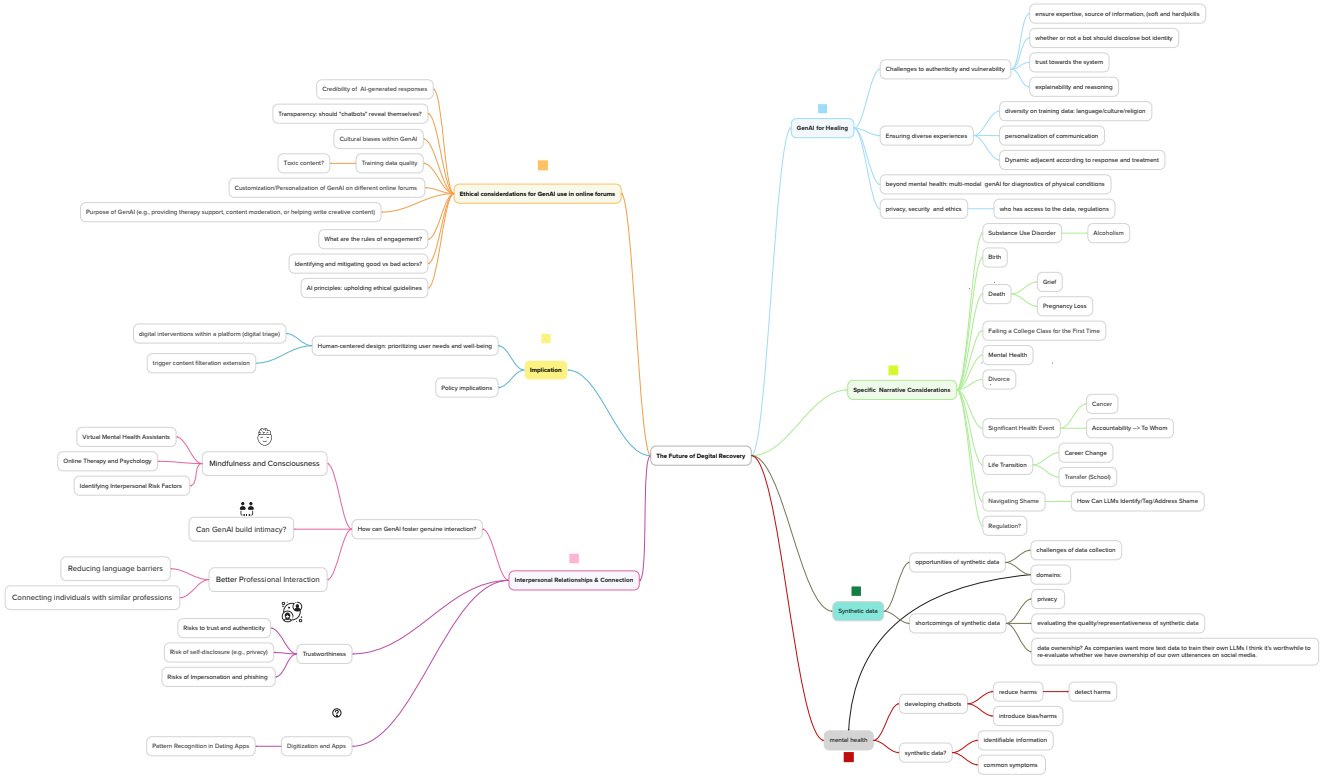

Figure 1: Result of collective visioning mind-map activity on the Future of Digital Recovery.

sociation for Computational Linguistics ACL 2024, 6131–6148. Bangkok, Thailand and virtual meeting: Association for Computational Linguistics.

Bouzoubaa, L.; and Rezapour, R. 2024. Euphoria's Hidden Voices: Examining Emotional Resonance and Shared Substance Use Experience of Viewers on Reddit. In *Proceedings of the Workshop on Data for the Wellbeing of Most Vulnerable at the 18th International AAAI Conference on Web and Social Media (ICWSM)*, 22.

Bouzoubaa, L.; Young, J.; and Rezapour, R. 2023. Exploring the landscape of drug communities on reddit: A network study. In *Proceedings of the International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 558–565.

Brubaker, J. R.; Morris, M. R.; Doyle, D. T.; Fiesler, C.; Gibbs, M.; and McGrenere, J. 2024. AI and the Afterlife. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, 1–5.

Chandrasekharan, E.; Jhaver, S.; Bruckman, A.; and Gilbert, E. 2022. Quarantined! Examining the effects of a community-wide moderation intervention on Reddit. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 29(4): 1–26.

Das, S.; and Srihari, R. K. 2024. Improving Dialog Safety using Socially Aware Contrastive Learning. In *SCI-CHAT-Workshop on Simulating Conversational Intelligence in Chat*, 4.

Doyle, D. T.; Brahm, C. B. R.; and Brubaker, J. R. 2024. " I hate you. I love you. I'm sorry. I miss you." Understanding Online Grief Expression Through Suicide Bereavement Letter-Writing Practices. *Proceedings of the ACM on Human-Computer Interaction*, 8(CSCW1): 1–27.

Ganti, A.; Hussein, E. A. H.; Wilson, S.; Ma, Z.; and Zhao, X. 2023. Narrative Style and the Spread of Health Misinformation on Twitter. In Bouamor, H.; Pino, J.; and Bali, K., eds., *Findings of the Association for Computational Linguistics: EMNLP 2023*, 4266–4282. Singapore: Association for Computational Linguistics.

Huang, Y.; and Wang, W. 2022. When a story contradicts: Correcting health misinformation on social media through different message formats and mechanisms. *Information, Communication & Society*, 25(8): 1192–1209.

Kornfield, R.; Sarma, P. K.; Shah, D. V.; McTavish, F.; Landucci, G.; Pe-Romashko, K.; and Gustafson, D. H. 2018. Detecting recovery problems just in time: application of automated linguistic analysis and supervised machine learning to an online substance abuse forum. *Journal of medical Internet research*, 20(6): e10136.

Li, A.; Huang, X.; Jiao, D.; O'Dea, B.; Zhu, T.; and Christensen, H. 2018. An analysis of stigma and suicide literacy in responses to suicides broadcast on social media. *Asia-Pacific Psychiatry*, 10(1): e12314.

Madani, N.; Saha, S.; and Srihari, R. 2024. Steering Conver-

sational Large Language Models for Long Emotional Support Conversations. *arXiv preprint arXiv:2402.10453*.

Madden, S.; and Alt, R. A. 2021. Know her name: Open dialogue on social media as a form of innovative justice. *Social Media+ Society*, 7(1): 2056305120984447.

McAuliffe, C.; Slemon, A.; Goodyear, T.; McGuinness, L.; Shaffer, E.; and Jenkins, E. K. 2022. Connectedness in the time of COVID-19: Reddit as a source of support for coping with suicidal thinking. *SSM-Qualitative research in health*, 2: 100062.

Mendelsohn, J.; Budak, C.; and Jurgens, D. 2021. Modeling Framing in Immigration Discourse on Social Media. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2219–2263.

Mittal, S.; Foriest, J. C.; Horne, B. D.; and De Choudhury, M. 2024. News Media and Violence Against Women: Understanding Framings of Stigma. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 18, 1068–1081.

Rai, S.; Zaveri, K. Z.; Havaldar, S.; Nema, S.; Ungar, L.; and Guntuku, S. C. 2024. A cross-cultural analysis of social norms in bollywood and hollywood movies. *arXiv preprint arXiv:2402.11333*.

Raza, S.; Schwartz, B.; Lakamana, S.; Ge, Y.; and Sarker, A. 2023. A framework for multi-faceted content analysis of social media chatter regarding non-medical use of prescription medications. *BMC digital health*, 1(1): 29.

Sayantan, P.; Souvik, D.; and Rohini, S. 2023. Mitigating Clickbait: An Approach to Spoiler Generation Using Multitask Learning. In *Proceedings of the 20th International Conference on Natural Language Processing (ICON)*, 486–490.

Scott, C. F.; Marcu, G.; Anderson, R. E.; Newman, M. W.; and Schoenebeck, S. 2023. Trauma-Informed Social Media: Towards Solutions for Reducing and Healing Online Harm. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–20.

Shing, H.-C.; Nair, S.; Zirikly, A.; Friedenberg, M.; Daumé III, H.; and Resnik, P. 2018. Expert, crowdsourced, and machine assessment of suicide risk via online postings. In *Proceedings of the fifth workshop on computational linguistics and clinical psychology: from keyboard to clinic*, 25–36.

Sougata, S.; and Rohini, S. 2023. Consolidating Strategies for Countering Hate Speech Using Persuasive Dialogues. In *Proceedings of the 20th International Conference on Natural Language Processing (ICON)*, 378–392.

Tran, T.; Valecha, R.; Rad, P.; and Rao, H. R. 2020. An investigation of misinformation harms related to social media during humanitarian crises. In *Secure Knowledge Management In Artificial Intelligence Era: 8th International Conference, SKM 2019, Goa, India, December 21–22, 2019, Proceedings 8*, 167–181. Springer.

Yoo, D. W.; Bhatnagar, A.; Ernala, S. K.; Ali, A.; Birnbaum, M. L.; Abowd, G. D.; and De Choudhury, M. 2023. Discussing Social Media During Psychotherapy Consultations: Patient Narratives and Privacy Implications. *Proceedings of the ACM on Human-Computer Interaction*, 7(CSCW1): 1–24.