How to Handle Online Risks? Discussing Content Curation and Moderation in Social Media

Donghee Yvette Wohn
New Jersey Institute of Technology
Newark, NJ 07102, USA
wohn@njit.edu

Munmun De Choudhury
School of Interactive Computing
Georgia Tech
Atlanta, GA 30332
munmund@gatech.edu

Casey Fiesler
Department of Information Science
University of Colorado Boulder
Boulder, CO
casey.fiesler@colorado.edu

J. Nathan Matias
MIT Media Lab
Cambridge, MA 02139
jnmatias@mit.edu

Libby Hemphill
Department of Humanities
Illinois Institute of Technology
Chicago, IL 60616
lhemphil@iit.edu

CHI'17 Extended Abstracts, May 06-11, 2017, Denver, CO, USA
ACM 978-1-4503-4656-6/17/05.
http://dx.doi.org/10.1145/3027063.3051141.

Abstract
Amidst proliferation of online negativity and harmful content such as fake news and harassment on social media, this panel will be an active discussion of the potential roles of various actors in sociotechnical systems in curating, moderating, and studying content. Should companies intervene? Why or why not, and if so, to what extent? What role do academics play in this process and how does it affect research processes? Our multidisciplinary panelists represent humanities, computer science, law, and media psychology. They will share perspectives based on their own research and interact with the audience to discuss varied perspectives around these central questions. A primary goal is to think about how, moving forward, these issues affect HCI research.

Author Keywords
moderation; content curation; social media; fake news; online harassment

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

Introduction

Is technology neutral? Should it be? This age-old question has become an increasingly pressing societal issue amidst concerns over potentially harmful phenomena such as the proliferation of fake news and online harassment. Companies that initially advocated free speech by positioning themselves as a vehicle for user-generated content rather than a producer of content themselves (e.g., Twitter) have lately been more reactive, building ways of removing, reporting, and detecting damaging online content. They therefore have an agenda-setting role in deciding what issues are important—a role that has been traditionally taken on by legacy media [10]. Changing responses to content moderation raise interesting questions – e.g., whether and how companies should intervene, what role(s) academics play in the mitigating online risk, and how moderation should unfold. We address these and similar questions in this panel by focusing on the roles of various actors in controlling online content.

The case for intervention

Massive Impact and Societal Wellbeing

A common argument made for intervention is general social wellbeing and the potential for positive impact. For example, should social media companies play a stronger role in content moderation and curation because of how much influence they have on people? Given that more Americans are relying on social media as their primary news source [5], the potential benefits and damages of "fake news" are huge, for instance. Numerous studies also show how intertwined our wellbeing is with our social media usage [4,5].

In some countries, the government may be making decisions about content moderation. Government-initiated intervention has been used both for censorship (e.g., anti-government sentiment, spread of liberal ideals in socialist countries) and as a measure to increase wellbeing (e.g., countering online harassment, stopping circulation of disturbing images). For example, when a video of a South Korean missionary being beheaded in Iraq in 2004 was distributed online, the Korean government forced local Internet service providers to ban access to any website that contained the video. The country also introduced a real-name policy in 2007 after a slew of suicides related to online harassment (it lasted 4 years). For both interventions, they cited the wellbeing of Korean citizens as their reason for intervening.

Adaptive Structuration

It is also important to keep in mind that once technology is released to the public, it will likely evolve (or descend) into a direction unforeseen by its creator. Adaptive structuration theory [4] posits that technology is usually designed with certain intentions by the maker but that societies may use them in unintended ways. Thus, technology influences society, but society also influences how the technology is used and potentially how they are developed. Therefore, a social system that has the most positive intentions could change into something that has negative effects. For example, location-based check-ins on social media were not intended as a device for burglars knowing one was away from home and the "Like" button on Facebook has evolved into interpretations beyond the literal "liking" of content [11]. Because of this cycle of change, technologies need to constantly adapt.

Harmful Content
Another argument toward intervention is simply that some content can be directly harmful to individuals—for example, cyberbullying or content that encourages self harm. With respect to platforms, some take a reactive stance on negative content (e.g., removing it after the fact or allowing users to flag inappropriate content), while others have taken proactive positions to try prevent the negative content from appearing in the first place. For example, multiplayer online gaming platforms (notorious for offensive language) have been working on this problem for some time. The game Overwatch will replace offensive (typed) statements with positive ones automatically, and Disney's Club Penguin, a virtual world primarily for young people, has very limited options for conversing with other players. Twitch, a livestreaming service where people post live videos that are accompanied with a real-time chat platform, enables users to have their own moderation settings to filter or prevent certain words or phrases.

There is also a potential role for researchers in this space. For instance, through browser plugins and sandboxes, researchers can experiment with a variety of system interventions such as automated moderation approaches, user block lists, flow controls, quarantined message queues, and content substitutions. We can also help understand the development of norms around acceptable and unacceptable content and behavior in various spaces and.

The case for non-intervention

Free speech and transparency
One common argument for non-intervention (by any actor) is the importance of free speech. In reality, there are very few online platforms that allow complete "freedom of speech." Some governments may be constrained from censorship by ideas of free speech, but (despite pervasive misunderstandings of what free speech entails) technology companies can "censor" content at will without giving a reason. Whether or not this is the *right* thing to do based on ethical obligations and/or the needs and wants of a user community, is more complex. Whether or not a company should take its users needs into consideration or forge their own path and philosophies, is also a complicated issue.

Also, if platforms moderate content (either algorithmically or manually), there have to be standards for what kind of content is inappropriate. Unfortunately, platform definitions of what constitutes behavior like "harassment" are incredibly vague and also highly inconsistent across platforms [9].

There is therefore an argument for making content moderation transparent both so that users are not unsettled by unexpected users of their content, and also to encourage norms around content that is acceptable and unacceptable. For example, if you just delete all of the fake news on Facebook and users never see it, then they won't be any closer to understanding what fake news is and why it's harmful.

Technology is not the only solution
It is increasingly common to jump to technology solutionism [8] for all of society's problems, particularly when the problem itself relates to technology. For example, after the 2016 presidential election, the media gave a lot of attention to the problem of "fake news" on Facebook, and in particular, to the idea that Facebook was obligated to solve this problem, typically through technological means.

But technology may not be the optimal solution, particularly if we ignore the underlying problem. Though it is good that technology developers are approaching hard social problems, it could be that a more nuanced approach is needed, beyond "solve it with technology." As Ethan Zuckerman [12] points out, sometimes you need to step back and consider whether you're solving the right problem. If you use technology to mitigate the symptoms of a deeper problem, then you may calcify it and make it more difficult to change. *However*, this critique does not mean that technology can't be part of a solution; but successful technology approaches to solving social problems also require changes to things like social norms and policy.

For example, is it more important to use technology to suppress fake news, or to examine the underlying social and psychological factors that lead to people creating it, believing it, and spreading it? Instagram bans searches on several pro-eating disorder hashtags, but this does not prevent people from uploading related content nor prevent users from finding ways to circumvent these restrictions [3]. In the case of cyberbullying, self harm, and other harmful content, are there ways beyond or in addition to content moderation, to help users who might be suffering or to help shift norms in a more positive direction? For example, Tumblr made the choice with community input, rather than banning harmful content, to present numbers for hotlines (e.g., suicide prevention) to users searching for certain hashtags [6].

Conclusion

We have discussed a number of perspectives on the issue of intervening into and controlling online content. As researchers, we are in a position to study these processes as they happen, or even to intervene ourselves. What do these decisions suggest for how we study and design these systems moving forward?

Panelists

Our panelists come represent multiple disciplines including the humanities, computer science, law, and media psychology, and have varied backgrounds as engineers, programmers, and journalists among others.

D. Yvette Wohn (moderator) is an assistant professor of informatics at New Jersey Institute of Technology where she heads the Social Interaction Lab. Co-founder of university social network sites Ewhaian (2001) and NJIT Buddy (2016), she studies how social systems facilitate social support and development of social capital. She has an ALM in journalism from Harvard University and a PhD in Media and Information Studies from Michigan State University.

Casey Fiesler is an assistant professor in the department of Information Science at the University of Colorado Boulder. Her research focuses on law, ethics, and social norms in online spaces. She holds a JD from Vanderbilt Law School and a PhD in Human-Centered Computing from Georgia Tech.

Libby Hemphill is an associate professor of communication and information studies at Illinois Institute of Technology who studies how people use social media in service of social change. One of her current related projects involves developing software tools to forecast imminent cyberbullying threats to facilitate community interventions. She earned her M.S. in Human-Computer Interaction and Ph.D. in Information from the University of Michigan.

Munmun De Choudhury is an assistant professor at the School of Interactive Computing, Georgia Tech. Munmun's research interests are in computational social science, with a focus on reasoning about personal and societal well-being from social digital footprints. She was a faculty associate with the Berkman Center for Internet and Society at Harvard, a postdoctoral researcher at Microsoft Research, and obtained her PhD in Computer Science from Arizona State University.

J. Nathan Matias is a Ph.D. Candidate at the MIT Media Lab Center for Civic Media, an affiliate at the Berkman-Klein Center at Harvard, and founder of CivilServant. He conducts independent, public interest research on flourishing, fair, and safe participation online. These include research on harassment reporting, volunteer moderation online, behavior change toward equality, social movements, and networks of gratitude.

Panel Session Format

The panel will begin with the moderator introducing the general topic of intervention versus non-intervention and the roles of various actors. Panelists will briefly state positions and how their research relates to the topic. For the bulk of the panel, the moderator will facilitate initial discussion among the panelists and audience based mainly on the following questions:

- Should social media companies engage in content moderation/curation? (arguments from many sides)
- What factors should be taken into consideration when making decisions about moderation/curation?
- What role(s) (if any) do academics have in intervening or impacting industry practices?
- What research agendas can better examine the risks and effects of possible interventions?

The discussion will take place in a debate-like fashion, where we will alternate hearing from people of opposing opinions, followed by a panelist (if there is a panelist who wishes to speak at that time, if not, we will continue with audience comments). Audience members will be able to take active roles as speakers. Whoever has the floor will only be allowed to speak for 1 minute; the moderator will make sure that no one goes substantially over the time limit: we will have a large timer displayed in the corner of the screen (or second screen, based on the configuration of the room). We will encourage active backchannel conversation via Twitter using a common hashtag. This conversation will be displayed on the main screen. The panelists will also have final remarks with the moderator wrapping up discussion.

References

1. Moira Burke, Cameron Marlow, and Thomas Lento. 2010. Social network activity and social well-being. *Proceedings of the 28th international conference on Human factors in computing systems CHI 10*, ACM Press, 1909–1912. http://doi.org/10.1145/1753326.1753613

2. Caleb T. Carr, D. Yvette Wohn, and Rebecca A. Hayes. 2016. [Thumbs up emoji] as social support: Relational closeness, automaticity, and interpreting social support from paralinguistic digital affordances in social media.

Computers in Human Behavior 62: 385–393. http://doi.org/10.1016/j.chb.2016.03.087

3. Stevie Chancellor, Jessica Annette Pater, Trustin A Clear, Eric Gilbert, and Munmun De Choudhury. 2016. #thyghgapp: Instagram Content Moderation and Lexical Variation in Pro-Eating Disorder Communities. *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing - CSCW '16*, ACM Press, 1199–1211. http://doi.org/10.1145/2818048.2819963

4. Gerardine DeSanctis and Marshall Scott Poole. 1994. Capturing the Complexity in Advanced Technology Use: Adaptive Structuration Theory. *Organization Science* 5, 2: 121–147. http://doi.org/10.1287/orsc.5.2.121

5. Jeffrey Gottfried and Elisa Shearer. 2016. *News Use Across Social Media Platforms 2016 | Pew Research Center*. Retrieved from http://www.journalism.org/2016/05/26/news-use-across-social-media-platforms-2016/

6. Matthew Heston and Jeremy Birnholtz. 2016. (In)visible Cities: An Exploration of Social Identity, Anonymity and Location-Based Filtering on Yik Yak. *iConference 2016 Proceedings*, iSchools. http://doi.org/10.9776/16152

7. Heewon Kim. 2014. Enacted social support on social media and subjective well-being. *International Journal of Communication* 8: 2340–2342.

8. Evgeny Morozov. 2014. *To save everything, click here: The folly of technological solutionism*. Public Affairs.

9. Jessica A. Pater, Moon K. Kim, Elizabeth D. Mynatt, and Casey Fiesler. 2016. Characterizing Online Harassment: Comparing Policies Across Social Media Platforms. *Proceedings of the 2016 ACM Conference on Supporting Group Work*.

10. Donghee Yvette Wohn and Brian J. Bowe. 2014. How social media facilitates social construction of reality. *Proceedings of the companion publication of the 17th ACM conference on Computer supported cooperative work & social computing*, ACM Press, 261–264.

11. Donghee Yvette Wohn, Caleb T. Carr, and Rebecca A. Hayes. 2016. How Affective Is a "Like"?: The Effect of Paralinguistic Digital Affordances on Perceived Social Support. *Cyberpsychology, Behavior, and Social Networking* 19, 9: 562–566. http://doi.org/10.1089/cyber.2016.0162

12. Ethan Zuckerman. 2016. The Perils of Using Technology to Solve Other People's Problems. *The Atlantic*. Retrieved from https://www.theatlantic.com/technology/archive/2016/06/tech-and-other-peoples-problems/488297/