Massachusetts Institute of Technology

Mapping Methodologies When the Platform is on Fire

Dave Jones
Independent Researcher

Michael Trice
Massachusetts Institute of Technology
mtrice@mit.edu

Liza Potts
Michigan State University
lpotts@msu.edu

Sweta Baniya
Virginia Tech
baniya@vt.edu

ABSTRACT

This extended abstract focuses on the methodologies used to research, examine, and understand content moderation policies on social media platforms during times of crisis.

CCS CONCEPTS

• **Human-centered computing**; • **Information systems**;

KEYWORDS

User experience, Platforms, Twitter, Content strategy, Moderation, Knowledge work, Disaster, Methodologies

1 INTRODUCTION

Researching digital spaces provides a host of challenges, especially during times of crisis when content is rapidly moving across systems and people are unsure what to believe. On social media platforms, content moderation is always tethered to both the policies that describe content governance and the ways in which those policies are applied. In these spaces, such policies and their applications are often held within what actor-network theory calls a black box, which Spinuzzi [1] describes as a force for stabilizing various complex relationships within a network. Ideally, a black box takes the complexities of processes like moderation and reduces it to "something that resembles an organized whole" [2]. Such moderation often impacts marginalized communities and hinders access to information. We link this to an established tradition in technical communication examining moderation and rhetorical practice from fora to social media [3–5].

Given that so many of these policies and actions are hidden by platform governors, how can researchers discover, uncover, and interrogate what is often an inconsistent, muddied set of practices? Or, as Ranade and Swarts [6] discuss, how do users and those providing

the infrastructure of moderation generate the principles of oversight, trust, feedback, and adaptation within such a black box? In this brief abstract, we discuss how we modeled our methodologies and adjusted our methods during times of crisis, disaster, and upheaval. Using Twitter's recent upheavals, we outline how changes in ownership, policy, and application has led to chaos while providing opportunities for other micro-blogging sites. While our phenomena of study vary across a spectrum of the design of communication (political, cultural, technological), Twitter's descent into chaos allowed for several key moments for research and invention. Building upon prior work regarding communication design [7, 8], this paper maps the functionality of Twitter and other microblogging sites (such as Post, Hive, Mastodon) to explore how genre features and innovations regarding content moderation are possible across emerging platforms within the social media ecosystem.

2 METHODOLOGIES IN CHAOS

Below, we briefly discuss how we mapped our methodologies and adjusted our methods during these turbulent moments.

2.1 Content Moderation

As technical communication scholars, we tend to ask questions about how people communicate across digital spaces, how content is constructed in these spaces, and what that means for the various policies and procedures in place. Studying content moderation often requires various methods to understand the processes, values, and rules on any given platform. For example, we might employ methods to review and code moderator rules, conduct a landscape analysis comparing moderation policies across platforms, or survey and/or interview content moderators themselves. This research requires stability on the platform itself, as the shifting sands of content moderation cannot be coded with any reliability that the same policy will be there the next day. As researchers who examine digital spaces, we are always writing a history of some kind. In the case of Twitter as it currently exists in 2023, even the historians of science and technology studies can find themselves lacking in reliable, consistent policy data and implementation examples.

2.2 Moderation as Black Box

Moderation practices are often encased within a black box, meaning the processes and policies used to moderate content on social media platforms are at least partially hidden. This capacity can serve to help that network operate more smoothly [2] by hiding complexity; these black boxes can simplify the network that people and technologies interact with. However, such obfuscation installs an immense amount of power in black boxes—hiding processes also

gives ample opportunity to avoid transparency. In cases such as Twitter, participants have recently dealt with a network in which this black box is employed to hide moderator decisions, even as Twitter's owner openly discusses his platform as a space for "absolute free speech." These conflicting messages have made moderation on Twitter a more distinctly political process, often re-platforming accounts that had been banned previously for violating policies that are ostensibly still applicable.

2.3 Impacting and Hindering Access

When Twitter created upheaval with the changes of the policies that not only impacted a large number of communities but also it hindered access of people to get information. Over the many years, people have relied on social media for the information as soon as the disaster happens, however the new changes made it so that people can no longer have an easy access as with the algorithms restricting ways in which people can access and participate in conversations. Likewise, the extent of misinformation that was seen on Twitter during the COVID-19 pandemic and the Russia-Ukraine War has also created a lot of challenges for both the information receivers, creators, and researchers. Moreover, during the Turkey-Syria earthquake, Twitter was banned by the Turkish government hindering survivor, activist, and disaster responder access to information. As the crisis continues in the world, and when our platform that we rely on is hindering access of information, saving lives and coalition building among each other becomes difficult. Furthermore, coalition building is the major ways marginalized and vulnerable groups survive the disaster [9] and hindering them the access of use and information means hindering their survival in a disaster.

2.4 Emerging Communities

New ownership has used the language of "transparency" to re-establish previously banned accounts and content while simultaneously limiting or banning others. In addition, new processes that ask users to pay for verification and access to some new features have altered perceptions of expertise and content integrity. This process has led to the Second Twitter Exodus with sites like Post, Blue Sky, and Mastodon rising as alternatives much as Gab, Truth

Social, and Parler did for far-right conservatives in 2018–2020. Examining what features are copied and what features are discarded, especially moderation features, offers insights into the evolution of social media as the lethargy of Web 2.0 finally begins to wane in text-based social media.

3 CONCLUSION

Our aim is to encourage discussions about the malleability of methods and methodologies when researching emerging, immediate, and critical events such as war, disaster, or similar events. As researchers studying in-the-moment technological change, we are forever interrogating our methodologies, approaches, and methods. In this paper, we want to open a discussion about this kind of work, what it means for technical communication research, and how we can better describe our methods in terms of reproducibility and trustworthiness to our academic fields and industry partners.

REFERENCES

[1] Clay Spinuzzi. 2008. Network: Theorizing Knowledge Work in Telecommunications. Cambridge University Press, New York, NY, USA.
[2] Bruno Latour. 1987. Science in Action: How to Follow Scientists and Engineers Through Society. Harvard University Press, Cambridge, MA, USA.
[3] Lauren. E. Cagle and Carl Herndl. 2019. Shades of denialism: Discovering possibilities for a more nuanced deliberation about climate change in online discussion forums. Communication Design Quarterly Review 7, 1 (April 2019), 22–39. https://doi.org/10.1145/3331558.3331561
[4] Chen Chen and Xiaobo Wang. 2020. # Metoo in China: Affordances and constraints of social media platforms. In J. Jones and M. Trice (Eds.), Platforms, Protests, and the Challenge of Networked Democracy, 253–269. Palgrave Macmillan Cham. https://doi.org/10.1007/978-3-030-36525-7
[5] Jordan Frith. 2017. Forum design and the changing landscape of crowd-sourced help information. Communication Design Quarterly Review 4, 2 (March 2017), 12–22. https://doi.org/10.1145/3068698.3068700
[6] Nupoor Ranade and Jason Swarts. 2022. Infrastructural support of users' mediated potential. Communication Design Quarterly Review 10, 2 (November 2022), 10–21. https://doi.org/10.1145/3507857.3507859
[7] Jennifer Sano-Franchini. 2018. Designing outrage, programming discord: A critical interface analysis of Facebook as a campaign technology. Technical Communication 65, 4 (November 2018), 387–410. Retrieved May 31, 2023 from https://www.stc.org/techcomm/2018/11/08/designing-outrage-programming-discord-a-critical-interface-analysis-of-facebook-as-a-campaign-technology
[8] Jordan Frith (Ed.). 2021. Special issue: Business and technical communication and COVID-19: Communicating in times of crisis. Journal of Business and Technical Communication 35, 1 (January 2021).
[9] Sweta Baniya. 2022. Transnational assemblages in disaster response: networked communities, technologies, and coalitional actions during global disasters. Technical Communication Quarterly 31, 4 (January 2022) 326–342. https://doi.org/10.1080/10572252.2022.2034973