Seminar on Creativity and Infrastructures, 30th November, 2023, Center for Technology & Society, TU Wien

Content moderation as a sociotechnical infrastructure problem

Kevin M. Blasiak
TU Wien & Center for Technology & Society (CTS)
kevin.blasiak@tuwien.ac.at

One of the most pressing challenges for social media platforms and their users is dealing with harmful or unwanted content, such as hate speech, misinformation, or extremism. As social media platforms and their practices have become a topic of interest for the public, journalists, and policymakers worldwide, content moderation research has also attracted much academic attention. The interest in content moderation and, in particular, solutions that address content moderation issues is not surprising, given content moderation's implications for free speech, polarization, and democratic society (Douek, 2022). Prevalent solutions come from various aspects of research and include legal or policy-driven solutions (e.g., NetzDG, Online Safety Bill), technical solutions based on the detection of harmful content (i.e., through the use of artificial intelligence), psychology-driven solutions using inoculation, and finally sociotechnical solutions (e.g., using bots, labels or crowdsourcing) to shape how users experience social media (Goldman, 2021).

However, the infrastructures set in place prevent creative, innovative solutions from emerging and solving the content moderation problem. Novel solutions face a myriad of restraints. Solutions are dismissed because they challenge the prevalent business models used by the industry, because of a lack of access to platforms and their data, and because of the politicization of content moderation and ethical considerations. As Gillespie (2023) notes, alternatives to prevalent content moderation are squeezed out of the way because they challenge the existing infrastructure, forcing researchers to solve the problems that the industry created but having to solve them on platform terms. This can lead us to overlook the underlying issues that the industry does not invite us to solve, the communities the industry tends to ignore, the exploitative business models, and the dilemmas that are not solvable but are meant to be contested.

Moreover, the prevalent paradigm, dictated by industry, ignores the underlying infrastructure while focusing on solutions at the top of the stack, taking the characteristics of the underlying infrastructures for granted (Bowker & Star, 2000). This ignores the crucial role that the characteristics of both the social and technical infrastructures have on the development of solutions, services, and applications (Tilson et al., 2010). For instance, proposals like algorithmic middleware

(Fukuyama, 2021) or “magic APIs” (Keller, 2021) are refreshing approaches to the content moderation issue because they approach the subject through infrastructure as a means of bringing about change.

Thus, this presentation aims to identify the pillars of the prevalent infrastructure that makes up content moderation. We highlight how infrastructures limit the exploration of innovative solutions and where they can catalyze change. We do this based on our experience researching and designing alternative content moderation measures to counter online extremist content on social media. We establish an overview of the existing infrastructure and show how social and technical infrastructures limit the development of creative, innovative solutions development. By doing this, we strive to enable the development of alternative solutions within the boundaries of this infrastructure and challenge existing infrastructures where necessary. This has implications for research on content moderation, as it calls for a more critical and reflexive approach that considers the infrastructural constraints and possibilities of content moderation, as well as the ethical and social implications of designing and implementing alternative solutions.

References

Bowker, G. C., & Star, S. L. (2000). *Sorting things out: Classification and its consequences*. MIT press.

Douek, E. (2022). Content Moderation as Systems Thinking. *forthcoming Harvard Law Review*, *136*, 1-82.

Fukuyama, F. (2021). The Future of Platform Power: Solving for a Moving Target. *Journal of Democracy*, *32*(3), 173-177.

Gillespie, T. (2023). The Fact of Content Moderation; Or, Let’s Not Solve the Platforms’ Problems for Them [content moderation; governance; industry research; platforms]. *2023*, *11*(2), 4. https://doi.org/10.17645/mac.v11i2.6610

Goldman, E. (2021). Content Moderation Remedies. *Michigan Technology Law Review*, *28*(1), 1-60. https://doi.org/10.2139/ssrn.3810580

Keller, D. (2021). The Future of Platform Power: Making Middleware Work. *Journal of Democracy*, *32*(3), 168-172.

Tilson, D., Lyytinen, K., & Sorensen, C. (2010, 5-8 Jan. 2010). Desperately Seeking the Infrastructure in IS Research: Conceptualization of "Digital Convergence" As Co-Evolution of Social and Technical Infrastructures. 2010 43rd Hawaii International Conference on System Sciences,