

Being watched means much more than being seen. This forum investigates information flows of sensing culled from sources as diverse as temperature check and iris scans to sound and movement sensors across terrains. After Veillance discusses how these systems distribute risk unevenly and shape the lives of populations across the globe. — Sareeta Amrute, Editor

Twitter, Facebook, and Google All Have a Trust Problem

Sareeta Amrute, University of Washington and Data & Society Research Institute

Elon Musk premised his Twitter takeover on a promise to bring “free speech” back to the social media platform. What that means in practice is unclear, and Musk’s one-sided understanding of the concept—which emphasizes unchecked expression and skepticism of content moderation—is not reassuring. Experts and democratic activists point out that Twitter’s problem is not that there is too little speech, but rather that too much of it is dominated by the same kind of voices, which are often spouting abuse and rhetoric that has the effect of shutting out broader perspectives [1,2].

Meanwhile, Facebook’s cross-check feature was deployed without the full knowledge of its own oversight board, which is supposed to act as an ethical backstop for the company. “XCheck” gives an added layer of review to politician and celebrity accounts, allowing their posts to remain on the site after they are reported and using what might be considered newsworthy, popular, or a PR risk to the company as criteria for whether the content should be taken down. As Chinmayi Arun writes, “Facebook’s Product Policy Team decided to presume that politicians’ speech is newsworthy unless” a risk of harm outweighs public interest [3].

Over at Google, the company has come under significant criticism for canceling a scheduled talk in honor of Dalit History Month by Dalit social movement leader Thenmozhi



Soundararajan; the Google employee who organized the event, Tanuja Gupta, resigned.

Insights

- While social media and search companies pay lip service to creating trust and safety on their platforms, users are better served by treating these spaces with distrust.
- A real investment would require both training staff in complex social issues and sending clear and consistent messages to staff about the continuing importance of creating trustworthy platforms for all users.

These and a litany of previous misdeeds suggest that tech companies have been unable to make their platforms worthy of trust, especially by publics who are absent in closed-door discussions among the powerful. These problems only magnify in contexts removed from the U.S.

According to a recent report from the Centre for Internet and Society [4], social media companies have not meaningfully invested in making their platforms safe and trustworthy for vulnerable users. For one, they have invested very little in understanding regional contexts and languages, which has made abuse on their platforms harder to control.

They also fail to train employees to recognize hate in local idioms, relying instead on the piecemeal free labor of minority groups themselves in the form of advisory sessions. Finally, companies fail to match their rhetoric on respecting diversity, equity, and inclusion with internal hiring, training, and retention policies for their employees. Although companies like Google claim to create a “world where everyone belongs” and “anything is possible,” their record demonstrates erasures of the work of women of color like Timnit Gebru and Tanuja Gupta, and deference to existing systems of power (<https://about.google/belonging/>).

In a 2018 op-ed [5], Soundararajan called for Twitter, then under Jack Dorsey’s leadership, to pay attention to the privacy and safety of the platform’s vulnerable users. That call is no less relevant today for all tech companies: While companies pay lip service to creating trust and safety on their platforms, users are better served by treating these spaces with distrust. The billions of dollars at play in Musk’s offer nowhere account for the damages done to vulnerable communities, nor do they include a plan for ameliorating them.

Platforms must ensure that all content moderation teams are trained in the nuances and practices of bias, especially entrenched biases like caste discrimination, misogyny and LGBTQ+ hate, anti-Black racism, and anti-Muslim sentiment that have become commonplace on the platforms. Companies should center the safety of minoritized groups on their

platforms at a structural level, moving, for instance, to create clear channels for communities to alert the companies and to quickly address organized campaigns that are aimed at suppressing their ability to communicate online. Finally, companies should end the practice of relying on users to offer insight into the harms they are facing, which their own paid workforces—with access to streams of data in real time—seem unable to surface on their own. To accomplish this, companies must encourage their employees to bring in experts to train staff in the social, economic, and political geographies—like caste—that are outside their usual practices.

ENDNOTES

1. Shanmugavelan, M. Caste-hate speech: Addressing hate speech based on work and descent. International Dalit Solidarity Network, 2021; <https://ruralindiaonline.org/en/library/resource/caste-hate-speech-addressing-hate-speech-based-on-work-and-descent/>
2. Franks, M.A. The free speech black hole. Knight First Amendment Institute, 2019; <https://knightcolumbia.org/content/the-free-speech-black-hole-can-the-internet-escape-the-gravitational-pull-of-the-first-amendment>.
3. Arun, C. Facebook’s faces. *Harvard Law Review* 135 (2022), 236–264; <https://harvardlawreview.org/2022/03/facebook-faces/>
4. Kain, D., Narayan, S., Sarkar, T. and Grover, G. Online caste-hate speech: Pervasive discrimination and humiliation. Centre for Internet and Society, Dec. 15, 2021; <https://cis-india.org/internet-governance/blog/online-caste-hate-speech-pervasive-discrimination-and-humiliation-on-social-media>
5. Soundararajan, T. Twitter’s caste problem. *New York Times*. Dec. 3, 2018; <https://www.nytimes.com/2018/12/03/opinion/twitter-india-caste-trolls.html>

Social media companies have not meaningfully invested in making their platforms safe and trustworthy for vulnerable users.

❶ **Sareeta Amrute** is an anthropologist who studies the relationship between race, work, and data. She is a principal researcher at the Data & Society Research Institute and an affiliate associate professor at the University of Washington in Seattle. She is the author of *Encoding Race, Encoding Class: Indian IT Workers in Berlin*.
→ amrutes@uw.edu

Distinguished Speakers Program

A great speaker can make the difference between a good event and a WOW event!

Students and faculty can take advantage of ACM’s Distinguished Speakers Program to invite renowned thought leaders in academia, industry and government to deliver compelling and insightful talks on the most important topics in computing and IT today. ACM covers the cost of transportation for the speaker to travel to your event.

speakers.acm.org



Association for
Computing Machinery



DOI: 10.1145/3567950 COPYRIGHT HELD BY OWNER/ AUTHOR