Deadline Extended: Moderating Data, Moderating Lives: Debating visions of (automated) content moderation in the contemporary

Announcement published by Cheshta Arora on Tuesday, April 12, 2022
Type:
Call for Papers
Date:
April 18, 2022
Location:
India
Subject Fields:
Digital Humanities, Linguistics, Popular Culture Studies, South Asian History / Studies, Sociology

Moderating Data, Moderating Lives: Debating visions of (automated) content moderation in the contemporary

Artificial Intelligence (AI) and Machine Learning (ML) based approaches have become increasingly popular as "solutions" to curb the extent of mis-, dis- mal-information, hate speech, online violence and harassment on social media. The pandemic and the ensuing work from home policy forced many platforms to shift to automated moderation which further highlighted the inefficacy of existing models (Gillespie, 2020) to deal with the surge in misinformation and harassment. These efforts, however, raise a range of interrelated concerns such as freedom and regulation of speech on the privately public sphere of social media platforms; algorithmic governance, censorship and surveillance; the relation between virality, hate, algorithmic design and profits; and social, political and cultural implications of ordering social relations through computational logics of AI/ML.

On one hand, large-scale content moderation approaches (that include automated AI/ML-based approaches) have been deemed "necessary" given the enormity of data generated (Gillespie, 2020), on the other hand, they have been regarded as "technological fixtures" offered by the Silicon Valley (Morozov, 2013), or "tyrannical" as they erode existing democratic measures (Harari, 2018). Alternatively, decolonial, feminist and postcolonial approaches insist on designing AI/ML models that centre voices of those excluded to sustain and further civic spaces on social media (Siapera, 2022).

From the global south perspective, issues around content moderation foreground the hierarchies inbuilt in the existing knowledge infrastructures. First, platforms remain unwilling to moderate content in under-resourced languages of the global south citing technological difficulties. Second, given the scale and reach of social media platforms and inefficient moderation models, the work is outsourced to workers in the global south who are meant to do the dirty work of scavenging content off these platforms for the global north. Such concerns allow us to interrogate the techno-solutionist approaches as well as their critiques situated in the global north. These realities demand that we articulate a different relationship with AI/ML while also being critical of AI/ML as an instrument of social empowerment for those at the "bottom of the pyramid" (Arora, 2016).

The seminar invites scholars interested in articulating nuanced responses to content moderation that take into account the harms perpetrated by algorithmic governance of social relations and irresponsible intermediaries while being cognizant of the harmful effects of mis-, dis- mal-information, hate speech, online violence and harassment on social media.

We invite abstract submissions that respond to these complexities vis-a-vis content moderation models or propose provocations regarding automated moderation models and their in/efficacy in furthering egalitarian relationships on social media, especially in the global south.

Submissions can reflect on the following themes using legal, policy, social, cultural and political approaches. Also, the list is not exhaustive and abstracts addressing other ancillary concerns are most welcome:

- Metaphors of (content) moderation: mediating utopia, dystopia, scepticism surrounding AI/ML approaches to moderation.
- From toxic to healthy, from purity to impurity: Interrogating gendered, racist, colonial tropes used to legitimize content moderation
- Negotiating the link between content moderation, censorship and surveillance in the global south
- Whose values decide what is and is not harmful?
- Challenges of building moderation models for under resourced languages.
- Content moderation, algorithmic governance and social relations.
- Communicating algorithmic governance on social media to the not so "tech-savvy" among us.
- Speculative horizons of content moderation and the future of social relations on the internet.
- Scavenging abuse on social media: Immaterial/invisible labour for making for-profit platforms safer to use.
- Do different platforms moderate differently? Interrogating content moderation on diverse social media platforms, and multimedia content.
- What should and should not be automated? Understanding prevalence of irony, sarcasm, humour, explicit language as counterspeech.
- Maybe we should not automate: Alternative, bottom-up approaches to content moderation

Seminar Format

We are happy to welcome abstracts for one of two tracks:

Working paper presentation

A working paper presentation would ideally involve a working draft that is presented for about 15 minutes followed by feedback from workshop participants. Abstracts for this track should be 600-800 words in length with clear research questions, methodology, and questions for discussion at the seminar. Ideally, for this track, authors should be able to submit a draft paper two weeks before the conference for circulation to participants.

Coffee-shop conversations

In contrast to the formal paper presentation format, the point of the coffee-shop conversations is to

enable an informal space for presentation and discussion of ideas. Simply put, it is an opportunity for researchers to "think out loud" and get feedback on future research agendas. Provocations for this should be 100-150 words containing a short description of the idea you want to discuss.

We will try to accommodate as many abstracts as possible given time constraints. We welcome submissions from students and early career researchers, especially those from under-represented communities.

All discussions will be private and conducted under the Chatham House Rule. Drafts will only be circulated among registered participants.

Please send your abstracts to workshops@cis-india.org.

Timeline

1. Abstract Submission Deadline (Extended): 18th April
2. Results of Abstract review: 25th April
3. Full submissions (of draft papers): 25th May
4. Seminar date: Tentative 31st May

References

Arora, P. (2016). Bottom of the Data Pyramid: Big Data and the Global South. *International Journal of Communication*, *10*(0), 19.

Gillespie, T. (2020). Content moderation, AI, and the question of scale. *Big Data & Society*, 7(2), 2053951720943234. https://doi.org/10.1177/2053951720943234

Harari, Y. N. (2018, August 30). *Why Technology Favors Tyranny*. The Atlantic. https://www.theatlantic.com/magazine/archive/2018/10/yuval-noah-harari-technology-tyranny/568330/

Morozov, E. (2013). *To save everything, click here: The folly of technological solutionism* (First edition). PublicAffairs.

Siapera, E. (2022). AI Content Moderation, Racism and (de)Coloniality. *International Journal of Bullying Prevention*, *4*(1), 55–65. https://doi.org/10.1007/s42380-021-00105-7

Contact Info:

Cheshta Arora, Centre for Internet and Society, Bengaluru, India

Contact Email:
cheshta@cis-india.org
URL:
https://cis-india.org/internet-governance/blog/cis-seminar-series