Feminist Perspectives on Social Media Governance

Synthesis Report

August 2022

FEMINIST PERSPECTIVES

ON SOCIAL MEDIA GOVERNANCE

Synthesis Report of Roundtable

by

IT for Change and InternetLab

August 2022

Contents

Contents

Acknowledgements

Synthesis of Key Points

Introduction

The original promise of freedom and democracy offered by the internet as a space for networked communication, self-expression, political participation, and public interaction beyond national borders has swiftly come and gone. The past decade has witnessed a steady rise in the amount of extreme content on online platforms in the form of cyberstalking, bullying, harassment, hate speech, disinformation, and propaganda. The intensity, scale, and ubiquity of these problems have brought to the fore the need to reimagine pathways for building a global communicative justice paradigm along feminist principles.

Today, we are confronted with misogynistic, hateful, and dangerous content on an unprecedented global scale.[1] Developing a feminist ethic of social media governance to counter such dangerous rhetoric, is therefore an urgent and pressing need. An important step in this direction is to recognize the function that extreme speech plays in deterring equal political participation and promoting further marginalization of historically oppressed sections of the society. The lack of recourse to address violence inflicted by hate speech further pushes us to probe the problem through an examination of the harms that have become normalized as part of digital culture. As part of our efforts to challenge and check the normalization of such extreme speech, we at IT for Change, have been working to tackle and understand the scale of the problem of pervasive online misogynistic speech. In April 2022, along with InternetLab Brazil, we organized a roundtable conversation among lawyers, academics, scholar-practitioners, and activists centered around the following three questions:

1. What do empirical studies of platform regulation (and self-regulation) tell us about addressing sexism and misogyny online?

2. What national legal-institutional frameworks may be appropriate to check gendered censorship and promote gender-equal participation in the online public space?

3. What kind of global responses may be relevant towards nurturing gender-inclusive online publics?

---

[1] http://webfoundation.org/docs/2020/03/WF_WAGGGS-Survey-1-pager-1.pdf

The two days of engaging discussion, and the cross-pollination of ideas from diverse disciplinary standpoints provided us with innovative ways to think about online misogyny—its social embeddings and the trajectories of legal responses to it.

This report synthesizes the key thematic threads that emerged in the course of discussion of the works presented at the roundtable, attempting to capture a snapshot of their variety and breadth. The curation of important ideas and arguments is selective, and not a comprehensive representation of the views of the participants. Nonetheless, we hope it can provide a flavor of the rich proceedings and the critical questions that were addressed during the two days about developing feminist approaches to social media governance.

Theme 1: Publicness, Visibility, and Performance

One of the overarching themes that cut across all presentations at the roundtable was the evolving conception of the public sphere and the ways in which women occupy public space. The interventions shared a recognition of the increasingly platformized conditions under which emergent forms of public action take shape in the social media context. To this effect, it was considered important to view the communicative infrastructure of the platforms as part of capitalist relations; and indeed, how the communication networks, conditions for algorithmic virality, and terms of (in)visibility themselves are shaped and governed by the logic of capital. A growing awareness of these mean that we collectively need to revisit our frames of reference and notions that presuppose the separateness of public and private spheres. In order to be able to more accurately apprehend the contemporary moment in which pervasive gender-based violence on social media is the brute reality, it is crucial to appreciate that the normative codes of public life implicate the personal in insidious ways.

Navigating this blurry distinction, some presenters drew our attention to the ways in which women's visibility in the online public sphere gets mediated by a conditional access, circumscribed as it is by patriarchal trade-offs. Panelists, Mardiya Siba Yahaya, and Anne Njathi and Rebeccah Wambui, presented works that dealt with how women content creators in the African region get compelled to tenuously navigate the contentious terrain of the online creator economy. By making patriarchal bargains that shape their choices of appearance in public

life and mold their representational self-identities in the online space, women delicately balance the wheels of visibility and violence. In other words, instead of participating in this emergent creator economy as digital entrepreneurs on equal footing, women content creators choose ways of presenting their online selves in a way that can avoid the violence resulting from self-presentations that pose a challenge to patriarchal gender norms.

Anne Njathi and Rebeccah Wambui's presentations provided examples from Kenya[2], alluding to the ways in which women receive gendered backlash in the form of "criticism" that takes aim at their bodies and character. Women content creators are forced to accept the gendered status quo, where they must resign themselves to the consequences of online visibility. Not only are women content creators subject to harassment, trolling, cyberbullying, doxxing, threats, shaming, and stalking, they are also victims of the gender pay divide, receiving less than their male counterparts. The presentation also pointed out that while platform responsiveness and accountability towards its users remains low, the possibility of counteraction against social activists criticizing the racialized system remains high. With their accounts shadow-banned or suspended for ostensibly spreading 'misinformation' and their reach getting limited, women content creators are trapped in the racial-colonial matrix of communicative capitalism.

Mardiya Siba Yahaya's work has sought to understand women's strategies of resistance in the face of such patriarchal control and surveillance in masculine online spaces. Her research, focused on Muslim influencers and content creators, interrogates the notions of visibility and moral surveillance. Challenging the patriarchal assertion that women should be hidden, these women, instead, center themselves and their identity and make money from "selling beauty".[3] At the same time, they are forced to choose ways of self-censoring to avoid the patriarchal violence that seems designed to constantly remind women that they must not take up space.

A recurring theme in the discussions was that social media is both a site of pleasure and danger. Drawing connections to the feminist assertions of the right to 'loiter' as a means of occupying and asserting ownership over public spaces, it was demonstrated how women choose myriad paths of resistance to claim space in the online public sphere. Helani Galpaya's research on online users in

---

2 Read the blurb and draft text of Anne Njathi and Rebeccah Wambui's essay here.
3 Read the blurb and draft text of Mardiya Siba Yahaya's essay here.

Myanmar provides an interesting case study on this phenomenon. The survey showed that women often use their husband or brother's Facebook account, as this is considered a safer option to having their own. Participants in the roundtable similarly presented subversive accounts of how women turn the security paradigm on its head by using their male kin's digital devices—designed to surveil them—to access the internet.

Helani Galpaya's research also showed that a statistically significant proportion of women respondents in the study use multiple social media accounts. Women often take up an alternative identity to access political debates. This alternate identity is invariably that of a *Bama* Buddhist male, the dominant ethnic and religious group in the country.

"So when you go to an online discussion forum where political issues are being debated, it often looks like a group of men are arguing, and yet this vibrant public sphere is occupied by women who appear to be men."

This is reflective of the ways in which the privileges of social location afford some respite to women who would otherwise be subjected to online vitriol. Reflecting on this finding, Galpaya spoke about the positive of how "there is a lot of 'agency' being explored by women in creating these alternate identities", but at the same time wrestled with the contradiction that the conditions enframing such exercise of agency fundamentally rests on a compromise. Even though people should be free to choose to express their publicness the way they want to, here their alternate assumed identity reveals more about the systemic erasure of those who have to rely on anonymity to exercise minimal liberties.

The cultural politics of presence, participation, and publicness are therefore very complex and deeply imbricated in contexts in which they arise. As numerous participants lamented, women are impelled by the system to make compromises, strike a "digital patriarchal bargain" to remain online, even as they pay a heavy price to visibly occupy public identities.

Theme 2: Gendered Political Violence

Multiple participants presented works that showed just how hostile the online public sphere is to women who are active in public-political life, whether in the field of formal politics or in online political discourse. One of the findings across presentations was that the degree and nature of abuse varied significantly

depending on women's social location—with those from marginalized groups having to bear a disproportionate price for the act of expressing themselves on political issues.

Shehla Rashid Shora, an activist with a vocal online presence, made an auto-ethnographic presentation in which she exhibited some instances of abuse directed at her on Twitter. In the screenshots of the vitriolic and harrowing messages directed at her on a daily basis, abusive and vitriolic tweets expressed hate towards her simply for being a Muslim woman in India. She noted that despite the fact that many of these abusive tweets contained offensive terms that are routinely used to attack Muslim women, they were allowed to remain on the platform, sometimes even after these tweets had been reported. While concluding her presentation, she pointed out that platforms like Facebook and Instagram invariably did not act on abusive content. The bar for platform accountability remains abysmally low while the threshold for violence for its users is really high!

Anita Gurumurthy and Amshuman Dasarathy presented the findings of their work on the hateful, abusive, and problematic speech on Twitter, directed at Indian women in public-political life. In their research they found that the trolling of women in political life runs rampant and remains completely unchecked by platforms despite its pervasiveness. They found that trolls rarely engaged with the stated political positions of women, resorting instead to gendered attacks on their character and bodies. Misogynistic trolls devised ways of collectively exploiting the affordances of social media platforms such as virality and anonymity to achieve their political goals. Another finding was that women from the Muslim community and political dissenters received a disproportionate amount of abuse on the platform. Perhaps most distinctively, the research showed that a large amount of abusive or violent speech directed at women in public-political life was in the nature of supposedly light-hearted jibes, misogynistic memes, wordplay, and regressive and stereotypical jokes about the place of women in society. Humor was thus perversely used as means to shroud violence against women beneath a thin veneer of political incorrectness.

Yasmin Curzi de Mendonça also presented her ongoing research based on interviews with women who are the targets of online gender-based violence in Brazil, specifically, activists, politicians, and journalists. She set out to

demonstrate the strategies of resistance, limits of the current legal-politico framework in responding to online gender-based violence (oGBV), and platform complicity in perpetuating such oGBV. She also sought to understand how such attacks impact women's lives through their digital participation. She found that in the Brazilian context, the degree of violence is contingent on determinate categories of race, gender identity, and sexuality of the person who is subjected to attack. She highlighted that for many women with public-facing work, the recourse is to silence themselves and reduce their participation in online spaces for their own safety. When confronted with oGBV, women find themselves in an isolating position, receiving neither any aid from the platform, nor from their employers. The law itself in most instances is inaccessible or is unable to translate the injury into legal norms. In order to challenge an individuated response to systemic violence, her recommendations argued for collectively confronting the problem, while creating support systems for victims.

Fernanda K. Martins presented the findings of qualitative research on online violence against journalists in Brazil that were intended to discredit, intimidate, and humiliate media professionals. Her presentation echoed the consensus that the attacks on women were both gendered and disproportionately directed towards women from minority groups. Martins explored the connection between disinformation and misogyny, and contended that the separated and isolated concepts of hate speech and disinformation are not adequate to understand the distinct nature of the violence. She proposed, instead, the necessity to think of "gendered disinformation". Elaborating on this point, she highlighted some salient narratives that were peddled to advance certain false claims about issues such as abortion, homosexuality, and feminism.

Another point that seemed to be common across multiple presentations was the differences in the nature of online abuse received by men and women. Martins briefly mentioned the findings of the MonitorA study which showed that while women politicians were abused on the basis of their appearance or their moral character, their male counterparts received messages criticizing their policies. In a different context, Galpaya's research findings show that more percentage of men in Myanmar say they have been harassed online than women. She explained that phone hacking is a big problem in Myanmar, and that when men's accounts are hacked they are asked to pay some money to have their accounts restored to them, while women's photos are uploaded and photoshopped onto nude images. Similarly, Suzie Dunn's large-scale study of technology facilitating gender-based

violence across 18 countries also found that many men are self-identifying as experiencing online harm. However, women experience higher sexualized violence based on their gender or sex. Another finding from her study is that men are predominantly the perpetrator of harm regardless of the target.

Theme 3: Contextual Content Governance

What naturally follows from a discussion of the kinds of violence that women face on social media is the question of what steps platforms can take to stop or at least lessen the amount of violent and hateful speech. The broad swathe of issues discussed under this thematic grouping dealt with the central issue of context specificity in content governance. Discussions on contextual approaches to content governance, and indeed to speech regulation more broadly, often fall prey to the unending tug of war between the absolute conceptual poles of universalism and relativism. These debates, while valuable, often tend to turn into tedious and unproductive ways of approaching the problem at hand. By choosing instead to devote sustained and closer attention to the actual experiences of online gendered violence in countries like India, Brazil, Kenya, and Azerbaijan, the participants were able to sidestep these purely conceptual discussions in favor of more grounded explorations that point to the place of 'context' in regulating online content.

Under this broader umbrella, one thematic grouping interrogated the role and efficacy of platform terms of service, or community guidelines, in tackling the issue of online gender-based violence in India. Divyansha Sehgal shared her research on how Indian platforms handle gender-based violence in their community guidelines. She noted how the guidelines act as material artifacts that set a baseline for the kind of behavior that platforms are willing to tolerate and/or encourage. Her work specifically focused on Indian short form video platforms like Moj and MX Takatak, which have risen to prominence after the ban of TikTok in India. She noted that her interest in these newer forms of visual social media grew out of an impulse to investigate whether these platforms have learned any lessons from the health and safety concerns that have arisen on social media platforms over the past decade. Are they spending more time thinking about how they want their communities to grow? A central finding based on her study of platform terms of service documents was that platforms do not place much weight on the issue of online gender-based violence, choosing instead to prioritize issues like misuse of technology and copyright violations. She also noted that there was little specificity

in the guidelines despite these being homegrown Indian platforms hosting a lot of regional language content.

Damni Kain and Shivangi Narayan's work also studied how popular social media platforms moderate caste-based hate speech and abuse. They analyzed the community guidelines, policies, and transparency reports of Facebook, Twitter, YouTube, and Clubhouse, and found that many platforms whose forms for reporting harmful content listed gender and race, did not list caste. Drawing attention to the distinctive nuance of caste-based hate speech, which is often based on more covert forms of humiliation and intimidation, they stressed the importance of being keenly aware of context to properly situate and recognize the iterations of online caste-based abuse. For example, they noted that trolls often mix caste names and words from different languages together, so that their comments appear meaningless to individuals who are not acutely aware of that particular local context. With caste not even mentioned in platform content policies, the apathy of platforms to tackle such pervasive forms of abuse and discriminatory speech in India is painfully clear.

Arjita Mital's work analyzing the secondary literature on platform community guidelines and governing policies also places the consideration of context at its center. She argues that although these governing documents are instated with the intent of protecting communities from violence and harm, they often end up reproducing the same patriarchal conditions that allow that harm to take place. She discusses how notions of obscenity and propriety are culturally variant norms and have vastly different meanings depending on the cultural context. When these concepts are used by platforms in content governance policies largely informed by an American sensibility, it works to privilege a particular way of seeing gendered bodies, and to deem forms of expression that do not fit this narrow mold as 'obscene.' Anne Njathi and Rebeccah Wambui raised a similar point in the African context by posing the question, "Where do platforms get their governing values from?" Arjita Mital also raised the problem of the vague language used in platform rules, and the lack of uniform enforcement. Giving an example, she stated:

"Twitter's rules say that behavior that crosses a line into abuse is prohibited. It could be argued that the lack of specificity and the mention of this ambiguous line encourages users to interpret it based on their own cultural contexts."

Such virtue and vice-signaling achieves little when it is not attached to a clear and definite regulatory system for online content.

Participants also spoke about the need for affected communities to participate in the development of governance standards. Divyansha Sehgal and Malavika Rajkumar noted that more external stakeholders should be involved in the creation of community guidelines, and that external audits of these documents should be periodically carried out by oversight bodies. Stressing the importance of involvement of the Dalit community in developing standards for content moderation, Shivangi Narayan spoke about how those unaffected by casteism are very often not even aware that a certain term is casteist.

Building on these presentations that critiqued the lack of contextual specificity in the content governing policies of platforms, another related group of presentations shed light on the unevenness in platform governance efforts in different parts of the globe. Analyzing the issue from a macro-level perspective, Graciela Natansohn traced the parallel (and intersecting) movement of historical dynamics of colonialism, racism, and patriarchy. Her presentation demonstrated a connection between contemporary movements toward neoliberalism and the techno-political organization of the internet today. She noted that the historical development of capitalism towards neoliberalization and digitalization solicits a dialectical reading of patriarchy—its contemporary form and contradictions.

She noted that although these are distinct spheres, they mutually determine each other, and therefore, online gender-based violence is not merely "an extension of the long term patriarchal violence that has marked the history of gender relations, but a different means to perpetuate it, demanding a specific analysis".[4]

Situating the debate within a historical-temporal context serves the purpose of recognizing why certain regions and populations are more vulnerable to online violence than others. In this context, Mashinka Hakopian presented her work on 'algolinguicism', a concept which relates to automated processes that minoritizes language users outside the Global North as well as speakers of non-Western languages, thus obstructing their access to emerging technologies and conditioning their participation on digital platforms. Such linguistic bias gets inscribed in the design-infrastructure of algorithms to produce systematic exclusion of people from non-Western contexts, excluding them from governance

---

4 Read the blurb and draft of Graciela Natansohn's essay here.

considerations. Grounded in an examination of digital platforms' failure to intervene in Azerbaijan's state-executed digital offenses that have resulted in the large-scale displacement and dispossession of Armenian people in the Republic of Artsakh, she showed how certain populations are disproportionately subject to algorithmic harm. Her research focused broadly on the question, "Whose languages are digital platforms taught to speak?"

Presenting her work on the exploitative content moderation industry, Quito Tsui critiqued the way in which social media companies are able to offset their governance responsibilities, distancing the human costs of content moderation from the geographies of their origin. On the invisibilized, underpaid, and precarious work that content moderators are made to perform, or the feminization of content moderation, she argued how,

> *"In pledging to protect online spaces, tech companies have chosen a form of platform governance that merely transfers harms from one part of the system to another."*[5]

These dynamics of content moderation being offshored to the Global South recall the 'race to the bottom' narratives that have long characterized the rapacious profiteering of large corporations. Tsui's intervention sought to explore the ways in which feminist care practices and understandings of communities of care can offer a holistic lens for the re-designing of social media governance.

Theme 4: Human-in-the-loop and Community-led Moderation Systems

Another set of issues discussed with regard to context specificity in content governance was the use of machine learning and AI-based approaches to detect online gendered violence. Shehla Rashid Shora highlighted how even the most standard abusive terms used to attack Muslim women are not flagged or taken down by platforms, contrasting this to the more developed abuse detection systems in English, by which even some common misspellings of abusive words are recognized as such by automated filtering tools. This means that even before we arrive at more nuanced and sophisticated discussions of human-in-the-loop systems and their various intricacies, platforms are not doing even the bare minimum to prevent abuse against women in the Global South by engendering bias in the AI systems.

---

5. Read the blurb and draft of Quito Tsui's essay here.

Shora also stressed the importance of inclusivity in the way machine learning models are trained. Refering to Twitter's quality filters, she noted that if marginalized women were consulted in the training of models such as these, they would be able to identify certain abusive words routinely used to attack women. Presenting their work on developing a user-facing browser plugin to detect online abusive or violent speech in three Indian languages, Cheshta Arora and Mahalakshmi J. spoke about the complex considerations to be borne in mind while building a human-in-the-loop system in consultation with the community it is intended to serve. They stressed the importance of disagreements (even within members of a particular community) being baked into the design of the tool, as reaching complete consensus on what constitutes violence is not an achievable target.

Participants also discussed the tactics used by aggressors to evade automated detection. Damni Kain spoke about how trolls mix caste names and words from different languages together, which makes it difficult for automated filtering systems to detect such content. Amshuman Dasarathy spoke about how trolls use special characters, alternate spellings, rhyme, and deeply-embedded cultural references to evade automated content filters. Yasmin Curzi de Mendonça also noted that trolls use other means to adapt their modes of attack and retain a degree of control over the popular narrative. She found that trolls tended to screenshot posts by women and post them from their own profiles, rather than using the reshare/retweet function, so as to reduce the women's ability to participate in the discourse.

Another interesting thematic thread saw participants discussing the merits and demerits of community-led moderation, in response to platform inaction in combating online gender-based violence. Arjita Mital mentioned Facebook's Community Chat and Twitter's Birdwatch feature, both of which are platform efforts at encouraging community-led approaches to content moderation. Mashinka Hakopian spoke about the guerilla regulation strategies used by the She Loves Collective, a feminist Armeninan diaspora and art collective that designed a counteroffensive—an ad hoc digital campaign to label state oppression and disinformation online. She noted that as mainstream platforms abdicate responsibility for abuses and harms, this project steps in to take up the reins of content moderation. While recognizing the necessity for such community-led efforts in the face of utter platform apathy and the crucial importance of human moderation for an appreciation of context, Anita Gurumurthy raised the counterpoint that maintaining the decency and civility of the online public sphere ought not to be left to feminist volunteerism.

Mahishaa Balraj's work in Sri Lanka to monitor and report crimes carried out on social media platforms and track platform response rates has attempted to push for better terms of transparency reporting and safer services for users. Balraj noted that such pressure and lobbying with platforms has even met with some limited success with an increase in take down of online gender-based violence content on Tik Tok.

Participants seemed to reach a consensus on the crucial importance of human moderators in detecting and tackling gendered violence on social media platforms. Interventions from Mariana Valente and Quito Tsui served as an important reminder that simply parroting the human-in-the-loop line, without any material changes to the conditions of employment for human content moderators not only perpetuates their exploitative working conditions, but also does little in terms of countering the problem of the lack of attention to context in moderation decisions. As Mariana Valente noted,

> *"Just because there's a human appraising the content doesn't necessarily mean there will be an assessment of context."*

Human moderators often spend only a matter of seconds appraising a piece of content and usually do not provide much context with which to assess it.

Theme 5: The Role of Law

In terms of calibrating the most appropriate legal-institutional responses to the problem of pervasive online gender-based violence, many participants at the roundtable recognized that a central and distinctive dimension of such online violence had to do with the unique distribution and circulation logic of social media platforms. Anita Gurumurthy's presentation of hateful, abusive, and problematic speech on Twitter showed that a majority of violent speech directed at Indian women in public-political life were acts of trolling of a "seemingly milder" variety, and derived its toxicity and potency from its volume and frequency, rather than only its locutionary content. Both Kim Barker and Suzie Dunn raised a similar point about how women receive innumerable messages that are clearly harassing or dehumanizing, but don't meet the evidentiary threshold of grave and imminent danger. As Suzie Dunn noted, "Sometimes it's not just the content, but the quantity of it." In the development of their user-facing tool to identify gendered

violence, Cheshta Arora and Mahalakshmi J. also wanted to identify such forms of content that, when appraised on a case-by-case basis may seem harmless, but cumulatively cause "everyday fatigue", and shrink the space for engagement.

As Anita Gurumurthy also stated, pervasive misogynistic speech legitimizes the discriminatory treatment of women in the online public sphere, and should therefore be construed as a harm that is relevant to law. Kim Barker pointed to how a more developed understanding of the harms of online gender-based violence is a crucial step to be taken to evolve an appropriate legal response, referring to a typology of 12 online harms to capture the breadth and severity of different forms of violence. According to Barker,

> *"Categorizing the harms caused by behaviors which are not (always) recognised as problematic by enforcement and judicial bodies, or by other stakeholders (including platforms) offers an opportunity to trigger responses and proactiveness in tackling responsibilities for online misogyny and online violence against women."[6]*

On this point of developing a deeper understanding of the harms of online gender-based violence, Shehla Rashid Shora spoke about the discrepancy between the platform take-down mechanisms for content that is dangerous, violent, or abusive, as opposed to content that is potentially a copyright infringement. While copyright infringing material is taken down immediately, dangerous or hateful content is allowed to remain on the platform, often despite repeatedly being reported by aggrieved users. She gave the example of the controversy surrounding an anti-Islamic short film titled "Innocence of Muslims", which was uploaded to YouTube, and caused riots and violent demonstrations in different countries across the globe. Despite numerous complaints to take down the film on the grounds of public order, incitement to violence, and hate speech, the film was finally taken down only on the basis of a copyright complaint. Shora also gave an example from her own experience on Twitter, in which she was unable to get the platform to take down an abusive tweet directed at her on the grounds that it was abusive. She was however, able to get it removed on the basis of a copyright claim because the abusive tweet contained her photograph. This reveals that in order to get platforms to act on something, a strong-armed pursuit of private-property claims works best.

---

6 Access the draft and blurb of Kim Barker's paper here.

From a legal perspective, Kim Barker raised the point that the stark difference in platform responsiveness between copyright enforcement and takedowns for gender-based violence is partly because there is a clear 'financial' harm recognized by judicial bodies for IP issues, while this understanding is much less developed in terms of a recognition of the harms of online gender-based violence. As was pointed out by the participants, this is evidence of a lack of any clear political will to disturb the existing business model of social media platforms and tackle the problem of pervasive online gender-based violence head-on, as this would entail disrupting their highly profitable ad-tech revenue model.

Kim Barker and Esther Lee, both highlighted the importance of recognizing platforms' complicity in fostering hostile communicative environments. Barker underscored the need for a move away from well-rehearsed statements that platforms are "mere conduits" and simply "host" the content. Similarly, Lee, in speaking about the recent revisions to South Korean law, which attempt to shatter the pervasive assumption of platforms as passive intermediaries, also raised the point about how populist opposition to these amendments argue for platform immunities from liability as a way of maintaining safe harbor for misogyny.

Esther Lee also highlighted the importance of legislative norm-setting, as violence against women has been banalized in the platform publics. Laying out the South Korean context where non-consensual intimate image distribution (NCIID) is a distressingly pervasive form of digital gendered violence, she described the "Nth Room" incident in South Korea, in which non-consensual intimate imagery extorted from women was traded across multiple social media platforms. Notably, Mahishaa Balraj's work in the Sri Lankan context also similarly found that the maximum cases were under the category of NCIID.

Lee spoke about some shortcomings of the proposed changes to the Telecommunications Business Act, also known as the Nth Room Prevention Law. One ground of critique related to a lack of clearly laid out moderation standards, and an absence of discussion or debate about who sets these standards, and whether or not they replicate systems of oppression. She brought up an example from the Korean penal code containing a provision relating to non-consensual intimate imagery, which states that a video will be deemed illegal if it creates sexual stimulus, which is an arbitrary, even ridiculous, legal standard. The question of norm-setting to address online gender-based violence through legal

mechanisms is therefore of crucial importance. Along similar lines, Nandini Chami highlighted the need to undertake a normative benchmarking exercise at the multilateral level to evolve common regulatory standards for content governance across social media platforms.

In the context of feminist norm-setting for the governance of the digital sphere, Patricia Angélica Peña Miranda's presentation dealt with the proposed changes to Chilean legal frameworks and the inclusion of new digital rights under the new draft Constitution. Against the backdrop of the widespread feminist demonstrations in Chile in 2018, which raised awareness about the pervasiveness of online gender-based violence against women, girls, and LTGBQI+ people, these proposed changes to Chilean law signal an important historical moment. Miranda highlighted two important norms that feminist organizations must promote in the new Constitution, one relating to access and connectivity, and the other, a norm that guarantees digital spaces free of violence.

While unequivocally emphasizing the importance of the law in combating online gender-based violence, both Suzie Dunn and Kim Barker reiterated that law is only one tool, and that there is a need for a more integrated, joined-up approach to address the problem. Referring to the work of Brenda Cossman, Amshuman Dasarathy noted that it is important to unsettle the exclusive authority that the discipline of law holds over the definition of gendered harm. Along similar lines, Kim Barker submitted that "while there is symbolic benefit in having the law recognize a particular societal challenge or behavior by enacting legislation to—for example—criminalize the behavior and/or harm, that sets expectations that the law alone will address the acts."[7]

Suzie Dunn raised the important point that women are often reluctant to approach the police or the criminal justice system, maybe due to the internalized misogyny in the systems. She noted that legislative change is not the only thing to aim for, but also changes in public service support, education, etc. For instance, funding for organizations to help people get content taken down. Government support for getting content taken down and regulating social media companies themselves is key. Fernanda Martins Sousa also raised her reservation about using criminal law as the primary tool to tackle online gender-based violence. Kim Barker highlighted the need to use "public health, and educational approaches … in addition to enforcement, monitoring, oversight, and legal consequences."[8]

7 Access the draft and blurb of Kim Barker's paper here.
8 Access the draft and blurb of Kim Barker's paper here.

Conclusion

The varied and even divergent opinions on what constitutes a feminist mode of doing social media governance signal the amount of work required to take meaningful steps toward a more gender-just online public sphere. The roundtable was an effort in this direction.

While the scale of the challenge is evident, the debates about increased platform accountability for online harms; unsettling the foundations of the online advertising industry; new and emerging approaches to imbue content moderation with an appreciation of context; and the role of law in combating online gender-based violence, all provided important signposts for future interventions. Feminist interventions must critically engage with each of these dimensions in order to ensure a holistic approach to social media governance.