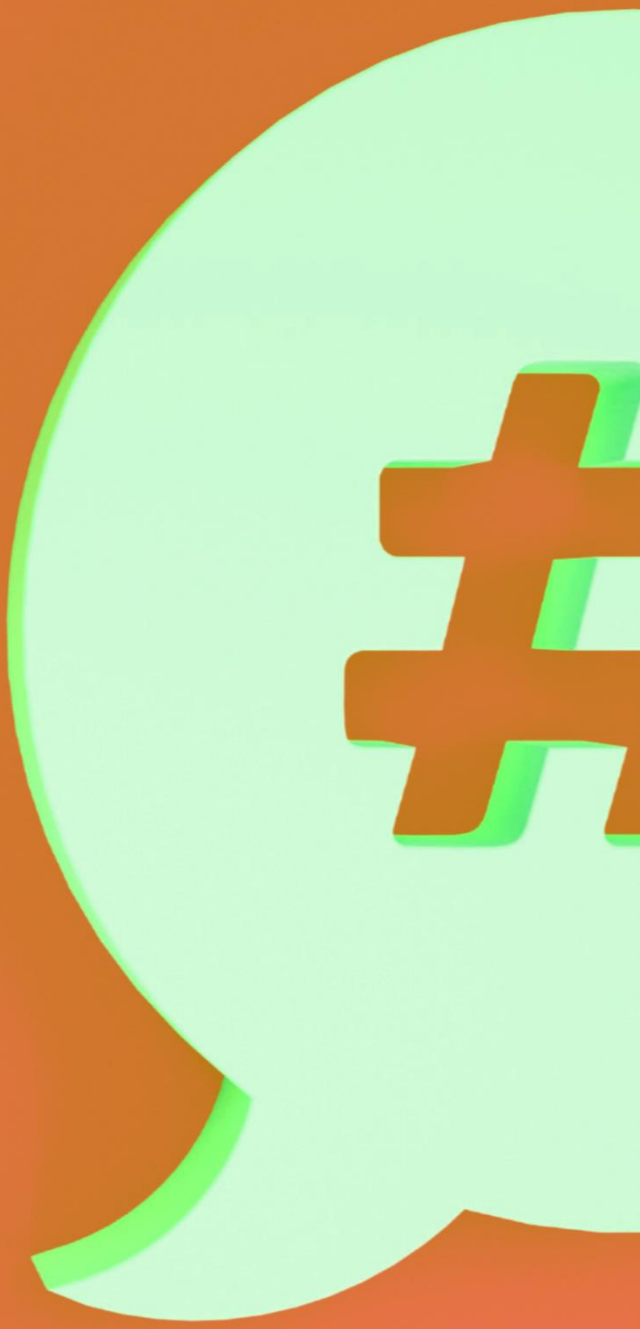
Rethinking Legal-Institutional Approaches to Sexist Hate Speech in India

Legislating an Absolute Liability Standard for Intermediaries for Gendered Cyber Abuse

Arti Raghavan



IT for Change
February 2021

Legislating an Absolute Liability Standard for Intermediaries for Gendered Cyber Abuse

Arti Raghavan

Arti Raghavan is an advocate practicing at the Bombay High Court.

This paper is part of a series under IT for Change's project, Recognize, Resist, Remedy: Combating Sexist Hate Speech Online. The series, titled Rethinking Legal-Institutional Approaches to Sexist Hate Speech in India, aims to create a space for civil society actors to proactively engage in the remaking of online governance, bringing together inputs from legal scholars, practitioners, and activists. The papers reflect upon the issue of online sexism and misogyny, proposing recommendations for appropriate legal-institutional responses. The series is funded by EdelGive Foundation, India and International Development Research Centre, Canada.

February, 2021

Conceptualisation

Anita Gurumurthy, Nandini Chami, Bhavna Jha

Editors

Anita Gurumurthy, Bhavna Jha

Editorial Support

Amay Korjan, Ankita Aggarwal, Sneha Bhagwat, Tanvi Kanchan

Design and Layout

Sneha Bhagwat

Arti Raghavan

Legislating an Absolute Liability Standard for Intermediaries for Gendered Cyber Abuse

"To say what one thought – that was my little problem – against the prodigious Current; to find a sentence that could hold its own against the male flood."

– Virginia Woolf

Introduction

Catherine MacKinnon, in her seminal (and highly provocative) book Only Words,[1] argued that pornography ought not to be afforded protection as free speech, as it deprived women of their right to free speech by having a subordinating effect on them. Social inequality, MacKinnon argued, is created and enforced through words and images.[2] Pornography, as a form of speech, through its degrading and dehumanising depiction of women, constructs the oppressive social reality that women occupy.

While MacKinnon's arguments in the context of pornography (and its regulation by the state) continue to be vigorously contested, the underlying philosophy, i.e., how speech acts create the reality we occupy, and how the exercise of free speech by some members of society infringes upon the rights, dignity, and autonomy of others, remains relevant. A commitment to free speech cannot be divorced from a commitment to equality[3] and must recognise that certain people are better positioned to exercise their right to free speech. When the right to free speech is exercised under conditions of inequality, it results in the distribution of power in society becoming "[...] more exclusive, coercive, and violent as it has become more and more legally protected."[4] This argument

[1] Catherine MacKinnon, *Only Words* (Harvard University Press, 1993).
[2] *Ibid* at pp. 31-33.
[3] *Ibid* at pp. 71.
[4] *Ibid* at pp. 72.

applies with considerable force in the context of the internet and gendered cyber abuse. It is imperative that harmful speech acts are deprived of the legitimacy of protected speech.

Gendered cyber abuse (which is a term this paper applies to describe a range of abusive or violent acts over the internet that are targeted on gendered lines, including rape and death threats, misogynistic hate speech, the non-consensual sharing of private images, cyber stalking) is not merely a reflection and reinforcement of existing structures of violence and oppression in the physical world. The anonymity and amplification provided to aggressors on online platforms serve to encourage toxic and violent behaviour that individuals might ordinarily have refrained from engaging in.

> The anonymity and amplification provided to aggressors on online platforms serve to encourage toxic and violent behaviour that individuals might ordinarily have refrained from engaging in.

Our highest constitutional court has affirmed the integral nature of the internet to the exercise of our fundamental right to freedom of speech and expression.[5] The internet, and social media platforms in particular, hold tremendous democratising potential. However, for this to be realised, it is imperative that the *causes* and the oppressive effects of gendered cyber abuse be recognised, before an appropriate legal framework is conceived to counter it.

This paper seeks to make an argument in favour of legislative intervention to counter gendered cyber abuse. Such a law must recognise that in the context of gendered cyber abuse, internet and social media platforms are not merely neutral hosts of content, but are in fact designed in a manner that enables, encourages and amplifies such harmful actions. It must be noted though that the complicity of internet platforms is not restricted to acts of gendered cyber abuse alone. However, this paper is confined in scope to addressing the issue of violence on internet platforms that is targeted along gender lines. The arguments and proposals could however be extended to other forms of online abuse including racial, religious, ethnic and caste-based forms.

The paper first examines the current legal framework (both legislative, and developed through judgments) and demonstrates why recourse to constitutional courts as a forum for *creating* law is problematic. In the next section, the paper highlights why, in light of emerging research, it is imperative that the regulation of gendered cyber abuse must recognise and address how social media and internet companies are directly implicated in such criminal acts. Specifically, the paper attempts to counter frequently invoked justifications such as feasibility and neutrality that have long been relied on by internet platforms and intermediaries to evade liability. Finally, the paper proposes a

[5] *Anuradha Bhasin v. Union of India*, (2020) 3 SCC 637, at paras 32-35, 160.2.

broad regulatory framework that holds such platforms liable for the harm caused by gendered cyber abuse.

Current Legal Framework

There are two aspects to the framework dealing with gendered cyber abuse: (i) defining the actions that constitute gendered cyber abuse; and (ii) identifying the actors who are culpable, and would face legal consequences for such unlawful actions.

On the first aspect, India has taken limited legislative steps towards regulating gendered cyber abuse by criminalising certain acts (including cyber stalking, voyeurism, circulation of private images non-consensually, and sexual harassment over the internet).[6] There is currently no law that criminalises misogynistic hate speech. This paper adopts the formulation for hate speech as set out in the 267th Law Commission Report.[7] A further elaboration on the constitutional, social and political justifications for such a law is beyond the scope of the present paper.

In respect of the second aspect (identifying actors who are liable), there is an almost singular regulatory focus on criminalising individual transgressions. This approach misses the woods for the trees. Gendered cyber abuse is not merely the result of acts of misogynistic or abusive internet users. Internet and social media platforms are implicated in these acts, as they have consciously designed spaces that enable abuse. The law as it stands, though, fails to recognise and address their culpability.

The Information Technology Act, 2000 (IT Act) provides intermediaries with a 'safe harbour' under Section 79, shielding them from liability for content posted by third parties. This is subject to the intermediaries "observ[ing] due diligence",[8] functioning merely as neutral 'conduits',[9] and taking down offending content within 36 hours of being notified of the same.[10] The obligation to take down illegal content is *only* upon the intermediary becoming notified of such content through a court order or by the appropriate government or its agency.[11]

In addition to the statutory framework, large internet and social media platforms such as Facebook, Instagram, Twitter and YouTube have community standards[12] that set out rules for appropriate user behaviour (and these standards *inter alia* proscribe gendered online abuse, including gendered hate

[6] See for instance Sections 354A, 354C, 354D of the Indian Penal Code.
[7] Law Commission of India, "Hate Speech", Report No. 267, March 2017, at pp. 51-53.
[8] Section 79(2)(c) read with Rule 3 of the IT Rules.
[9] Section 79(2) (a) and (b) of the Information Technology Act, 2000; *See also* Bahl, V. S., Rahman, F., and Bailey, R. (2020). *Internet Intermediaries and Online Harms: Regulatory Responses in India* (Data Governance Network Working Paper 06); *See also* Torsha Sarkar, "A Deep Dive into Content Takedown Time Frames", The Centre for Internet and Society, 30 November, 2019, at p. 48.
[10] Section 79 of the Information Technology Act, 2000 read with Rule 3 (2) and (4) of the Information Technology (Intermediaries guidelines) Rules, 2011 (IT Rules).
[11] *Shreya Singhal v Union of India*, (2015) 5 SCC 1, at paragraph 122.
[12] Facebook Community Standards, available at: https://www.facebook.com/communitystandards/; YouTube Community Guidelines, available at: https://www.youtube.com/intl/ALL_in/howyoutubeworks/policies/community-guidelines/; The Twitter Rules, available at: https://help.twitter.com/en/rules-and-policies/twitter-rules; Instagram Community Guidelines, available at: https://www.facebook.com/help/instagram/477434105621119

speech). However, even when notified by users about content that violates these community standards, there are no legal consequences that intermediaries face for failing to take appropriate steps.[13] Indeed, there are extensive accounts of rampant cyber abuse and unchecked hate speech that stand testimony to the tendency of these platforms to wilfully look the other way.[14]

Courting Trouble

To understand the flaws in the current legal framework dealing with intermediary liability and gendered cyber abuse, it is necessary to examine the problematic contributions by Indian constitutional courts.

Under the IT Rules as originally formulated in 2011, intermediaries were required to disable content that was proscribed by Rule 3(2) of the IT Rules (which includes content that would amount to gendered cyber abuse) within 36 hours of "[…] obtaining knowledge by itself or been brought to actual knowledge by an affected person […]".[15] This requirement was diluted by the 2015 Supreme Court judgment in *Shreya Singhal v. Union of India* (*Shreya Singhal*).[16] The Court read down the phrase "brought to actual knowledge" to mean that intermediaries were only required to act on court orders and/or a notification by the appropriate government or its agency, and not notifications by individual users.[17] The Court justified this on the basis that "[…] otherwise it would be very difficult for intermediaries like Google, Facebook, etc. to act when millions of requests are made and the intermediary is then to judge as to which of such requests are legitimate and which are not."[18] Interestingly, neither Google nor Facebook were arrayed as parties in the matter. Despite this, the purported practical difficulties faced by large, resource-rich foreign corporations such as these struck a chord with the Bench in determining the extent of an intermediary's liability. This misplaced empathy and trust in large internet platforms has been a recurring theme before our constitutional courts.

On 3 April 2019, the Madurai Bench of the Madras High Court passed an *ex parte ad-interim* order in Public Interest Litigation (PIL) proceedings instituted by one Mr. S. Muthukmar. The order prohibited

---

[13] Rule 3(11) of the IT Rules requires that intermediaries designate a grievance officer who users may contact in the event that content proscribed under the IT Rules (Rule 3(2) in particular) is hosted by the intermediary. A complaint is to be redressed within 30 days, with no clear mechanism prescribed for such redressal. Rule 3(11) adopts a 'one-size-fits-all' procedure for all types of complaints, ranging from those regarding content that harms minors and child pornography, to intellectual property related issues or grievances regarding content that invades privacy. There are no qualifications prescribed for such a grievance officer (clearly designated to perform a rather onerous, quasi-judicial function), nor a mechanism to assail or appeal a decision made upon such redressal.

[14] *See for instance* Newley Purnell and Jeff Horwitz, "Facebook's Hate-Speech Rules Collides with Indian Politics", Wall Street Journal, 14 August 2020, available at: https://www.wsj.com/articles/facebook-hate-speech-india-politics-muslim-hindu-modi-zuckerberg-11597423346; Andrew Marantz, "Why Facebook Can't Fix Itself, New Yorker, 12 October 2020, available at: https://www.newyorker.com/magazine/2020/10/19/why-facebook-cant-fix-itself; "NWMI Demands Action Against Online Abuse of Journalist Kavin Mallar", 1 September 2020, available at https://www.nwmindia.org/statements/against-sexual-harassment/nwmi-demands-action-against-online-abuse-of-journalist-kavin-malar/

[15] Rule 3(4) of the IT Rules.

[16] *Shreya Singhal v. Union of India*, (2015) 5 SCC 1, at paragraph 122.

[17] *Ibid*.

[18] *Ibid*.

the TikTok application ("app") from being downloaded, and the videos created using the app from being telecast. The Court observed that TikTok was "[…] degrading culture and encouraging pornography besides causing pedophiles [sic] and other explicit disturbing content, social stigma and other mental health issues between teens […]".[19]

The 'ban' on TikTok was eventually lifted on 24 April 2019.[20] The judgment dated 24 April 2019 records detailed submissions by ByteDance (the parent company controlling the app) as to the various safety features and community guidelines adopted by TikTok.[21] These broadly mirrored those applied by other large platforms such as Google and Facebook, and include a combination of artificial intelligence enabled mechanisms and human intervention to monitor unlawful content. In light of these features, and given the statutory remedies available to affected parties (presumably under the Information Technology Act, 2000), the *ex parte ad-interim* order imposing the ban on the app was vacated. The Court warned that "The reply affidavits filed by the [TikTok's parent companies] are treated as an undertaking that negative and inappropriate or obscene materials would be filtered and if any violation is found later, this Court would seriously view it as contempt of Court."[22] In the year that followed, graphic, violent and abusive content, including acid attack videos, continued to be freely circulated on the platform, without any legal consequences for the content creators, disseminators or the platform.[23]

This script has played out (with minor variations) in various proceedings before constitutional courts in India, with well-intentioned petitioners seeking sweeping directions from courts, but failing to achieve an effective remedy against the widespread circulation of abusive or violent content over online platforms.

Graphic, violent and abusive content, including acid attack videos, continued to be freely circulated on the platform, without any legal consequences for the content creators, disseminators or the platform.

In Re: Prajwala Letter Dated 18.2.2015 Videos of Sexual Violence and Recommendations[24] (Prajwala) was a PIL before the Supreme Court. The proceedings were instituted on the basis of a letter

[19] *S. Muthukumar v. Telecom Regulatory Authority of India & Ors,* Order dated 3 April 2019 in W.P. (MD) No. 7855 of 2019 before the Madurai Bench of the Madras High Court. The Court noted that since TikTok's launch in countries outside China, it had "[...] spread like a virus and 500 millions [sic] of people are using [TikTok]". The Court also directed the Union of India to respond as to whether it intended to enact a legislation similar to an enactment in the United States of America titled the Children's Online Privacy Protection Act.
[20] *S. Muthukumar v. Telecom Regulatory Authority of India & Ors.*, 2019 SCC OnLine Mad 24317. The Indian entity that managed TikTok in India – Bytedance (India) Technology Private Limited – was impleaded in the matter.
[21] *Ibid.* at para 9.
[22] *Ibid.* at para 13.
[23] *See* Rakhi Bose, "TikTok is full of Videos that Promote Acid Attack and Sexual Abuse. App Ban Is Still Not the Answer", News18, May 19 2020, https://www.news18.com/news/buzz/banning-acid-attack-video-from-tiktok-is-a-start-but-gender-violence-is-common-trope-on-social-media-2626811.html; *See also* Shreya Chauhan, "Some TikTok Content Normalises Violence Against Women Among Other Things; Banning A Solution?", IndiaTimes, 19 May 2020, available at: https://www.indiatimes.com/trending/social-relevance/some-tiktok-content-normalises-violence-against-women-among-other-things-banning-a-solution-513633.html
[24] *In Re: Prajwala Letter Dated 18.2.2015 Videos of Sexual Violence and Recommendations*, Supreme Court, (2018) 15 SCC 573.

addressed by an NGO to the Court, seeking intervention on the issue of the circulation of videos depicting sexual violence and child pornography. An ad hoc committee was constituted by the Court, that included representatives from internet giants such as Google, WhatsApp, Facebook, Yahoo, and Microsoft. The mandate of the committee was to assist and advise the Court "[…] *on the feasibility of ensuring that videos depicting rape, gang rape and child pornography are not available for circulation*, apart from anything else, to protect the identity and reputation of the victims and also because circulation of such videos cannot be in public interest at all" (emphasis added).[25]

Following discussions and the submission of proposals by the members of the Committee, the Court only considered those proposals and recommendations where there was a *consensus* by all the members of the ad hoc committee (and in particular, the large social media companies). Notably, none of these proposals or recommendations fixed any legal responsibility on the platforms to proactively identify and take down illegal content. At best, the proposals can be seen as suggestions or good faith undertakings to better monitor content that is uploaded on these platforms. For instance, on the issue of whether platforms can provide a separate link or mechanism to report child pornography and videos depicting rape and gang rape, the represented companies evasively stated that they "[…] are continuously working on improving processes for reporting content including child pornography and videos depicting rape and gang rape that violate their policies or applicable laws. The Committee noted the same."[26] It is entirely unsurprising that these entities did not consent to stricter standards, nor agree to assume any legal responsibility for the content. What is particularly problematic, though, was the Court's complicity in lowering the bar for their conduct.

> It is imperative that the exercise of formulating laws to counter gendered cyber abuse be wrested away from courts, and be undertaken in a transparent, consultative manner through the legislature.

The proceedings before the Supreme Court in *Sabu Mathew v. Union of India & Ors.*[27] represent another judicial misadventure. Here, a writ petition was filed for the effective implementation of certain provisions of the Preconception and Prenatal Diagnostic Techniques (Prohibition of Sex Selection) Act, 1994 (PNDTA). The central concern was advertisements for prenatal sex determination (that is prohibited under law) that were published on internet platforms. Google, Yahoo and Microsoft were arrayed before the Court, and a series of orders were passed.[28] The Court gave short shrift to its adjudicatory role. Instead, the proceedings unfolded through a series of muddled orders

[25] *Ibid* at para 95.
[26] *Ibid* at para 95 (at item 11).
[27] *Sabu Mathew v. Union of India & Ors.*, (2017) 7 SCC 657; *Sabu Mathew v. Union of India & Ors*., (2018) 3 SCC 229.
[28] *Sabu Mathew v. Union of India & Ors.*, (2015) 11 SCC 545; *Sabu Mathew v. Union of India & Ors.*, (2016) 14 SCC 418; *Sabu Mathew v. Union of India & Ors.*, (2017) 2 SCC 514; *Sabu Mathew v. Union of India & Ors.*, (2017) 7 SCC 657.

and measures. These included having internet companies respond to questionnaires as to their willingness and ability to ensure compliance with the law,[29] directing the companies to adopt a process of "auto-blocking"[30] (that required the application of certain keywords to pre-empt the publication of infringing content), subsequently watering down that direction, and then mandating the constitution of a Nodal Agency to intimate the platforms of content that violated the PNDTA.[31] The Court failed to undertake the basic exercise of examining and applying the existing framework of law, including the procedure to be followed for the takedown of infringing content under the Information Technology (Procedure and Safeguards for Blocking for Access of Information by Public) Rules, 2009.[32] Similarly, there was a failure to reconcile the sweeping orders regarding "auto-blocking" with the law as laid down in *Shreya Singhal* (that limited intermediaries' takedown obligations to instances where they received notice of a court order or government directive to that effect).[33]

As demonstrated above, multiple problems have emerged in the attempts by the judiciary to regulate cyber harms: (i) a failure to clearly articulate the scope of platform liability by applying procedure and penalties *as they exist* under the law. This exercise is instead jettisoned in favour of judicially legislating new laws, thus creating further uncertainty; (ii) a failure to recognise the active culpability of platforms in the generation and dissemination of such content; and (iii) the undue deference to *feasibility* in the context of moderation of harmful content on the internet. It is therefore imperative that the exercise of formulating laws to counter gendered cyber abuse be wrested away from courts, and be undertaken in a transparent, consultative manner through the legislature.

The Fallacy of Neutrality

As explained in a recent paper by Luke Munn, "Just as the design of urban space influences the practices within it […] the design of platforms, apps and technical environments shapes our behaviour in digital space. This design is not a neutral environment that simply appears, but is instead planned, prototyped, and developed with particular intentions in mind."[34] Unfortunately, much of the evidence that bears out the particular intentions of internet corporations is not in the public domain, and is only available to those within these entities. However, in the recent past, there have been

[29] *Sabu Mathew v. Union of India & Ors.*, (2017) 2 SCC 514, at para 6 of the order dated 19 September 2016.
[30] *Sabu Mathew v. Union of India & Ors.*, (2017) 2 SCC 514, at para 7-10 of the order dated 19 September 2016.
[31] *Sabu Mathew v. Union of India & Ors.*, (2017) 2 SCC 514, at para 21 of the order dated 11 April 2017.
[32] Amrita Vasudevan and Anita Gurumurthy, "Costly Ambiguities – A Gender Based reading of the last orders of the Supreme Court in the Sabu Mathew George v. Union of India case", available at: https://itforchange.net/index.php/gender-based-reading-of-the-Sabu-Matthew-Case; *See also* "Statement of Concern on the Sabu Mathew George Case: Don't Auto-Block Online Expression", Internet Freedom Foundation, 20 February 2017, available at: https://internetfreedom.in/statement-of-concern-on-the-sabu-mathew-george-case-dont-auto-block-online-expression/
[33] "Statement of Concern on the Sabu Mathew George Case: Don't Auto-Block Online Expression", Internet Freedom Foundation, 20 February 2017, available at: https://internetfreedom.in/statement-of-concern-on-the-sabu-mathew-george-case-dont-auto-block-online-expression/
[34] Luke Munn, "Angry by Design: Toxic Communication and Technical Architecture", *Humanit Soc Sci Commun* 7, 53 (30 July 2020), at p. 2. Available at https://doi.org/10.1057/s41599-020-00550-7

"confessional moments" from those responsible for the design of these platforms,[35] including admissions as to how they are designed to be addictive and that they exploit negative triggers.[36]

There is mounting evidence for how hate speech is encouraged by internet platforms. News feeds on platforms such as Facebook and Reddit are designed to promote content with the most engagement.[37] It has been observed that negative and primal emotions (including anger and fear) draw the most engagement, and consequently get the most visibility.[38] Under normal circumstances, studies have shown that people are not typically inclined to accept views or beliefs that are seen as extreme.[39] The enhanced visibility provided to toxic or negative content on online platforms (through news feed algorithms) results in such harmful or toxic views or beliefs no longer being perceived as fringe or extreme. Users are thus presented with an algorithm-altered reality where fringe views and extremist speech are amplified and boosted, and are more likely to be considered acceptable, as they appear 'mainstream'.

News feeds on platforms such as Facebook and Reddit are designed to promote content with the most engagement. It has been observed that negative and primal emotions (including anger and fear) draw the most engagement, and consequently get the most visibility.

An example of this was the popularity of fora on Reddit that targeted feminists and persons who were considered overweight. These sub-reddits had users indulging in crude, demeaning speech and inciting hatred against feminists and 'fat people', and despite the viciously demeaning content, the posts on these fora were extremely popular (until they were banned by Reddit).[40] As observed by Fisher and Taub, "Such ideas can naturally proliferate on social media algorithms, by indulging anger against vulnerable targets and us-versus-them tribalism."[41] Speech that would otherwise be considered offensive or unacceptable in society gains currency when endorsed by thousands (or millions) of internet users.

[35] *Ibid.* at p. 3.
[36] Paul Lewis, "'Our minds can be hijacked': The tech insiders who fear a smartphone dystopia." The Guardian, 6 October 2017, available at https://www.theguardian.com/technology/2017/oct/05/smartphone-addiction-silicon-valley-dystopia
[37] Ibid 34 at p. 3.
[38] Fisher, M. and Taub, A. (2018) "How everyday social media users become real-world extremists". New York Times, October 10, 2018; available at https://www.nytimes.com/2018/04/25/world/asia/facebook-extremism.html; *See also* Fan R, Xu KE, Zhao J, "Higher contagion and weaker ties mean anger spreads faster than joy in social media", 5 November 2018, available at http://arxiv.org/abs/1608.03656; Jessie Daniels, "The Algorithmic Rise of the alt-right", 28 March 2018, available at: https://contexts.org/articles/the-algorithmic-rise-of-the-alt-right/; Luke Munn, "Angry By Design: Toxic Communication and Technical Architecture", *Humanit Soc Sci Commun* 7, 53 (30 July 2020), at p. 2, available at https://doi.org/10.1057/s41599-020-00550-7
[39] *Ibid*. 38.
[40] *Ibid*. 38.
[41] *Ibid*. 38.

The inherently warped architecture of such platforms also normalises, encourages, and amplifies harmful conduct by facilitating easy dissemination of such information.[42] It has been contended that control and amplification of content, as well as manipulative algorithms, are themselves protected as a form of free speech.[43] However, when such acts produce social harms that directly affect public order, they fall within the permissible grounds to regulate the freedom of speech and expression, in terms of Article 19(2) of the Indian Constitution. In the case of gendered cyber abuse, such acts do not merely have law and order implications, but also affect public order as they have a profound impact on society and the community at large. For instance, the issuance of death and rape threats, or the publication of private images would affect women in general, by making them feel unsafe and vulnerable to similar attacks, should they choose to be participants on various internet platforms. The regulation of such speech is thus constitutionally defensible as laid down by various judgments of the Supreme Court.[44]

It is particularly perverse that the problematic architecture of these digital spaces and its consequences have been brought to the attention of senior executives of these platforms, but have been ignored. For instance, when internal studies at Facebook demonstrated how divisive and polarising content was being actively fed to users (in an attempt to grab more attention), the company's executives ignored the findings.[45] It is therefore essential that the law provides the necessary deterrents and penalties to ensure that internet companies and social media platforms – particularly the ones with large user bases and deep pockets – effectively deploy their resources and efforts to ensure that such harmful acts are curbed.

The Feasibility Bogey

A recurring theme in discussions on the regulation of online content is the issue of *feasibility*, given the volume of data and the number of users on various internet and social media platforms.

The judgment in *Shreya Singhal*, as explained above, made it difficult for individual users to seek swift remedies to block content that amounts to gendered cyber abuse. No matter what the nature of the unlawful content was – whether it infringed copyrights or published private details or images of a woman – it could only be disabled by recourse to courts or appropriate government agencies. Users who may be subject to rape or death threats or vicious hate speech could not hold the relevant intermediary liable for failing to disable the content when notified of it. Instead, they were required to

[42] Luke Munn, "Angry by Design: Toxic Communication and Technical Architecture", *Humanit Soc Sci Commun* 7, 53 (30 July 2020), at p. 3. Available at https://doi.org/10.1057/s41599-020-00550-7
[43] Clyde Wayne Cruz Jr., "The Case Against Social Media Content Regulation", Competitive Enterprise Institute, June 2020, at p. 8, available at: https://cei.org/sites/default/files/WayneCrewsTheCaseagainstSocialMediaContentRegulation.pdf
[44] *Arun Ghosh v. State of West Bengal*, (1970) 1 SCC 98, at pp. 99 - 100 that was cited with approval in Shreya Singhal at para 38.
[45] Jeff Horwitz and Deepa Seetharaman, "Facebook executives shut down efforts to make the site less divisive." Wall Street Journal, 26 May 2020, available at: https://www.wsj.com/articles/facebook-knows-it-encourages-division-top-executives-nixed-solutions-11590507499

approach law enforcement agencies, obtain an order from an appropriate court and then have the intermediary disable the objectionable content. The degree of harm caused for every hour that such content continues to be available (often publicly, and in circulation) causes immeasurable harm to the victim: harm that the Court was impervious to.

> Users who may be subject to rape or death threats or vicious hate speech could not hold the relevant intermediary liable for failing to disable the content when notified of it. Instead, they were required to approach law enforcement agencies, obtain an order from an appropriate court and then have the intermediary disable the objectionable content.

On 22 June 2020, the European Commission (EC) released the results of its fifth evaluation of the 2016 Code of Conduct[46] on countering illegal hate speech online. Data from 23 member countries and the United Kingdom was collected for a sample period of six weeks in 2019 to assess the responsiveness of social media companies to notifications of hate speech. The notifications studied included those made over channels available to general users, and those by trusted flaggers and reporters. The EC observed that "On average 90% of the notifications are reviewed within 24 hours and 71% of the content is removed." Facebook – the platform that received the largest number of notifications for hate speech – was found to have assessed notifications in less than 24 hours in 95.7% of the cases and 3.4% in less than 48 hours. Unsurprisingly, Facebook brandished these findings in its newsroom, with a declaration that "We don't allow hate speech on Facebook."[47] Similarly, a study of the turnaround time of large incumbent intermediaries to notifications of unlawful content under the Germany Network Enforcement Act or NetzDG demonstrated that they were able to respond to a substantial number of complaints within a 24-hour window.[48]

There is some debate as to what these numbers represent (particularly in terms of the accuracy of the responses and the volume of un-notified hate speech that remains on the platforms).[49] But what is clear is that in jurisdictions with strong regulatory oversight (such as the European Commission) and where there is the threat of large fines (such as those prescribed under the NetzDG[50]), internet and social media platforms take the exercise of content moderation seriously, and do not hide behind

[46] European Commission, *Countering Illegal Hate Speech Online: Fifth Evaluation of the Code of Conduct*, 22 June 2020, available at: https://ec.europa.eu/commission/presscorner/detail/en/ip_20_1134
[47] Guy Rosen, "New EU Report Finds Progress Fighting Hate Speech", Facebook, 23 June 2020, available at https://about.fb.com/news/2020/06/progress-fighting-hate-speech/
[48] On the basis of transparency reporting conducted for the period January – June 2018. [C.f. Torsha Sarkar, "A Deep Dive into Content Takedown Time Frames", The Centre for Internet and Society, November 30 2019, at pp. 10 – 11].
[49] J. Fergus, "Facebook Will Protect Users from Itself for a Limited Time Only", Input, 31 October, 2020, available at https://www.inputmag.com/tech/facebook-will-protect-users-from-itself-for-a-limited-time-only
[50] Network Enforcement Act available at: https://germanlawarchive.iuscomp.org/?p=1245

arguments of feasibility when presented with strict requirements for reviewing potentially infringing content.

The burden for devising appropriate mechanisms to combat abusive content online ought not to be borne by regulators or civil societies, but must be placed firmly at the doorstep of the enterprises that unleashed – and profit from – this problem.

Therefore, while crafting a legal framework to regulate such intermediaries, excessive deference to the professed feasibility issues of moderating content must be abandoned. Platforms such as Facebook, Instagram, WhatsApp, YouTube, and Twitter, that see the most significant volume of hate speech and abuse, are owned and controlled by resource-rich corporations with billions of dollars in turnover. The burden for devising appropriate mechanisms to combat abusive content online ought not to be borne by regulators or civil societies, but must be placed firmly at the doorstep of the enterprises that unleashed – and profit from – this problem.

Proposed Regulatory Framework for Gendered Cyber Abuse

A useful legal principle to understand this approach is that of absolute liability, as created and affirmed by the Indian Supreme Court: an enterprise that is engaged in a hazardous or inherently dangerous activity that harms anyone (even by accident) is *absolutely liable* to compensate those who are affected by such accident. The compensation payable ought to be commensurate to the means available to the corporation. This principle was conceived and applied in the judiciary's populist pronouncements, such as in the context of industrial disasters like the Oleum Gas Leak.[51] It would be apposite in the context of large internet platforms and social media companies that are, in effect, engaged in the business of manipulating human behaviour for profit. These platforms have deliberately designed spaces that encourage harmful behaviour that has direct and prejudicial consequences on women, amongst other vulnerable communities. When business enterprises enable toxic masculinity, permit the issuance and wide dissemination of death and rape threats, and thus have a chilling effect on the participation of women in society on account of fear of abuse both online and offline, they cannot be absolved of the consequences of such harms.

Accordingly, this paper proposes a regulatory framework with the specific intent of addressing the issue of gendered cyber abuse. A narrow, targeted legislation (aimed at only regulating gendered cyber violence) would potentially avoid the pitfalls of over regulation that the NetzDG suffers from. In

[51] *M.C. Mehta vs. Union of India (UOI) and Ors.* (1987) 1 SCC 395, at paras 31-32. *See also Union of India (UOI) v. Prabhakaran Vijaya Kumar & Ors*, (2008) 9 SCC 527, at paras 18-37.

particular, the over-broad definition of what amounts to "violating content" in NetzDG is of particular concern,[52] as is the absence of any procedure to challenge content removal decisions.[53]

One significant issue in the current legal framework is the lack of proper classification (in terms of scale, reach and resources), and targeted regulation of intermediaries.[54] A law that is intended to hold intermediaries liable for gendered cyber abuse would not apply to all intermediaries that host third-party content. It would apply to those large intermediaries that enjoy a significant user base and substantial turnover. This is not to suggest that gendered cyber abuse on platforms with relatively smaller user bases and revenue is not pernicious. The differential classification and regulation are justifiable on the grounds of their wider reach and potential for greater harm. Further, it ensures that smaller or newer intermediaries are not forced out of the market because they lack the deployable resources to address gendered cyber abuse. To force such smaller platforms out would have profound implications on free speech over the internet, by strengthening the monopoly of larger, entrenched platforms. The harm caused by this would arguably outweigh the benefits of also targeting gendered cyber speech on smaller platforms.

The other key aspects of the proposed legislation are:

1. That the targeted intermediaries would be required to respond to all notifications in respect of gendered cyber abuse – whether by individual users, government agencies or authorities, or a court order – within a limited period of time (potentially 48-72 hours), with a more restricted response period (possibly 24 hours) in the cases of patently harmful content such as death threats, rape threats, videos of sexual abuse or violence and child pornography. Such intermediaries must make available an accessible reporting mechanism on the platform interface so that users may notify it of unlawful content. These intermediaries would also be required to (i) confirm the receipt of a notification or complaint; (ii) intimate the author of the content complained of; and (iii) intimate both complainant and author of an eventual decision in respect of the content.
2. The constitution of a statutory regulator who would prosecute errant companies before specially constituted tribunals. Particularly given the subject matter of the legislation, it is important that the appointments to such a regulator and Tribunal are representative in terms of gender composition (though this alone does not secure gender just outcomes of the judicial process).

[52] ibid 52 at p. 17.
[53] Germany: The Act to Improve Law Enforcement of the Social Networks, at p. 17 August 2017, available at: https://www.article19.org/wp-content/uploads/2017/09/170901-Legal-Analysis-German-NetzDG-Act.pdf
[54] Bahl, V. S., Rahman, F., & Bailey, R. (2020). *Internet Intermediaries and Online Harms: Regulatory Responses in India* (Data Governance Network Working Paper 06); *See also* Torsha Sarkar, "A Deep Dive into Content Takedown Time Frames", The Centre for Internet and Society, 30 November, 2019, at p. 24.

3. The regulator could act *suo moto*, or upon complaints from citizens, NGOs, feminist organisations or civil society groups. This is essential, as the punishing nature of the criminal justice process frequently deters women from formally seeking action in cases of gendered cyber abuse. It is also important that the regulator functions in a transparent manner. For instance, audits by a committee (that includes external members), reviewing its prosecutorial decisions and responses to complaints, may be considered.

4. That in the event of repeated failures to respond to notifications in time, the regulator may seek the imposition of a penalty before a specially constituted tribunal. It is essential that – in a departure from the negligible penalties prescribed under the IT Act[55] – the sanctions imposed be significant (taking a cue from the German NetzDG that imposes penalties up to EUR 50 million[56]). This is to ensure that platforms take the threat of regulatory action seriously and invest their resources in improving the platform's ability to detect and remove offensive content.

There may be legitimate concerns that imposing significant penalties may result in 'over-regulation' of online content by private enterprises.[57] In fact, this criticism has been raised in the context of the NetzDG in Germany.[58] It is also argued that private entities are ill-equipped to make determinations as to content illegality.[59] Given the reach and user base that these platforms command, having intermediaries make unassailable, unilateral decisions in respect of content, or suspending or deleting accounts does have a tangible, detrimental and often irreversible consequence on the freedom of speech.[60]

However, as things stand, the private, arbitrary regulation of speech online – whether *suo moto* by the intermediaries or at the behest of states – is a reality,[61] even in the absence of a regulatory regime of a law prescribing penalties for unlawful content. The problem of arbitrary regulation of online content by such private enterprises can only effectively be countered by laws that require

[55] *See for instance* Sections 44-45, and 71-73 of the IT Act.
[56] Ibid. 50.
[57] Intermediary Liability 2.0: A Shifting Paradigm, Software Freedom Law Centre, 15 March 2019, available at: https://sflc.in/intermediary-liability-20-shifting-paradigm
[58] Germany: The Act to Improve the Enforcement of the Law in Social Networks", Article 19, August 2017, at p. 19, available at: https://www.article19.org/wp-content/uploads/2017/09/170901-Legal-Analysis-German-NetzDG-Act.pdf
[59] Article 19, "*Responding to Hate Speech: Comparative Overview of Six European Countries*", 2018, Page 11, (available at: https://www.article19.org/wp-content/uploads/2018/03/ECA-hate-speech-compilation-report_March-2018.pdf); United Nations Human Rights Council, "Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression", 11 May 2016 (para 40-44).
[60] Zeynip Tufecki, "The Problem with All Tech Hearings", October 28, 2020, available at: https://zeynep.substack.com/p/the-problem-with-all-the-tech-hearings
[61] *See for instance* Vidhi Doshi, "Facebook under fire for 'censoring' Kashmir-related posts and accounts", The Guardian, 19 July 2016, available at: https://www.theguardian.com/technology/2016/jul/19/facebook-under-fire-censoring-kashmir-posts-accounts; Gilad Edelman, "Surprise! The Section 230 hearing Wasn't about Section 230", 28 October 2020, available at https://www.wired.com/story/section-230-hearing-wasnt-about-section-230/; Nosheen Iqbal, "Instagram row over plus-size model forces change to nudity policy", The Guardian, 25 October 2020, available at: https://www.theguardian.com/technology/2020/oct/25/instagram-row-over-plus-size-model-forces-change-to-nudity-policy

more accountability, transparency and better processes to challenge such decisions.[62] Concern regarding private regulation of speech ought not to jettison efforts to fasten liability on these platforms for the harm caused by cyber abuse. It only calls for greater regulation and accountability in respect of the decision-making aspect as well.

Conclusion

Gendered cyber abuse is by no means the only (or most) serious issue to contend with while considering the governance of the digital world. Ethnic,[63] religious[64], and sexual minorities have been victimised over internet and social media platforms,[65] with the violence being enabled in much the same manner as with gendered cyber abuse. Countries have also been contending with the problem of fake news, and the abuse of digital spaces to influence electoral outcomes.[66] One of the biggest problems confronting the governance of digital spaces is the vested political interest in preserving un-moderated online spaces so that political leaders may take advantage of them.[67] The role and responsibility of internet platforms for each of these issues is similar to their culpability for acts of gendered cyber abuse, and necessitate similar regulatory intervention. However, the regulation of gendered cyber abuse (by holding intermediaries liable for failing to disable infringing content) may be a legislative project that is less polarising, and may not provoke as much fear of its potential for misuse for private or political censorship. While it may be politically expedient to pursue such targeted legislation (instead of a more comprehensive but politically fraught regulatory framework to address all forms of hate speech over the internet), every effort must be made to avoid the risk of fracturing or compromising the larger cause of democratising the internet for all vulnerable and marginalised persons.

Further, a statutory framework that imposes strict penalties on large intermediaries for repeated failures to respond to notifications of gendered cyber abuse is no silver bullet to addressing the issue.

---

[62] Gennie Gebhart, 'Who Has Your Back? Censorship Edition 2019' (Electronic Frontier Foundation, 12th June 2019), available at https://www.eff.org/wp/who-has-your-back-2019#scope; The Santa Clara Principles on Transparency and Accountability in Content Moderation, available at: https://santaclaraprinciples.org/; Torsha Sarkar, "A Deep Dive into Content Takedown Time Frames", The Centre for Internet and Society, 30 November, 2019, at p. 23.

[63] Zachary Laub, "Hate Speech on Social Media: Global Comparisons", Council on Foreign Relations, 7 June 2019, available at: https://www.cfr.org/backgrounder/hate-speech-social-media-global-comparisons

[64] "Facebook Admits it was Used to Incite Violence in Myanmar", New York Times, 6 November 2018, available at: https://www.nytimes.com/2018/11/06/technology/myanmar-facebook.html

[65] Ben Hunte, "Transgender people treated 'inhumanely' online", British Broadcasting Corporation, 25 October 2019, available at: https://www.bbc.com/news/technology-50166900

[66] Turgay Yerlikaya, "Social Media and Fake News in the Post-Truth Era: The Manipulation of Politics in the Election Process." *Insight Turkey* 22, no. 2 (2020): 177-96.

[67] The recent hearings before the Senate Commerce Committee of the United States of America regarding Section 230 of the 1996 statute, the Communications Decency Act bears this out. Section 230 grants interactive computer services broad legal immunity for user generated content, while also allowing them to moderate content (without any liability for decisions made in that regard). It was revealing that the primary focus of the hearings appeared to question content moderation decisions taken by internet giants (and in particular Twitter, which ironically has a far more minuscule user base than Facebook or Google). There was regrettably little focus (and consequently limited disagreement between Senate leaders) as to issues of fastening liability on these interactive computer services for third-party generated content. (*See* Gilad Edelman, "Surprise! The Section 230 hearing Wasn't about Section 230", 28 October 2020, https://www.wired.com/story/section-230-hearing-wasnt-about-section-230/)

Patriarchy and misogyny are deeply entrenched in the law enforcement apparatus, and we have repeatedly witnessed how laws intended to protect women are weaponised and used against them. The real battle is to ensure that the legislative and judicial process is imbued with gender sensitivity and a feminist consciousness.

> Patriarchy and misogyny are deeply entrenched in the law enforcement apparatus, and we have repeatedly witnessed how laws intended to protect women are weaponised and used against them. The real battle is to ensure that the legislative and judicial process is imbued with gender sensitivity and a feminist consciousness.

Ours is a country where rape and sexual abuse victims are vilified[68] and sometimes incarcerated,[69] where attempts to de-sexualise female bodies are met with criminal sanctions,[70] where misogynistic hate speech often finds its place in pronouncements by our constitutional courts,[71] and where women are not part of critical decision-making bodies and institutions.[72] The law has demonstrably not been a great ally to the Indian feminist cause. This sobering reality need not mean that the legislative reform be abandoned as a lost cause, but that the law be approached with an unwavering, critical eye, and a more tempered expectation of the liberation and freedom that it can deliver.

---

[68] See for instance "CJI Gogoi's Misuse of SC as a Platform Sets Back #MeToo Movement", Quint, 23 April 2019, available at: https://www.thequint.com/voices/opinion/cji-ranjan-gogoi-sc-hearing-sexual-harassment-allegations-abuse-of-power; Rohit Parihar, "Girl who accused Asaram of sexual assault is mentally disturbed, says Ram Jethmalani", India Today, 16 September 2013, available at: https://www.indiatoday.in/featured/story/asaram-bapu-lawyer-ram-jethmalani-girl-is-mentally-disturbed-211207-2013-09-16

[69] Anumeha Yadav, "Araria Rape Survivor Shares Her Story", Article 14, 25 July 2020, available at: https://www.article-14.com/post/araria-rape-survivor-shares-her-story

[70] V. Venkateshan, "Rehana Fathima's Struggle Against Gender Stereotypes Should Be Celebrated, Not Punished", Wire, 29 July 2020, available at: https://thewire.in/women/rehana-fathima-pocso-body-painting

[71] Anupriya Dhonchak & Namita Bhandare, "Courts' Misogynistic Rules for Rape Survivors", Article 14, 29 June 2020, available at: https://www.article-14.com/post/the-indian-courts-misogynistic-handbook-for-rape-survivors

[72] Veera Mahuli, "Why Is the Home Ministry's Committee on Criminal Law Reform Functioning in Secrecy?", Wire, 9 October 2020, available at: https://thewire.in/law/criminal-law-reform-committee-transparency; Shruti Sundar Ray "The Higher Judiciary's Gender Representation Problem", Article 14, 31 August 2020, available at: https://www.article-14.com/post/the-higher-judiciary-s-gender-representation-problem