Submission to Ministry of Electronics and Information Technology, Government of India, on the draft Information Technology (Intermediary Guidelines) Rules 2018 with special reference to gender-based cyberviolence against women

IT for Change

December 2019

Even as the digital sphere has emerged as a site for new forms of community building, online sexism, misogyny and gender-based cyberviolence are incontrovertibly acknowledged to diminish the ability of women users and gender minorities to have meaningful and equal access to digital spaces.[1] Research suggests that one in ten women has already experienced some form of cyberviolence since the age of fifteen.[2] The threat of violence contributes to the perpetuation of the gender gap in connectivity. In South Asia, for instance, women are *de facto* minorities in the digital sphere, and are 58% less likely than men to use the mobile internet, due to safety and security concerns.[3]

Ranging from cyberstalking to non consensual sharing of intimate images (NCII), gender-trolling and doxxing, the internet has proven to be as treacherous as offline spaces for women and gender minorities to navigate, in their quest for fundamental freedoms. The latest NCRB figures on cybercrimes (2017) show that 245 cases were registered for violation of (bodily) privacy (66E), and 542 incidences of cyberstalking of women/children (354D IPC). Undoubtedly, this is but the tip of the iceberg.

Parallel to the notification of the draft Intermediary Guidelines, two cases have seen developments in respect to intermediary responsibility for content governance:

- In the suo moto petition Re: Prajwala (SMW (Crl) No. 3/2015), which was concerned with the proliferation of rape videos online, the Supreme Court directed the state to expeditiously frame guidelines for eliminating “child pornography, rape and gang rape imageries, videos and sites in content hosting platforms and other applications.”[4]

- The Supreme Court recently transferred to itself a host of petitions relating to social media messaging accountability and traceability for a hearing in January 2020.[5] The issue at stake is compelling Facebook to identify the originator of a message, for the purpose of aiding investigation by

1 https://www.amnesty.org/en/latest/research/2018/03/online-violence-against-women-chapter-1/
2 https://www.livemint.com/Politics/L3okAh2OoXgHlYttSw8ftJ/Online-campaign-takes-on-sexual-harassment-in-India.html
3 https://www.gsma.com/mobilefordevelopment/resources/mobile-gender-gap-report-2019/
4 https://economictimes.indiatimes.com/tech/internet/meity-may-allow-more-time-to-spike-toxic-online-content/articleshow/71073425.cms?from=mdr
5 https://economictimes.indiatimes.com/news/politics-and-nation/sc-to-hear-whatsapp-message-traceability-petitions-in-january/articleshow/71715693.cms?from=mdr

law enforcement agencies. This gains importance not just in cases relating to terrorism and national security but also in the circulation of morphed images, NCII, doxxing, etc. for identifying the perpetrator of such violations to privacy and dignity of women.

IT for Change's empirical research with 881 young women aged 19-23 years across Kerala, Karnataka and Tamil Nadu in 2018-19 revealed that incidents of identity-based violence and sexual harassment in online spaces have become naturalized and normalized.[6] Researchers and practitioners working with women victims/survivors observe that gender-based violence in the digitally mediated context often lacks corresponding provisions in the law. This is because existing taxonomies in the law often fail to capture emerging, and hitherto unseen, forms of violations. For example, the traditional categories of 'heckling' or 'street sexual harassment' cannot adequately capture 'gendertrolling'.[7] In addition to legal reform to account for new forms of cyberviolence, we also require regulation to address the responsibility of internet intermediaries in combating digitally-mediated violence and harms. The proposed Intermediary Guidelines Amendment Rules provide an opportunity to create redressal mechanisms to combat not only misinformation and economic offenses, but also gender-based cyberviolence and sexist hate speech.

The Intermediary Guidelines Rules 2011 were conceived of as a fault-based liability mechanism under legally granted safe harbour exemptions for the limitation of liability as envisioned by Section 79 of the Information Technology Act 2000. Safe harbour provisions around the world have taken inspiration from Section 230 of the US Communications Decency Act.[8] With mass access to the internet and increasing seamlessness in online-offline experiences, the exact role of the intermediary in curating and moderating content is increasingly in the spotlight. The intention behind the offering of "safe harbour" to intermediaries, it must be noted, is not to provide blanket immunity for the content they carry, but to encourage them to take proactive steps to build standards for content governance.

In India, our experience of the decade since the previous IT Amendment Act came into force in 2008 has demonstrated that as the internet is increasingly socialized, the role of platforms becomes more everyday and complex. This means there is a need for changed scope and parameters for self-regulation. However, what we are seeing is that there is an erosion of the democratic fibre of the public sphere, notably politicization and the exclusionary nature of intermediary platforms, reflecting the crisis of self-regulation. That the law must step up to offer a) standards for intermediary self-regulation, and b) avenues for redress, is beyond doubt.

---

6 https://itforchange.net/sites/default/files/1662/Executive_Summary_Born%20digital-Born-free%20.pdf

7 For instance, in Cyber Cell v Yogesh Pandurang Prabhu where the accused had created a fake account with pornographic images and attached the victim's number on the account, the court creatively applied 66E, even though the images uploaded in the name of the victim were not of her. This demonstrates a lacuna in the law, as much as it presents ways in which courts can use existing legal provisions to provide redressal to victims of grievous reputational harms.

8 https://www.huffingtonpost.in/entry/online-harassment-section-230_n_5b4f5cc1e4b0de86f488df86?ri18n=true, "Chris Cox, then a Republican congressman from California, learned about the Stratton-Oakmont v. Prodigy decision from a newspaper on a cross-country flight and thought the decision was unfair. The ruling appeared to say that an internet platform could only be immune from liability if it did not moderate its content at all — not for spam, harassment, pornography or anything else. What kind of internet would that be, he thought."

Key Recommendations:

1. The Intermediary Guidelines Rules, 2018 must respect the fundamental rights guaranteed by the Indian Constitution. Part III of the Constitution defines social and political liberties in the form of fundamental rights, calling upon the state to particularly establish equality before the law and non-discrimination on the grounds of protected identities, including gender, as well as declaring the freedoms guaranteed to citizens and the reasonable restrictions imposed on them. However, the internet is proving to be a treacherous territory for women – a pathway to their self-actualization, yet, paved with the risk of violence and abuse. The inextricable intertwining of the internet and its disruption of social norms and behaviour, calls for a re-articulation of freedoms and their limits, particularly the balance between freedom of expression and freedom from violence online.[9] In this context, a new civil rights framework for the internet, akin to Brazil's *Marco Civil da Internet*, may be an urgent necessity.

2. The overarching notice-and-takedown regime recommended by the Intermediary Guidelines Rules, 2018 lacks a nuanced view of how liability differs according to type of content at issue (vertical differentiation) and according to the kind of function performed by the intermediary (horizontal differentiation).[10] As the Supreme Court of Argentina has observed in 2014, it is important to distinguish between infringing content and manifestly unlawful content, when determining intermediary liability. And for the latter, greater liability must be placed on internet intermediaries. For manifestly unlawful content such as rape videos and child pornography, intermediaries must be bound by a notice-and-takedown regime where they have to take down the offending content on being notified by any user, even when s/he is not an affected party. For all other forms of infringing content (cases of reported non-consensual circulation of intimate images, copyright violation, defamation etc.), a notice-and-notice regime should be adopted to guard against overcensorship. The Brazilian experience with *Marco Civil da Internet*, and Canadian copyright law, shows that rather than a notice-and-takedown regime, a notice-and-notice regime has the advantages that it: 1) provides the intermediary with more information to add context to a request for removal of content, 2) offers the author the opportunity to remove the content themselves, 3) triggers a fault-based liability on the intermediary only when there is a court order for restricting access.

3. The state must define public standards for algorithmic content management to be followed by intermediaries. Though algorithms could be used for flagging potential violating content, the final decision for content take-down should be human-supervised.

9 https://thecdd.wordpress.com/2014/03/28/marco-civil-da-internet-unofficial-english-translation/
10 https://sflc.in/intermediary-liability-20-shifting-paradigm

Draft Rule	Text	Comments
Rule 3(2)(b)	*"Such rules and regulations, privacy policy terms and conditions or user agreement [published by the intermediary] shall inform the users of computer resource not to host, display, upload, modify, publish, transmit, update or share any information that:* *(b) is grossly harmful, harassing, blasphemous, defamatory, obscene, pornographic, paedophilic, libellous, invasive of another's privacy, hateful, or racially, ethnically objectionable, disparaging, relating or encouraging money laundering or gambling, or otherwise unlawful in any manner whatever;"*	This provision has been present since the 2011 iteration of the Intermediary Guidelines. A concern that has been expressed is that under Indian legal jurisprudence, the *vires* of "hateful" "racially, ethnically objectionable" and "disparaging" are unclear, and what is "otherwise unlawful" cannot be gleaned through *ejusdem generis,* due to categorical dissimilarity. We recognize that the internet introduces new categories that either did not exist or do not have correlatives in pre-digital legislation, which necessitates the creation of legal nomenclature that can encompass new harms. Sexist hate speech for instance should be added to the prohibited categories of speech outlined in Rule 3(2)(b)and it should be legally defined. The prevalence of sexist hate speech has led to multiple instances where women journalists and politicians holding public positions are targeted exclusively for their gender. Sexist hate speech and gendertrolling has a chilling effect on women. Our self-administered survey of college students' experience of cyberviolence found that after being attacked, 39% of those women reported having reduced the use of their mobile phone and laptop, 38% deleted their social media accounts, and 26% expressed a fear of posting or sharing content. In the absence of robust frameworks of prevention and redressal of sexist hate speech and gender-based cyberviolence, survivors often tend to practice self-censorship.
Rule 3(5)	*"When required by lawful order, the intermediary shall, within 72 hours of communication, provide such information or assistance as asked for by any government agency or assistance concerning security of the State or cyber security; or investigation or detection or prosecution or*	The demand for traceability of originators for purposes of aiding law enforcement investigation assumes pertinence in the context of bringing perpetrators of gender-based cyberviolence to book. At the same time, the right to redress of victims of cyberviolence must be balanced against the imperative of providing citizens a "zone of privacy" that will protect them against arbitrary interference in the expression of free speech and opinions.[11] In the digital age, the right to

11 https://freedex.org/2018/07/06/uns-chief-freedom-of-expression-monitor-urges-states-companies-to-protect-encryption-online/

	prevention of offence(s); protective or cyber security and matters connected with or incidental thereto. Any such request can be made in writing or through electronic means stating clearly the purpose of seeking such information or any such assistance. The intermediary shall enable tracing out of such originator of information on its platform as may be required by government agencies who are legally authorised."	encryption is thus closely tied to the right to privacy. Where decryption is implicit within a demand for traceability, the UN Special Rapporteur for freedom of speech and expression has endorsed the idea that states are duty-bound to follow the legality, necessity and proportionality principles of international human rights law, and adopt least-restrictive-means in cases where decryption is sought to be pursued.[12] The current prescription contained in Rule 3(5) of "enabling tracing out of originator of information" seems to assume that it is possible "to trace the origin of the message without breaking encryption (of the content)".[13] However, the technical feasibility of facilitating such decryption involving originator information alone has been questioned.[14] In any case, as has been pointed out by digital rights organizations,[15] tracing origins of messages might simply lead investigators to troll armies/chat bots, rather than the actual originators. There is also a likelihood that enforcing traceability will be counteracted by the proliferation of commercial services that allow for masking origin, especially when deployed by actors breaking the law.
Rule 3(7)	*"The intermediary who has more than fifty lakh users in India or is in the list of intermediaries specifically notified by the government of India shall:* *(i) be a company incorporated under the Companies Act, 1956 or the Companies Act, 2013; (ii) have a permanent registered office in India with physical address; and(iii) Appoint in India, a nodal person of contact and alternate senior designated functionary, for 24x7 coordination with law*	In the Re: Prajwala proceedings, the Solicitor General sought the appointment of India-based contact officer and escalation officers for intermediaries providing services in India. The rationale offered for this demand was that this could counter the problem of applicability of Indian jurisdiction over intermediaries providing services here, in the absence of MLATs with the competing jurisdiction. However, mandating a permanent registered office in India with physical address may not fully address the game of "*Lexi Loci Server*" that intermediaries play to evade accountability.[16] In this context, it is imperative that India be a part of global conversations to facilitate international cooperation on cross-jurisdictional

12 https://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=16095&LangID=E
13 https://www.republicworld.com/technology-news/apps/whatsapp-tracability-violation-fundamental-right-privacy.html
14 https://www.medianama.com/2019/08/223-kamakoti-solution-for-traceability-whatsapp-encryption-madras-anand-venkatanarayanan/
15 https://www.medianama.com/2019/08/223-iff-response-kamakoti-submission-traceability-2/
16 https://sflc.in/intermediary-liability-20-shifting-paradigm

	enforcement agencies and officers to ensure compliance to their orders/requisitions made in accordance with provisions of law or rules."	cybercrimes such as the Budapest Convention. Articles 16 to 21 on the securing of electronic evidence and Articles 23 to 35, on jurisdiction, extradition and mutual assistance of the Budapest Convention weave an exemplary framework in this regard.
Rule 3(8)	*(8) The intermediary upon receiving actual knowledge in the form of a court order, or on being notified by the appropriate Government or its agency under section 79(3)(b) of Act shall remove or disable access to that unlawful acts relatable to Article 19(2) of the Constitution of India such as in the interests of the sovereignty and integrity of India, the security of the State, friendly relations with foreign States, public order, decency or morality, or in relation to contempt of court, defamation or incitement to an offence, on its computer resource without vitiating the evidence in any manner, as far as possible immediately, but in no case later than twenty-four hours in accordance with sub-rule (6) of Rule 3.*	Intermediary liability regimes need to make a distinction between infringing content and manifestly unlawful content. The blanket notice-and-takedown mechanism prescribed in Rule 3(8) lacks such a nuanced view. Its second limitation is that by restricting the definition of actual knowledge to court/executive order, it makes timely redress for those subject to harm by such unlawful content very difficult. For manifestly unlawful content such as rape videos and child pornography, intermediaries must be bound by a notice-and-takedown regime where they have to take down the offending content on being notified by any user, even when the user is not an affected party. For all other forms of infringing content (cases of reported non-consensual circulation of intimate images, copyright violation, defamation etc.), a notice-and-notice regime should be adopted, This is superior to a notice-and-takedown mechanism as it fulfills the *audi alteram partem* principle of natural justice. The rule in its current form has two major disadvantages: 1) It does not provide the alleged violator an opportunity for response. 2) It does not provide, for instance, in cases of NCII – where victims may seek immediate solutions to alleviate their trauma – the ease of access to redress mechanisms through a direct interface with intermediaries. A notice-and-notice mechanism will be able to address both these aspects, as New Zealand's experience under the Harmful Digital Communications Act demonstrates.[17] Also, intermediaries must be obligated to report specific data on what content requested to be removed was or was not taken down and on what grounds. This report should be subjected to periodic auditing for transparency by an independent body. This could be

17 http://www.legislation.govt.nz/act/public/2015/0063/latest/DLM5711810.html

		informed by the experience of Germany in placing transparency requirements through the NetzDG law.[18]
Rule 3(9)	*(9) The Intermediary shall deploy technology based automated tools or appropriate mechanisms, with appropriate controls, for proactively identifying and removing or disabling public access to unlawful information or content.*	Algorithmic filtering and other automated tools are a necessity at a time when Google executives point to the fact that 500 hours of content are uploaded on YouTube every minute. There is a consensus in progressive technology and policy communities that algorithms are currently not well-equipped to judge the appropriateness of content. We recommend that although algorithms should be deployed for identifying potentially violating content, algorithms should not take the decision of actioning content. Any action taken upon unlawful content should be based on human discretion. The Guidelines must go beyond requiring intermediaries to deploy automated tools, and also define "appropriate mechanisms" and "appropriate controls". The proposed framing of Rule 9 gives too much amplitude to private parties to decide what is "appropriate", without creating any accountability or protections for lawful uses of technology. The state should define public standards which must be followed by intermediaries who design algorithms for content moderation, taking into account gender-based experiences of violations and abuses online. Once algorithms are deployed, there must also be a dedicated institutional mechanism for public scrutiny and process audit for appropriate application of content related laws. A model for this exists in the form of New York city's algorithmic task force,[19] which has been set up to examine the automation systems, derived from machine learning, data processing or AI techniques, which are used to "make or assist in making decisions concerning rules, policies or actions implemented that impact the public."

18 The NetzDG law requires all platforms receiving more than 100 complaints for carrying "obviously unlawful content" per calendar year to publish biannual reports of their activities.https://policyreview.info/articles/analysis/reading-between-lines-and-numbers-analysis-first-netzdg-reports

19 https://www.theverge.com/2019/4/15/18309437/new-york-city-accountability-task-force-law-algorithm-transparency-automation