Response to the Draft of The Information Technology [Intermediary Guidelines (Amendment) Rules] 2018

January 31, 2019

By **Gurshabad Grover**, **Elonnai Hickok**, **Arindrajit Basu**, **Akriti Bopanna** and **Torsha Sarkar**

with inputs from **Pranesh Prakash**

Reviewed by **Pranav MB**

Research Assistance by **Vedika Pareek** and **Shweta Mohandas**

The Centre for Internet and Society, India

Introduction

The Centre for Internet and Society (CIS) is a non-profit organisation that undertakes interdisciplinary research on internet and digital technologies from policy and academic perspectives. The areas of focus include digital accessibility for persons with disabilities, access to knowledge, intellectual property rights, openness (including open data, free and open source software, open standards, open access, open educational resources, and open video), internet governance, telecommunication reform, digital privacy, and cyber-security. The academic research at CIS seeks to understand the reconfiguration of social processes and structures through the internet and digital media technologies, and vice versa.

With this submission, the Centre for Internet & Society (CIS) would like to respond to the Ministry of Electronics and Information Technology's invitation to comment and suggest changes to the draft of The Information Technology [Intermediary Guidelines (Amendment) Rules] 2018 (hereinafter referred to as the "draft rules") published on December 24, 2018.[1] CIS is grateful for the opportunity to put forth its views and comments.

In this response, we aim to examine whether the draft rules meet tests of constitutionality and whether they are consistent with the parent Act. We also examine potential harms that may arise from the Rules as they are currently framed and make recommendations to the draft rules that we hope will help the Government meet its objectives while remaining situated within the constitutional ambit.

High-level Comments

Below are our high-level comments to the proposed amendments to the Rules under Section 79 of the IT Act.

Need for holistic approach to disinformation

We acknowledge that the intention of the Ministry in planning these amendments, as stated by the Honorable Minister this July in the Rajya Sabha, is to ensure that intermediaries online platforms do not become the venue or conduit for *large-scale Misuse of Social Media platforms and spreading of fake News*.[2]

It is important to qualify that 'disinformation' can be broken down into different categories For example, UNESCO has made the following distinction:

---

[1] Comments/suggestions invited on Draft of "The Information Technology [Intermediary Guidelines (Amendment) Rules] 2018", Ministry of Electronics and Information Technology, 2018, <http://meity.gov.in/content/comments-suggestions-invited-draft-%E2%80%9C-information-technology-intermediary-guidelines>

[2] Ibid.

- *"**Disinformation**: Information that is false and deliberately created to harm a person, social group, organisation or country*
- ***Misinformation**: Information that is false but not created with the intention of causing harm*
- ***Mal-information**: Information that is based on reality, used to inflict harm on a person, social group, organisation or country."*[3]

We feel it is also important to understand that what qualifies as 'disinformation' can be heavily context dependent and solutions need to be able to accurately account for this. To this extent - a broad requirement for platforms to proactively filter unlawful content may be simpler for content such as pornography but would be more complicated for child pornography and vastly more difficult for fake content. This is especially true as emerging doctoring techniques, such as those utilised by "deep fakes", are increasingly indistinguishable from real content and require fact checking and verification to surface if they are or are not real.[4]

We also recognize that disinformation is a complex issue, and as such requires cooperation from multiple stakeholders including government, civil society, industry, the media, law enforcement authorities as well as the public. Similarly, solutions need to be multipronged with technical, legal, and individual components and need to seek to underscore multiple agendas simultaneously including that of cyber security, national security, democratic values, and the protection of human rights. There is also a significant need for research into disinformation in India.

There are a number of provisions in Indian law that can serve as legal tools for the Government in order to penalize disinformation or mal-information. These include Section 505 of the IPC, and if the disinformation is intended to cause communal strife then other provisions such as Sections 290 and 153A of the IPC are also available. The government furthermore has the ability to block content via Section 69A of the IT Act, intercept, monitor, and decrypt communications via Section 69(1) of the IT Act, and monitor and collect traffic data vis Section 69B of the IT Act. Recognizing that there are a number concerns with the Rules issued under that Section that CIS has previously pointed out,[5] we would recommend that the government with the guidance of a court apply these provisions as and when justified.

At the same time, mass public awareness needs to be built around disinformation in order to help curb the spread and societal impact of the same. Watchdog organizations and fact checking organizations such as Boom[6] or Factchecker.in[7] also play an important role in

[3] Journalism, 'Fake News' and Disinformation: A Handbook for Journalism Education and Training, UNESCO, (15th November 2018) <https://en.unesco.org/fightfakenews>
[4] Disinformation on Steroids: The Threat of Deep Fakes, Council on Foreign Relations, (16 October 2018) <https://www.cfr.org/report/deep-fake-disinformation-steroids>
[5] V Kharbanda, Policy Paper on Surveillance in India, The Centre for Internet and Society, (August 2015) <https://cis-india.org/internet-governance/blog/policy-paper-on-surveillance-in-india>; E Hickok, Policy Brief: Oversight Mechanisms for Surveillance, The Centre for Internet and Society, (November 2015) <https://cis-india.org/internet-governance/blog/policy-brief-oversight-mechanisms-for-surveillance>; Prakash, Pranesh. "How Surveillance Works in India." The New York Times, http://india.blogs.nytimes.com/2013/07/10/how-surveillance-works-in-india (2013).
[6] BOOM Live <https://www.boomlive.in/about-us/>

identifying misinformation. Indeed, the government should also focus on enabling mechanisms that verify the authenticity of content as opposed to removing content. Any approach to disinformation must also include robust accountability, oversight, and redressal mechanisms.

The current approach in the Rules places the responsibility of identifying unlawful content as well as the individuals spreading or creating such content fully onto private intermediaries. The Rules also attempt to place blanket and uniform requirements on domestic and foreign intermediaries regardless of function and size. Such a 'one size fits all' framework can risk harming individual freedom of expression and privacy and decentivises smaller intermediaries from setting up platforms as well as foreign intermediaries from operating in India.

Existing Concerns with the Rules

There are a number of concerns that the Centre for Internet and Society (CIS) had raised in 2011 on the draft rules released for consultation[8], and the 2011 Rules that were notified[9]. A number of these concerns still remain and/or have become compounded with the 2018 proposed amendments. We recommend that the following previous recommendations be carried over to the amendment Rules:

- Rule 3(2) makes unconstitutional obligations on intermediaries by compelling them to advise users not to post "unlawful" content that includes "disparaging", "racially, ethnically or otherwise objectionable", "relating or encouraging money laundering or gambling", which are restrictions beyond what is permissible by Article 19(2) of the Constitution.

  Rule 3(2), in placing the aforementioned obligation, also makes no distinction between different types of intermediaries. While these standard obligations may accommodate one type of intermediary, they would not be accommodative of all. For example, an intermediary relying on user-generated content (UGC), would have different terms of use, as opposed an intermediary providing communication services. Forcing umbrella terms of use negates this inherent differentiation, and therefore is impractical.

  It was recommended that Rule 3(2) in its entirety be deleted.

- Rule 3(4) (and now the proposed amendment to it), which compels the intermediary to inform its users that the intermediary has the right to terminate the users' service in case the terms of service are violated, assumes that all intermediaries are websites

[7] FactChecker.in <https://factchecker.in/about-us/>
[8] CIS Para-wise Comments on Intermediary Due Diligence Rules, 2011, The Centre for Internet and Society, (25 February 2011) <https://cis-india.org/internet-governance/blog/intermediary-due-diligence>
[9] Constitutional Analysis of the Information Technology (Intermediaries' Guidelines) Rules, 2011, The Centre for Internet and Society, (16 July 2012) <https://cis-india.org/internet-governance/constitutional-analysis-of-intermediaries-guidelines-rules>

or applications, has no rational nexus with questions of intermediary liability or due diligence to be observed by intermediary for the purpose of protection from liability, and is *ultra vires* the IT Act.

It was recommended that Rule 3(5) of the 2011 Rules, analogous to Rule 3(4) of the current Rules, be deleted.

- Rule 3(5) is ultra-vires Sections 69 and 69B of the IT Act, rules under which already specify a procedure with certain safeguards for agencies to intercept and monitor information held by intermediaries.

  It was recommended that Rule 3(7) of the 2011 Rules, analogous to Rule 3(5) of the current Rules, be deleted.

- Rule 3(10), which mandates intermediaries to report cyber security incidents to the Computer Emergency Response Team (CERT) has no nexus with intermediary liability, and should ideally be a rule issued under Section 70B of the IT Act.

  It was recommended that Rule 3(9) of the 2011 Rules, analogous to Rule 3(10) of the current Rules, be deleted.

- By not having a provision that requires intermediaries to inform users when their content is taken down, draft Rule 3(8) enables an "invisible" form of censorship that may be incompatible with the constitutional requirements of due process and natural justice.

Applicability to intermediaries

The current intermediary guidelines, notified in 2011, and the draft rules make no distinction between the different types of entities that qualify as intermediaries under the law, and thus creates uncertainty as to how these regulations apply to them.

For instance, in the 2011 rules, rule 3(2) compels intermediaries to inform their users to not share or upload certain information. We believe that the intention of the rule is to place the obligation primarily on intermediaries that host third-party content. However, the definition of intermediaries under the IT Act includes service providers which may exert zero or minimal control over the actual content they transmit, such as internet service providers, cyber cafes, content delivery networks and backbone networks. Thus, the rules create confusion as to whether these obligations apply to them equally.

Similarly, the draft rules make certain obligations (for instance, for proactively monitoring content under draft Rule 3(9), or for enabling traceability under draft Rule 3(8), etc.) that are only applicable to intermediaries that host third-party content.

We recommend that instead of adopting a one-size-fit-all approach to intermediary liability, the Government devise a separate definition for intermediaries primarily hosting third-party content, and start a consultation process as to how the obligations would differ for different types of intermediaries.

Unclear scope of the term 'unlawful'

The scope of the term 'unlawful' is undefined and used inconsistently throughout the Rules thus resulting in it potentially being broadly interpreted. It is used first in Rule 3(2)(b), as part of the due diligence duties of the intermediary, in consonance with several other terms which indicate the kind of subject-matter that the intermediary would be obligated to *not* host on its platform. Majority of these terms seem to go beyond the constitutional mandate of Article 19(2). Applying the principle of harmonious internal consistency within statutes, the term 'unlawful' also seems to assume a similar, overreaching context.

The next place where the term occurs is in Rule 3(8). Here, the intermediary is under the obligation, upon receipt of 'actual knowledge' [in lieu of the *Shreya Singhal* judgment], to remove content relating to 'unlawful acts relatable to Article 19(2). The third use of the term, in Rule 3(9), is again in relation to the duty of the intermediary to apply automated technology to remove 'unlawful' content.

These usages render a proper, harmonious reading of the rules difficult. Not only is the term 'unlawful' *not* defined in the Rules, or in the parent Act, its usage in two out of the three instances of its occurrence is overbroad. While the merit of the term 'unlawful' in relation to Rule 3(2)(b) was not explicitly discussed in the *Shreya Singhal* judgment, it would not imply that the acts mentioned in the rule, not overtly struck down by the judgment, continue to be constitutionally valid. Nevertheless, save Rule 3(8), the interpretation of the term violates the dictum of the *Shreya Singhal* judgment, which had laid down that unlawful acts beyond Article 19(2) cannot form part of the section 79.[10]

In relation to the third usage of the term, even if we assume that mandating intermediaries to use automated technology to flag down unlawful content is valid, this still does not lay down the scope of the intermediary's duty in this regard. This also does not define what is meant by unlawful content. The Indian Penal Code, and several other criminal statutes make certain conduct 'illegal' or 'unlawful', but there is no general definition of 'unlawful content'. (For example, even books that are "banned" are not "unlawful content" since there is no provision for declaring them as such: there are provisions for declaring their publication and distribution unlawful and there are provisions for seizing such books.) In other words, what kind of content would the intermediary be obligated to filter using this technology? Would it only be content that relates to unlawful acts as per Article 19(2)? Or would it also include unlawful content as per the interpretation of Rule 3(2)(b)? Or unlawful as per any other law in India? Without any definition, or limiting guidelines to the term therefore, the duties of the intermediaries vis-a-vis its users, and the government is ambiguous.

[10] Shreya SInghal v. Union of India, AIR 2015 SC 1523 (Supreme Court of India).

The *Shreya Singhal* judgment upheld the legal proposition that any restrictions not emanating from Article 19(2) could not find place in Section 79 of the Act and as an extension, it should be refrained from being imbibed under the guidelines rules as well. It was clarified that free speech comprises of three elements: discussion, advocacy and incitement; and however unpopular the former two might be, it is the last that can demand a restriction. There was cognizance of the fact that our constitution does not permit the State to place limits on freedom of speech in order to "*promote general public interest.*"[11] This is applicable for speech regardless of the mode of communication as supported by the precedent in *Ministry of Information & Broadcasting v. Cricket Association of Bengal*[12] case.

These thoughts have also found support in the 2013 report by the Parliamentary Standing Committee on Subordinate Legislation where the Committee stated that the terms in Rule 3(2) that have been defined under others laws should be incorporated in these rules and the undefined ones should be defined. Such a step would ensure that "*no new category of crimes or offences is created in the process of delegated legislation*".[13] Not defining all terms in the Rules is in direct contravention of the Committee's recommendations.

It is also important to note that"information or content" is not made unlawful under Indian laws, whereas specific **acts** are made unlawful. Even books that are "banned" are not "unlawful content", since there is no provision for declaring them as such: there are provisions for declaring their publication and distribution unlawful and there are provisions for seizing such books.

Recommendation

It is recommended that phrases employing the term 'unlawful' to define acts or speech be deleted in all three instances: draft rules 3(2)(b), 3(8) and 3(9).

Excessive delegation of legislative functions

Delegated legislation is a constitutionally accepted means by which the legislature may delegate a component of its function to an external authority[14], which may include an executive authority, such as the Ministry of Electronics and Information Technology (MEiTy) in this case. However, there are entrenched constitutional limitations on the extent of delegation. The legislature cannot delegate essential legislative functions which includes the

[11] Ibid., para 21.
[12] Para 78, The Secretary, Ministry of Information & Broadcasting v. Cricket Association of Bengal & Anr., (1995) SCC 2 161, (Supreme Court of India).
[13] Committee On Subordinate Legislation (2012-2013) (Fifteenth Lok Sabha) Thirty-First Report, Lok Sabha Secretariat, New Delhi, (March 2013) <https://sflc.in/sites/default/files/wp-content/uploads/2013/03/31-Report-_IT_.pdf>
[14] Vishwanathan, T. K. Legislative Drafting Shaping the Law For the New Millennium. p. 441-480 Indian Law Institute, New Delhi, 2015.

determination of legislative policy. They also cannot delegate the power to repeal, modify or alter the scope of an existing law.[15]

In State of Karnataka v. Ganesh Kamath[16] the Supreme Court held that "it is a well settled principle of interpretation of statutes that the conferment of rule-making power by an Act does not enable the rule-making authority to make a rule which travels beyond the scope of the enabling Act or which is inconsistent there with or repugnant thereto". In KSEB v.Indian Aluminium Company[17], it held that "subordinate legislation cannot be said to be valid unless it is within the scope of the rule making power provided in the statute"

As per Indian Express Newspapers Pvt. Ltd. v Union of India[18], a subordinate legislation can be challenged on any grounds that the parent legislation might also be challenged but also be vulnerable if it does not conform to the parent statute or fail to comply with constitutional requirements. Basically, the agency to which authority is delegated is merely supposed to fill in administrative and procedural details for implementation of the law, not re-write or enlarge its scope.

The original section 79 merely states that the intermediary will not be held liable for any information hosted by her if she complies with the requirements as per the law. The draft rules are not limited to implementing the legislative mandate or filling out details, but instead create a host of new obligations on intermediaries (including proactively filtering content and disabling access in a number of cases) that do not pertain directly to the hosting of information or disabling of the same. These obligations have potential consequences for the safeguarding of fundamental rights enshrined in the constitution, which we will discuss throughout the rest of the document. Even if these obligations were to become law, it would have to be through the passing of a new legislation by the Parliament rather than as an executive notification under Section 79 of the IT Act by a Ministry.

Recommendations:
Even if these obligations were to become law, it would have to be through the passing of a new legislation by the Parliament after legislative debate rather than as an executive notification under Section 79 of the IT Act by a Ministry.

Specific Comments

Rule 3(2)(j)

[15] Agricultural Market Committee v. Shalimar Chemical Works Ltd AIR 1997 SC 2502, (Supreme Court of India).
[16] (1983) 2 SCC 40, (Supreme Court of India).
[17] AIR 1976 SC 1031, (Supreme Court of India).
[18] AIR 1986 SC 515, (Supreme Court of India).

"3. (2) Such rules and regulations, privacy policy or user agreement shall inform the users of computer resource not to host, display, upload, modify, publish, transmit, update or share any information that —

(j) threatens public health or safety; promotion of cigarettes or any other tobacco products or consumption of intoxicant including alcohol and Electronic Nicotine Delivery System (ENDS) & like products that enable nicotine delivery except for the purpose & in the manner and to the extent, as may be approved under the Drugs and Cosmetics Act, 1940 and Rules made thereunder;"

Comments

The terms "threaten" or "public health or safety" are not defined under the Rules or in any of the laws referenced by the rules, and are therefore are left open to broad interpretation. Additionally, imposing restrictions on free speech for "public health or safety" interests is not reasonable under Article 19(2), and thus, the draft rule may be deemed unconstitutional.

There are three items whose promotion via an intermediary is prohibited by the draft rules, save as permitted by the Drugs and Cosmetics Act, 1940 (*hereinafter* D&C Act). These are: (i) cigarettes and any other tobacco products; (ii) consumption of intoxicant including alcohol; and (iii) ENDS and similar products.

However, the D&C Act *does not* regulate the promotion/advertisement of cigarettes and tobacco products, nor does it regulate promotion of alcohol. The only relevant matters in this regard under the scope of the Act are the sale of nicotine gum containing up to 2gm of nicotine (as per Chapter IV of the Act) and the regulation of ENDS and like products[19]. If the purpose of this clause is to extend the ban on the advertising of alcohol and tobacco products from television to the online platforms, then the clause should refer to the Rules and Notifications issued under the Cable Television Networks (Regulation) Act, 1995 and the Rules and notifications thereunder. The sub-rule, purporting to regulate online advertisements of the mentioned subject matter, however, does not seem to take into account *any* of the relevant regulations dealing with the same.

Moreover, use of the phrase 'promotion' instead of ' advertising' is over-reaching and therefore a cause for concern.As has been the case, several liquor companies indulge in surrogate advertising for the promotion of their products in the digital media.[20] This goes beyond mere product advertising, and results in in-film branding, association with sports events, hosting competitions and so on[21]. Without any limiting framework to the term

[19] Advisory on Electronic Nicotine Delivery Systems (ENDS) including e-Cigarettes, Heat-Not-Burn devices, Vape, e-Sheesha, e-Nicotine Flavoured Hookah, and the like products, Ministry Of Health & Family Welfare, (28 August 2018) <https://mohfw.gov.in/sites/default/files/ADVISORY%20ON%20ELECTRONIC%20NICOTINE%20DELIVERY%20SYSTEMS%20ENDS.pdf>

[20] Surrogate liquor advertising: Time for change?, Santosh Jangid, (2 October 2017) <http://www.indiantelevision.com/mam/marketing/mam/surrogate-liquor-advertising-time-for-change-171002>

[21] Liquor brands override ad bans by leveraging digital, R Maheshwari & PM Dasgupta, (26 November 2015)

"promotion", the draft rule may result in overbroad interpretations that go beyond standards even laid out by the Advertising Standards Council of India Code[22].

Recommendations

1) The entirety of (j) to be deleted as it does not fall within the limits of Article 19(2).

Rule 3(2)(k)

> *"3. (2) Such rules and regulations, privacy policy or user agreement shall inform the users of computer resource not to host, display, upload, modify, publish, transmit, update or share any information that —*
> *(k) threatens critical information infrastructure."*

Comments

The Government as per S.70 (1) of the IT Act, through its official gazette can notify any resource to be critical information infrastructure if "*the incapacitation or destruction of which, shall have debilitating impact on national security, economy, public health or safety*".

Threatening CII, ostensibly, can be read into the endangering national security. However, the use of the term "threatening" is of concern here, since it is unclear what constitutes threatening and how an intermediary would determine this. Further, the term 'threatening' is inconsistent with section 66F(iii) of the IT Act which, among other things, punishes acts that adversely affect critical information infrastructure and characterizes the same as cyber terrorism.[23] Moreover, section 70(3) of the IT Act already criminalizes unauthorized attempts to access critical infrastructure.

Recommendations

It is recommended that this clause be deleted as threats to critical infrastructure are already addressed through section 66F and 70 of the IT Act.

Rule 3(4)

<https://brandequity.economictimes.indiatimes.com/news/digital/liquor-brands-override-ad-bans-by-leveraging-digital/49923754>

[22] The Code for Self-Regulation of Advertising content in India, The Advertising Standards Council of India (September 2018) <https://ascionline.org/images/pdf/code_book.pdf>

[23] "...and by means of such conduct causes or is likely to cause death or injuries to persons or damage to or destruction of property or disrupts or knowing that it is likely to cause damage or disruption of supplies or services essential to the life of the community or adversely affect the critical information infrastructure specified under Section 70"

"3. (4) The intermediary shall inform its users at least once every month, that in case of noncompliance with rules and regulations, user agreement and privacy policy for access or usage of intermediary computer resource, the intermediary has the right to immediately terminate the access or usage rights of the users to the computer resource of Intermediary and remove noncompliant information."

Comments

This rule states that the intermediary has a duty of informing users that in case of non-compliance with rules and regulations, ToS and privacy policy, the intermediary can *terminate* the usage or access rights of the users. These policies are not directly related to intermediary liability exemptions bestowed by S. 79.

The suggested termination procedure also lacks a notice and appeal requirement. In other words, the intermediary is not obliged to give a notice to the concerned user before terminating the access or usage rights or provide them a mechanism to appeal the decision.

Recommendations

It is therefore recommended that this provision be deleted as account restriction does not directly pertain intermediary liability. If this requirement is included, the intermediary must also be required to provide a procedure of notice that includes the reason for termination to the users, and a procedure of appeal against such termination. We would recommend similar safeguards as those laid out by the Manila Principles for content restriction:

a. *"Before any content is restricted on the basis of an order or a request, the intermediary and the user content provider must be provided an effective right to be heard except in exceptional circumstances, in which case a post facto review of the order and its implementation must take place as soon as practicable.*
b. *Any law regulating intermediaries must provide both user content providers and intermediaries the right of appeal against content restriction orders.*
c. *Intermediaries should provide user content providers with mechanisms to review decisions to restrict content in violation of the intermediary's content restriction policies.*
d. *In case a user content provider wins an appeal under (b) or review under (c) against the restriction of content, intermediaries should reinstate the content.*
e. *Where content has been restricted on a product or service of the intermediary that allows it to display a notice when an attempt to access that content is made, the intermediary must display a clear notice that explains what content has been restricted and the reason for doing so."*[24]

Rule 3(5)

[24] Principle 5 and 6 of the Manila Principles. https://www.manilaprinciples.org/

"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 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."

Comments

On receipt of a 'lawful order', the intermediary is required to provide 'such information' and assistance as asked by 'any government agency'. In practice this provision could permit government agencies to request access to a broad range and large quantity of data held by intermediaries including both metadata and content data and at a lower standard than that mandated under section 5 and associated 419A rules of the Telegraph Act, section 69 and 69B and associated rules of the Information Technology Act, and section 91 and 92 of the CrPC. Further, if the corporations are not located in India, then Mutual Legal Assistance Treaties, and other treaties and laws would be applicable as well. There are four issues of concern here:

Process: First, the exact nature of a lawful order is unclear as is the process by which such order would be issued. It is also unclear which agencies are authorized agencies under the Rules.

Second, the terms 'such information' and 'assistance' are undefined and thus could encompass anything a governmental agency wishes to ask. Further, the grounds for such requests are too broad. For example, "protective or cyber security and matters connected with or incidental thereto" is undefined and is not found in other legal provisions.

Third, there are no clear oversight or review mechanisms as found in section 5 and associated 419A rules of the Telegraph Act, section 69 and 69B and associated rules of the Information Technology Act.

Fourth, the Rule further requires intermediaries to comply with orders for information and assistance within 72 hours. Depending on the size of the organization, location, and complexity of the request - it is unclear that all intermediaries would have the resources or the ability to comply with all orders within the 72 hour timeframe. The Rule also does not provide a procedure for an intermediary to request more time if needed. The pressure that this will place on intermediaries means that in practice they may not undertake the due diligence needed to verify requests and information and assistance shared. Furthermore, India's formal provisions around interception, monitoring, decryption, collection of traffic data, and access to stored information do not place similar timeframes on intermediaries.

Fifth, the Rule does not recognize the MLAT process or recent developments in the modalities of cross-border data sharing such as the US Cloud Act and the ability for the government to use those processes to access information and assistance.

Further, there are several issues with the obligation on intermediaries to enable "tracing out of [...] originator of information".

First, it is unclear what kind of information the intermediaries will have to share with authorized agencies to comply with such requests. The word "tracing" or the phrase "tracing out of [...] originator of information on its platform" are broad enough to include several kinds of information: for instance, it is unclear whether the Government is seeking to (a) provide particular content to an intermediary and request the identity of the creator of the content, or (b) request communication metadata. In either case, there is no specific reason why the information the Government is seeking under "tracing" cannot be provided under the first part of this provision, i.e. information or assistance requests.

Second, in either interpretation, several categories of intermediaries will be technically unable to comply with the traceability requirement. For instance, ISPs transmitting encrypted traffic from a user to a service have no access to its contents or granular information (say final intended recipient of content when the user is communication with an intermediary). In this respect, the word "platform" is used in the rule, but is left undefined. It is unclear whether the draft rule places obligations on just social media platforms and interpersonal messaging services, or all intermediaries as defined by the law. This vagueness has far-reaching implications on the services provided by internet service providers, backbone networks, cyber cafes, content delivery networks, and a host of intermediaries that exert little control over the content they transmit.

Even when we limit ourselves to communication applications, the current phrasing, i.e. "shall enable tracing [...] as may be required by government agencies [...]". This makes it unclear as to whether (a) all intermediaries have to enable "tracing" by default and comply with Government information requests in this regard, or (b) enable "tracing" when asked by the Government. For instance, Whatsapp claims that it does not retain logs (metadata) of delivered messages.[25] If the draft rule is interpreted as (a), then the draft rules force them to retain communication metadata at all times; and if it is (b), then the company only has to retain communication metadata of only certain individuals when requested by the Government.

In this context, it is useful to note that several privacy-preserving applications and software are technically designed to decrease the information available to the service provider. For instance, Signal messenger has a feature called "sealed sender", which prevents the Signal server from knowing the identity of the sender of messages, thus reducing the amount of

[25] Information for Law Enforcement Authorities, Whatsapp <https://faq.whatsapp.com/en/android/26000050/?category=5245250>

communication metadata available to them.[26] The proposed rules create uncertainty as to whether these services are in risk of losing their exemption from liability.

Additionally, tracing of the originator of the concerned information can be done by 'any authorized agency'. So the rule creates a dichotomy between government agencies who can request information and authorized agencies who can request tracing. This dichotomy must be removed and only a list of authorized agencies, priorly notified, must be able to perform either of these functions. It is unclear how this provision works with section 69 and associated rules of the IT Act which enables authorized agencies to request decryption keys from intermediaries.

Recommendations

We would recommend that this provision be deleted, and section 69(1) and 69B of the IT Act, section 5 and 419A rules of the TA, and section 91 and 92 of the CrPc be relied upon for access to information and assistance including traceability. If the information or assistance is required from a foreign intermediary - the MLAT system must be followed. As a note - CIS is cognizant of the challenges in the MLAT system and would also recommend India to start exploring solutions to the MLAT system, including potentially the negotiation of a multilateral data sharing agreement.[27]

We had recommended that India improve its position in diplomatic negotiations with the US by:

Utilising principles of International Law and concrete principles of human rights as a baseline tool for negotiations: Despite the uncertainty in the hierarchy of various permissive principles for extra-territorial jurisdiction, it is clear that Indian jurisprudence recognises these principles. International Law dictates that the hierarchy would need to be determined based on which country has a greater substantial connection to the rime at hand when deciding a conflicts situation. between a country, which is merely storing data as the processor is a company incorporated there and a country where the crime has been committed or whose citizens have been affected, it is clear that the latter would have a more substantive connection. Echoing these principles either in the MLAT agreement or any agreements entered into under the CLOUD Act should reflect this hierarchy. The argument can be made more cogently if these principles are referred to during the negotiations

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;*

[26] Technology preview: Sealed sender for Signal, J. Lund, (29 October 2018) <https://signal.org/blog/sealed-sender/>

[27] A. Sinha, E. Hickok, and Ors., Cross Border Data-Sharing and India: A Study in Processes, Content and Capacity, (27 September 2018) <https://cis-india.org/internet-governance/files/mlat-report>

(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 enforcement agencies and officers to ensure compliance to their orders/requisitions made in accordance with provisions of law or rules.

Comments

Section 79 is an exemption clause relating to intermediary liability; provisions dealing with registration under the Companies Act or having an office in India have no rational nexus with issues of intermediary liability. Thus, these requirements on intermediaries relating primarily to the Companies Act may exceed the scope of the powers of subordinate legislation conferred by the IT Act.

This rule lays down two criteria to identify intermediaries that must maintain a physical office in India, and appoint a nodal officer to work with and respond to requests from law enforcement.: *first,* the number of users and *second*, whether it is list of intermediaries notified by the Government under the rule. As a note, rule 13 of the rules framed under section 69A also require the intermediary to appoint a nodal officer to handle governmental blocking orders.

Unclear requirement of user base: Though it is possible to place requirements on intermediaries based on the size of the user base, it is unclear (i) if this number would encompass all users globally or only the India user base, and (ii) whether this number is the active number of users for a specific period or users registered in entirety. Usually, only the intermediary would be privy to its precise number of users. Thus, to implement this provision, intermediaries would need to be mandatorily required to report their user base on a set schedule. Furthermore, it is unclear how users would be calculated for different types of intermediaries. For example, would the number of "users" for a content delivery network (CDN) be the number of customers they have or the number of end-users they end up serving?

Lack of guidelines for notified list: No mechanism, threshold, or guidelines for the inclusion of intermediaries on the list of notification has been specified, and thus the arbitrariness can be used to target intermediaries that may or may not have the financial standing to maintain a local office in India or support a 24/7 legal team. The cost of incorporating a company or having a permanent registered office in India may also prove to be a deterrent from expanding services in India and stifle innovation and competition. Furthermore, it is unclear why all intermediaries (even those operating services without commercial interests) must register as companies as opposed to another type of entity like a trust.

No distinction between types of intermediaries: By including all intermediaries in its ambit, the draft rule fails to take into account certain intermediaries, such as content delivery networks and backbone networks, that primarily serve a network function and have minimal or zero control over the information that they transmit.

As an additional note: the use of the term ‘law enforcement’ is inconsistent with the term ‘authorized agencies’ used in other provisions in the Rules. Both of these terms - "law enforcement" and "authorized agencies" - leave the question of "who is authorized" unaddressed, leaving intermediaries guessing. Furthermore, the rules do not make any provisions for notifications listing out authorized agencies. Thus, the phrase "authorized agencies" is vague by talking of "authorized" without specifying how one is to recognize which agencies are "authorized" or by whom or under what law.

Recommendations

We recommend that draft rule 3(7) be deleted in its entirety as it exceeds the scope of delegated legislation permissible under Section 79. The nodal person already available to the Government under Section 69A could act as the contact for authorized agencies to seek the assistance of intermediaries for law enforcement purposes.

To achieve the Government’s stated objectives, we recommend exploring comprehensive legislation that recognizes the different kinds of intermediaries such as Internet Service Providers, search engines, social networks, content aggregators, etc. and accord responsibility (perhaps even incorporation and physical registration), if at all, on the basis of this differentiation. 2) For certain categories of intermediaries, formulate a criteria based on user size and annual turnover to determine whether or not an intermediary needs to maintain a local office, if at all. 3) Formulate principles by which exceptional cases could be taken into consideration by the government. We recommend that the Government start a consultation process to formulate legislation with the briefly-summarised framework we present here, to which CIS will be happy to provide detailed inputs and specific recommendations.

Rule 3(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.
Further the intermediary shall preserve such information and associated records for at least one hundred and eighty days for investigation purposes, or for such longer period as may be required by the court or by government agencies who are lawfully authorised.

Comments

This provision requires intermediaries to comply with court and governmental orders that are within the ambit of 19(2) of the constitution within 24 hours and extends the data storage

period by entities from 90 days to 180 days or as required by the Court or lawfully authorized government agencies. There are a number of concerns with this provision:

Firstly, this draft rule is in direct contravention of the Supreme Court's decision in *Shreya Singhal v. Union of India* which held that "actual knowledge" is only said to be accrue to the intermediary when it is informed of a court order or under asking it remove certain content.

Secondly, Though short time frames to comply with orders is a trend that a number of governments are adopting globally[28] research has yet to show the effectiveness of these timeframes, but research has demonstrated that it is extremely difficult for intermediaries to comply with all requests within 24 hours and still maintain a level of due diligence from their side.[29] As a note section 69A and associated rules do not place a similar time frame on intermediaries to comply with governmental orders, instead Rule 11 requires that intermediaries act 'expeditiously' but no later than seven days[30] and Rule 13 requires intermediaries to acknowledge the order within two hours of receiving the same.[31]

Additionally, the proviso that mandates the intermediary to preserve records for investigation purposes for 180 days does not specify the process for the extension of the retention period, nor does it make it clear who "lawfully authorised" agencies are, or under what law they need to be authorised.

Recommendations

We recommend that:

- The text "or on being notified by the appropriate Government or its agency" should be replaced with "or on being notified by the appropriate Government or its agency about a valid court order". A process for the government to issue such orders from a court to intermediaries should be established. This could be the same process as established under section 69A and associated Rules of the IT Act.
- The proviso "Further the intermediary shall preserve such information and associated records for at least one hundred and eighty days for investigation purposes, or for such longer period as may be required by the court or by government agencies who

[28] For example: 1) NetzDG gives 24 hours to remove content that is 'obviously illegal' and seven days for 'illegal' content. 2) DMCA does not have a particular time-frame, but research shows that the time period is somewhere in between 24-72 hours. 3) EU's code of conduct on countering online hate speech has a time-frame that is less than twenty four hours.

[29] V Munjal, A March towards Digitization, PSA E-Newsline, (December 2017) <http://www.psalegal.com/wp-content/uploads/2017/01/E-Newsline-December-2017.pdf>; A. Mohanty, An Open Letter to Kapil Sibal on Copyright and Free Speech, SpicyIP (18 May 2012) <https://spicyip.com/2012/05/dear-mr-sibal-youve-got-it-all-wrong.html>; S. Pathak, Information and Technology (Intermediaries Guidelines) Rules 2011: Thin Gain with Bouquet of Problems <http://docs.manupatra.in/newsline/articles/Upload/269ED933-8F47-4EB3-A6C3-DA326C700948.pdf>

[30] "11. Expeditious disposal of request.--
The request received from the Nodal Officer shall be decided expeditiously which in no case shall be more than seven working days from the date of receipt of the request."

[31] "(2) The designated person of the Intermediary shall acknowledge receipt of the directions to the Designated Officer within two hours on receipt of the direction through acknowledgement letter or fax or e-mail signed with electronic signature."

are lawfully authorised" should be modified to "Further the intermediary shall preserve such information and associated records for at least one hundred and eighty days for investigation purposes, or for such longer period as communicated to the intermediary through a court order."

- A procedure for the intermediary to challenge the notification should be established.
- Notification that results from ex-parte hearings should be challengeable by any interested party.
- Notifications should be published on a website like: accessremoval.meity.gov.in to allow for transparency, and so that such notifications may be appropriately challenged through an established legal framework.
- We recommend that the 24 hour timeframe is removed and instead, as in 69A and associated Rules, intermediaries be required to acknowledge receiving the order and act 'expeditiously'.
-

Rule 3(9)

"3.(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."

Comments

We have two broad sets of concerns regarding this draft rule. They can be classified as under:

1. **Constitutional and legal concerns**
    a. Vagueness and inaccuracy in the language of the provision
    b. Inappropriate delegation of a state's duty to a private actor
    c. Violation of the right to freedom of speech and expression under Article 19 of the Constitution, and international human rights laws that India is bound by
    d. Similar laws in Europe which have been criticised on grounds of violating the ICCPR, the Universal Declaration of Human Rights, and similar Europe-level human rights instruments
2. **Practical and technical concerns**
    a. Accuracy of automated technologies such as big data analytics and Artificial Intelligence
    b. Costs and sustainability of deploying automated technologies
    c. Accountability and oversight of decisions taken by automated technologies

1. Constitutional

(a) Vagueness in the language of the provision

In *Kartar Singh v. State of Punjab*[32], the Supreme Court held that as a basic principle of legal jurisprudence, an enactment is void for vagueness if the prohibitions it imposes are not

[32] (1994) 3 SCC 569 (Supreme Court of India).

clearly defined. Laws should give a person of ordinary intelligence a reasonable opportunity to know what is prohibited, so that he may act accordingly as vague laws are subject to manipulation and might not give fair warning to the innocent.

The wording of Rule 3(9) fails this test due to the absence of the definition of certain key terms. For example the phrase *"unlawful information or content"* is undefined. While "information" is defined in Section 2(1)(j), "unlawful" is not defined in the IT Act, 2000 or the draft rules. Further, there is no definition of 'automated technology' that might be used by the intermediary or definition of 'appropriate controls' and there is an absence of guidelines on the timelines imposed on the intermediary to take down the content or further information on a process that might be followed in pursuance of such removal or for appeals (automated or otherwise) for such automated removals.

As highlighted in the high-level comments above, it is also important to note that "information or content" is not made unlawful under Indian laws, whereas specific acts are made unlawful.

(b) Inappropriate delegation of a state's powers to a private actor

Shifting the burden of adjudicating what is 'unlawful' content onto a technology developed or procured by the intermediary is against the constitutional mandate of *Shreya Singhal*. The legislature cannot do indirectly what it cannot do directly.[33] This goes specifically against the interpretation given to section 79 by the Supreme Court in *Shreya Singhal*, *viz*. "*Section 79(3)(b) has to be read down to mean that the intermediary upon receiving actual knowledge that a court order has been passed asking it to expeditiously remove or disable access to certain material must then fail to expeditiously remove or disable access to that material.*" Further, the Supreme Court also stated that "The intermediary applying its own mind to whether information should or should not be blocked is noticeably absent in Section 69A read with 2009 Rules."[34] Therefore, since the section under which these Rules are issued itself has been qualified by the requirement of a court or governmental order, the Rules cannot revive the obligation to remove content in any manner other than through a court.

Further, by unconstitutionally delegating an act that could have potential implications for the freedom of expression to a private actor, the state is indirectly avoiding its responsibilities under Part III of the Constitution and shifting the same to a private actor. It was clearly stated in *Hamdard Dawkhana* that the legislature cannot do indirectly what it cannot do directly.[35] Whenever a government body performs a 'public function,' they are subject to the entire gamut of fundamental rights, which include the substantive and procedural due process requirements in Article 21, the Right to Equality in Article 14 and the Freedom of Speech and Expression in Article 19. Any individual is entitled to file a writ petition against the state for violation of its fundamental rights. However, judicial precedent on the horizontal application of fundamental rights has still not been clearly delineated. This effectively means that any individual whose content has been arbitrarily removed by the intermediary has no constitutionally viable means of enforcing her fundamental right as the specific act of identifying and evaluating the content as illegal and subsequently taking down

[33] Hamdard Dawakhana v. Union of India, 1960 AIR 554 (Supreme Court of India).
[34] Shreya Singhal, para 116
[35] Hamdard Dawakhana v. Union of India, 1960 AIR 554 (Supreme Court of India).

the material has not been done by the state. As effectively articulated by Seth Kreimer, expert on constitutional law at the University of Pennsylvania, this form of delegation effectively amounts to 'censorship by proxy.'[36]

It is also vital to note that legally requiring private actors to make determinations regarding content restriction, can often lead to over-enforcement as the intermediary is incentivised to err on the side of taking down content in order to avoid expensive litigation.[37] A study conducted by Rishabh Dara at CIS demonstrated this in the Indian context as it was found that six out of the seven intermediaries who were sent flawed take-down notices by private parties over complied even in cases where the notice had some debilitating flaws.[38] This could have a high social cost and an indirect chilling effect on the freedom of expression online, which is compounded by the information asymmetry that exists because the user continues to remain unsure about the process, reasoning and oversight that went into the takedown. As we discuss below,these concerns can become further compounded when the decision is taken by an automated tool without human oversight or intervention.

(c) Violation of the constitutional guaranteed right to freedom of speech and expression under Art. 19

The transgression of constitutionally guaranteed standards of free speech and expression commences with the use of the word 'unlawful'. As we discussed previously in the beginning of this submission**,**the use of the word "unlawful" in Section 79(3)(b) of the IT Act was challenged in Shreya Singhal on the grounds that it goes beyond the restrictions delineated in Article 19(2) of the Constitution. The Supreme Court clarified that "unlawful acts" which do not fit under one of reasonable restrictions to the freedom of speech and expression laid down in Article 19(2) cannot form any part of Section 79, and also read down Section 79(3)(b) on those grounds.[39]

As we discussed at the beginning of this submission,the restriction can only be incorporated through new legislation. **Further, whether the restriction is reasonable or not should be determined on a case-by-case basis. [40]This should be done to ensure that the "practical results" of such actions are duly considered before imposing disproportionate restrictions.**

(d) Lessons from International Law and Europe

---

[36] Kreimer, Seth F. "Censorship by proxy: the first amendment, Internet intermediaries, and the problem of the weakest link." U. Pa. L. Rev. 155 (2006): 11.

[37] Kraakman, Reinier H. "Gatekeepers: the anatomy of a third-party enforcement strategy." Journal of Law, Economics, & Organization 2, no. 1 (1986): 53-104.,., Lee, D. "Germany's NetzDG and the Threat to Online Free Speech." (2018). <https://law.yale.edu/mfia/case-disclosed/germanys-netzdg-and-threat-online-free-speech>.

[38] Dara, Rishabh. "Intermediary Liability in India: Chilling Effects on Free Expression on the Internet." (2011) <http://cis-india.org/internet-governance/intermediary-liability-in-india>.

[39] P. 117, 119, Shreya SInghal v. Union of India, AIR 2015 SC 1523 (Supreme Court of India).

[40] State of Madras v. V G Row [1952] SCR 597 (Supreme Court of India).

Laws like the NetzDG[41], or the 'fake news' law in France[42], mandate that the intermediary take down content that is 'manifestly' illegal. The NetzDG has attracted immense criticism from civil society activists. David Kaye, who is the UN Special Rapporteur on freedom of expression penned an open letter to the government of Germany arguing that the vague and ambiguous criteria used in the law is incompatible with Article 19 of the ICCPR which guarantees the right to freedom of expression.[43] Permissible restrictions on the internet should be judged on the same parameters as those offline.[44]

Indeed, under article 19(3) of the ICCPR which has been signed and ratified by India, restrictions on the right to freedom of speech and expression must be

1. Provided by Law: It is not sufficient if the restriction on the freedom of expression is formally enacted as domestic law. They must also be sufficiently, clear, accessible and predictable-something that the present guidelines are not due to the presence of vague and ambiguous terms.

2. Necessary for the rights and regulations of others: This incorporates an assessment of proportionality of the restrictions which should have the objective of ensuring that these restrictions " targets a specific objectives and do not unduly intrude upon the rights of targeted persons."[45] The interest being intruded upon must also be the least intrusive means possible. WIthout considering and undertaking extensive research and pilot projects on alternative means available to curb the 'fake news' or disinformation issues, the NetzDG, like Rule 3(9) violates the ICCPR.

2. Practical and Technical Concerns

(a) Accuracy of automated technologies such as big data analytics and Artificial Intelligence

The draft rule has suggested that automated technologies be used to conduct this filtering. It has been widely argued that automated technologies are inappropriate for conducting filtering as it lacks the human judgement to assess context. Further, outsourcing filtering to Artificial Intelligence driven technologies come replete with the problems to endemic to AI.

---

[41] E. Douek, Germany's Bold Gambit to Prevent Online Hate Crimes and Fake News Takes Effect, (31 October 2017) <https://www.lawfareblog.com/germanys-bold-gambit-prevent-online-hate-crimes-and-fake-news-takes-effect>

[42] M.R. Fiorentino, France passes controversial 'fake news' law, (22 November 2018) <https://www.euronews.com/2018/11/22/france-passes-controversial-fake-news-law>

[43] Mandate of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression (1 June 2017), <https://www.ohchr.org/Documents/Issues/Opinion/Legislation/OL-DEU-1-2017.pdf>

[44] Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Frank La Rue (A/HRC/17/27) <https://www2.ohchr.org/english/bodies/hrcouncil/docs/17session/A.HRC.17.27_en.pdf>.

[45] General comment No. 34, United Nations ICCPR (CCPR/C//GC/34) (12 September 2011) <https://www2.ohchr.org/english/bodies/hrc/docs/gc34.pdf>.

In a previous report[46], we had documented the possible sources of discriminatory decisions that may come with any decision made by Artificial Intelligence. The same systemic issues apply in this case. These include

1.Incomplete or inaccurate training data

The data being used for creating training data sets in the case of pro-active filtering might be incomplete or not reflect lacunae in the data collection process.This issue is most acute in the case of supervised learning systems that require labelled data sets, which proactive filtering mechanisms such as the one recommended in this rule would require.[47] As the labelling of datasets in new contexts, it is likely that the intermediary may use readily available sets that might not provide the complete picture. For example, many natural language processing systems use readily available training datasets from leading western newspapers, which may not be reflective of speech patterns in different parts of the world. A similar automated tool deployed for pro-active filtering by intermediaries raises similar concerns.[48]

For example, there is a growing body of research on the use of automated tools for copyright enforcement and the problems that arise with their use. Research has shown that the use of Digital Rights Management (DRM) systems can have wide sweeping impact on free speech and on fair use.[49]It has been stated that enforcement algorithms work on rules set in code created by programmers, which are distinct from laws are a made and interpreted.[50] Hence these tools might be programmed to remove infringing content but these tools lack the nuance to understand the context and verify whether the use comes under the fair use principle or if they are licensed.[51]There have been multiple cases where these systems have taken down content that were in protected under fair use.[52] Additionally, with the safe harbour provisions for the intermediaries to proactively remove infringing content it was observed that the intermediaries are at times using this as an excuse to over regulate, there

[46] A. Basu, E. Hickok, Artificial Intelligence in the Governance Sector in India, (14 September 2018) <https://cis-india.org/internet-governance/ai-and-governance-case-study-pdf>

[47] Danks, David, and Alex John London. "Algorithmic bias in autonomous systems." In Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, pp. 4691-4697. 2017; Discussion Paper on National Strategy for Artificial Intelligence | NITI Aayog | National Institution for Transforming India. (n.d.) <http://niti.gov.in/content/national-strategy-ai-discussion-paper>.

[48] D. Keller, Problems With Filters In The European Commission's Platforms Proposal, (5 October 2017) <http://cyberlaw.stanford.edu/blog/2017/10/problems-filters-european-commissions-platforms-proposal>

[49] Bar-Ziv, Sharon, and Niva Elkin-Koren. "Behind the Scenes of Online Copyright Enforcement: Empirical Evidence on Notice & Takedown." Conn. L. Rev. 50 (2018): 339.

[50] Perel, Maayan, and Niva Elkin-Koren. "Black Box Tinkering: Beyond Disclosure in Algorithmic Enforcement." *Fla. L. Rev.*69 (2017): 181.

[51] Depoorter, Ben, and Robert Kirk Walker. "Copyright false positives." *Notre Dame L. Rev.* 89 (2013): 319.. Where the example was given how the online broadcast of Neil Gaiman's acceptance speech was disrupted because the DRM software flagged the images from Doctor Who to be copyright infringement, even though the images were licensed for the use during the awards.

[52] Bar-Ziv, Sharon, and Niva Elkin-Koren. "Behind the Scenes of Online Copyright Enforcement: Empirical Evidence on Notice & Takedown." *Conn. L. Rev.* 50 (2018): 339.

are multiple examples of content that were taken down on grounds of copyright enforcement.[53]

2. Algorithmic Processing

An AI driven solution is an amorphous process-such as the 'risk profile' of an individual or the 'suspicious nature' of certain kinds of speech. While human may not be able to assess vast tracts of data to undertake the amorphous task of pro-active filtering, using source code enables a machine to do so. Through it's hidden layers, the machine generates an output, which corresponds to assessing the risk value of an individual, or in the case of pro-active filtering, certain forms of speech. Rouvroy further chastises 'algorithmic governmentality'-a phenomenon that ignores the subjective forms of speech and the embodied speaker. It reduces speech to quantifiable values-sacrificing inherent facets of dignity-such as their unique singularities,personal psychological motivations and intentions.[54]

A further problem with algorithmic processing comes at the stage of developing the technology as the human monitoring the trial-runs and incorporating the results into the decision trees might suffer from some pre-existing sources of bias.[55] Facebook, Twitter, Youtube all have used machine learning to to detect certain content on their platforms.[56] Google has also publicly committed to use machine learning algorithms to fight terrorism-related content.[57] Such techniques and commitments have in part arisen out of the government pressure or mounting number of content-takedown requests around the world (as the Transparency Reports of many of these intermediaries suggest) as well as the growing size of user generated content and user base. However, these have been their own commitments as opposed to compliance with governmental mandates to deploy automated techniques.[58] Usage of these tools also have had mixed results most of the time. While some have said that the tool has been useful in filtering out terrorist related content and spam, the same can not be said with hate speech[59], or adult content.[60]

---

[53] For example YouTube facilitated the removal of a documentary film, India's Daughter, based on the gang rape of a twenty-three-year-old student, the screening of which was banned in India due to copyright infringement allegations.YouTube also allowed the censorship of the satirical show Fitnah when it complied with DMCA takedown notices sent by the primary, state- funded Saudi TV channel, "Rotana." See. Perel, M.; Elkin-Koren, N. (2016). Perel, Maayan, and Niva Elkin-Koren. "Accountability in Algorithmic Copyright Enforcement." *Stan. Tech. L. Rev.* 19 (2015): 473.

[54] Rouvroy, Antoinette. "The end (s) of critique: data behaviourism versus due process." In *Privacy, Due Process and the Computational Turn*, pp. 157-182. Routledge, 2013.

[55] M. Sears, AI Bias And The 'People Factor' In AI Development, Forbes (13 November 2018) <https://www.forbes.com/sites/marksears1/2018/11/13/ai-bias-and-the-people-factor-in-ai-development/#1dfa830c9134>

[56] G. Rosen, F8 2018: Using Technology to Remove the Bad Stuff Before It's Even Reported Facebook Newsroom, (2 May 2018) <https://newsroom.fb.com/news/2018/05/removing-content-using-ai/>, How Content ID works, Youtube Support <https://support.google.com/youtube/answer/2797370?hl=en>, D. Harvey, D. Gasca, Serving healthy conversation, Twitter Blog, (15 May 2018) <https://blog.twitter.com/official/en_us/topics/product/2018/Serving_Healthy_Conversation.html>

[57] Content Regulation in the Digital Age Submission to the United Nations Human Rights Council, Special Rapporteur for Freedom of Expression (June 2018) <https://www.ohchr.org/Documents/Issues/Opinion/ContentRegulation/Witness.pdf>

[58] J. Vincent, Why AI isn't going to solve Facebook's fake news problem, The Verge, (5 April 2018) <https://www.theverge.com/2018/4/5/17202886/facebook-fake-news-moderation-ai-challenges>

[59] B. Dickson, The challenges of moderating online content with deep learning, TechTalks, (10 December 2018) <https://bdtechtalks.com/2018/12/10/ai-deep-learning-adult-content-moderation/>

Respecting individual autonomy means, at the very least, ensuring that users have knowledge, choice and control. Pervasive and hidden AI applications that obscure the process of content display, personalisation, moderation and profiling and targeting can undermine the ability of individuals exercise their right of freedom of opinion, expression and privacy. [61]

(b) Costs and sustainability of deploying automated technologies

To assess the scale and sustainability of any initiative, we need to look both into financial costs and extent of disruption the proposal causes to existing business processes.So far, application of automated technology to filter/monitor content on social media platforms, has only been undertaken by the largest companies, with large-scale resources acting as the prerequisite.[62] In light of this, mandating resort to these tools would be problematic because

- The research on the proper implementation of this technology remains incomplete
- Presumably (if the mixed results from the big companies is any indication), the resources and scale required for the smaller intermediaries to work this technology would be unreasonably high and unprofitable for their overall business.

Second, the requirement to "proactively" identify and remove "unlawful" content is technically impossible for certain intermediaries, such as ISPs, including Whatsapp transmitting encrypted traffic and interpersonal communication platforms which offer end-to-end encryption and would necessitate a rehaul of of their business practices and security protocols.

(c) Accountability and oversight of decisions taken by automated technologies

We accept that bias would exist if any decision outsourced to an algorithm were undertaken by a human being. The key difference between that and discrimination by AI lies in the ability of other individuals to compel the decision-maker to explain the factors that lead to the outcome in question and testing its validity against principles of human rights. A defining feature of Artificial Intelligence is the algorithmic 'black box' that processes inputs and generates usable outputs.[63] Ensuring accountability is an imperative that is challenging when the "values and prerogatives that the encoded rules enact are hidden within black boxes." However, given the metaphorical 'black box' that converts inputs into examinable outputs,

---

[60] H. Bergstrom, Should Artificial Intelligence Be Used to Moderate Online Content?, Diplomatic Courier, (12 December 2018) <https://www.diplomaticourier.com/2018/12/12/should-artificial-intelligence-be-used-to-moderate-online-content/>

[61] Promotion and protection of the right to freedom of opinion and expression (A/73/348) <https://undocs.org/A/73/348>.

[62] Content Regulation in the Digital Age Submission to the United Nations Human Rights Council, Special Rapporteur for Freedom of Expression (June 2018) <https://www.ohchr.org/Documents/Issues/Opinion/ContentRegulation/Witness.pdf>.

[63] Pasquale, F. (2015). The black box society: The secret algorithms that control money and information. Harvard University Press.

implementing workable accountability and evaluation standards for algorithms engaging in pro-active filtering remain a challenge.

The following gaps in accountability would exist if automated pro-active filtering by intermediaries were to be enabled:

- The reasoning and process followed in developing the algorithm
- The time limits, reasoning and process followed by the human beings on the moderation team in response to algorithmic output
- Appropriate avenues and processes for appeals and grievance redressal

Recommendations

We recommend that this provision be deleted in its entirety. There is a dire lack of research on the potential impacts of using automated technologies for pro-active filtering. We have outlined the adverse legal and societal impacts that this technology may have-all of which have been documented above. We also recommend that there must always be a human moderator taking the decision unless concrete research emerges showing that automation and the consequent creation of 'black-boxes' can generate more accurate and equitable patterns. We recognize that human moderation may not be able to keep up with the pace of discourse on social media and may be inaccurate but we hope that the mechanisms detailed below along with robust reinstatement systems providing clearer notification when content is removed and the reasons underpinning said removal.

We recognize, however, that the spread of fake news and misinformation via platforms needs to be curbed. There are three potential alternatives that might be considered, even though they are replete with potential concerns. Therefore, we recommend them as potential areas for research for government, civil society and industry, rather than as suggestions for implementation:

User-filtering:
As per a paper written by Ivar Hartmann advocating for this method, user filtering is a process that can be used for gatekeeping as it concerns the control of information flow.[64] In some ways, it re-configures power dynamics as the 'gated' become 'gatekeepers.' [65] Essentially, this decentralized process of filtering exists in a scenario where the users of an online platform collectively accomplish an objective that regulate the flow of information. Users collectively agree on a set of standards and general guidelines for filtering. [66]Rough consensus or 'incompletely theorized agreements' where users agree on a set of (relative)

[64] I. A. Hartmann, Let The Users Be The Filter? Crowdsourced Filtering To Avoid Online Intermediary Liability, <http://blogs.oii.ox.ac.uk/ipp-conference/sites/ipp/files/documents/IPP2014_Hartmann.pdf>
[65] Karine Barzilai-Nahon, Toward a Theory of Network Gatekeeping: A Framework for Exploring Information Control, 59 Journal of The American Society For Information Science and Technology 1493, 1496 (2008).
[66] I. A. Hartmann, Let The Users Be The Filter? Crowdsourced Filtering To Avoid Online Intermediary Liability, <http://blogs.oii.ox.ac.uk/ipp-conference/sites/ipp/files/documents/IPP2014_Hartmann.pdf>

particulars rather than a set of (relative) abstractions can promote coordination even among users that have widely disparate ideologies, convictions and identities.[67]

In addition to the potential fetters to achieving this 'incompletely theorized agreements,' Hartmann himself acknowledges two potential drawbacks of user-filtering:

1. **Incentives to engage in filtering:** This is linked to the problems of coordination. All users engaging in the filtering have a set of personal values that may not necessarily be shared. While clearly objectionable content such as child pornography, filtering certainly becomes more challenging in the context of hate speech.It remains to be seen how far community-centric standards can deal with this issue.
2. **Potential for over-filtering:** Hartmann conceives the possibility that as the power dynamics shift and users are given more power, they may apply stricter standards and filter more content. He cites the example of mothers who mobilized against the posting of breast-feeding pictures.[68]

In addition, in the user-filtering model, the issue of appropriate appeal and grievance redressal mechanisms also crops up. Legally valid mechanisms that can enable aggrieved persons to challenge take-down decisions must be conceptualized.

Self-Regulation

This would require conceptualizing a scenario where status quo continues and intermediaries regulate speech on their platforms, as Google and Facebook have been doing. This has its disadvantages as it effectively grants autonomy to intermediaries, who are large business corporations and might incorporate self-regulation as part of their business strategy calculus as opposed to an independent societal prerogative.

Ghonim and Rashbass have indicated three ways in which self-regulation might be made more transparent and accountable[69]:

1. The platform must publish all data related to all public posts so that the consumer is made aware of reach-both geographic and demographic and how a story attained 'viral' or 'trending' status.
2. They should publish the intricate details of their content regulation policies-including processes followed, hierarchies in the decision-making process followed, the substantive parameters involved and points-of-contact for grievance redressal.
3. Even if implemented effectively points 1 and 2 may not enable the public to keep pace with the existence and dissemination of posts on social media.[70] Therefore, Ghonim and Rashbass suggested that all platforms should develop an Algorithm Programming Interface (API) or 'Public Interest Algorithms' that capture the relevant

[67] Sunstein, Cass R., Incompletely Theorized Agreements (1995). Harvard Law Review, Vol. 108, No. 7, p. 1733, 1995.

[68] Emil Protalinski, Breastfeeding women protest outside Facebook offices. Available at: http://www.zdnet.com/blog/facebook/breastfeeding-women-protest-outside-facebook-offices/8673 (last visited Apr 26 2012)

[69]https://www.washingtonpost.com/news/democracy-post/wp/2017/10/31/its-time-to-end-the-secrecy-and-opacity-of-social-media/?utm_ter&utm_term=.6b7ef451d550

[70]https://thewire.in/tech/beyond-twitter-russia-make-social-media-incorporated-work-democracy

inputs and outputs used by the platform and make their data public so it may be easily consumed by the public.[71]

Co-Regulation: Models for multi-stakeholder co-operation on developing frameworks, standards and best practices for combating the issues that come with the use of social media in India today might be a useful starting point. The outcome may result in an universal code that guides a combination of self-regulation and user-centric filtering or in informal modes of cooperation. Either way, it is worth pursuing as a potential future research agenda.

[71]https://www.washingtonpost.com/news/democracy-post/wp/2017/10/31/its-time-to-end-the-secrecy-and-opacity-of-social-media/?utm_ter&utm_term=.6b7ef451d550