A deep dive into content takedown timeframes

Inputs from stakeholders and a survey of regulations to inform India's content takedown regulation

30 November, 2019

By **Torsha Sarkar**
Research assistance by **Keying Geng** and **Merrin Muhammed Ashraf**
Edited by **Elonnai Hickok, Akriti Bopanna, Gurshabad Grover**
With inputs from **Tanaya Rajwade**

The Centre for Internet and Society, India

Table of contents

Executive Summary

Since the 1990s, internet usage has seen a massive growth, facilitated in part, by growing importance of intermediaries, that act as gateways to the internet. Intermediaries such as Internet Service Providers (ISPs), web-hosting providers, social-media platforms and search engines[1] provide key services which propel social, economic and political development.[2] However, these developments are also offset by instances of users engaging with the platforms in an unlawful manner. The scale and openness of the internet makes regulating such behaviour challenging, and in turn pose several interrelated policy questions.

In this report, we will consider one such question by examining the appropriate time frame for an intermediary to respond to a government content removal request. The way legislations around the world choose to frame this answer has wider ramifications on issues of free speech and ease of carrying out operations for intermediaries. Through the course of our research, we found, for instance:

1. An one-size-fits-all model for illegal content may not be productive. The issue of regulating liability online contain several nuances, which must be considered for more holistic law-making. If regulation is made with only the tech incumbents in mind, then the ramifications of the same would become incredibly burdensome for the smaller companies in the market.[3]
2. Determining an appropriate turnaround time for an intermediary must also consider the nature and impact of the content in question. For instance, the Impact Assessment on the Proposal for a Regulation of the European Parliament and of the Council on preventing the dissemination of terrorist content online cites research that shows that one-third of all links to Daesh propaganda were disseminated within the first one-hour of its appearance, and three-fourths of these links were shared within four hours of their release.[4] This was the basic rationale for the subsequent enactment of the EU Terrorism Regulation, which proposed an one-hour time-frame for intermediaries to remove terrorist content.
3. Understanding the impact of specific turnaround times on intermediaries requires the law to introduce in-built transparency reporting mechanisms. Such an

---

[1] Article 19, 'Internet Intermediaries: Dilemma of Liability' (2013) <https://www.article19.org/data/files/Intermediaries_ENGLISH.pdf> accessed 31 October 2019

[2] OECD, 'The Role of Internet Intermediaries in Advancing Public Policy Objectives' <https://read.oecd-ilibrary.org/science-and-technology/the-role-of-internet-intermediaries-in-advancing-public-policy-objectives_9789264115644-en#page5> accessed 31 October 2019

[3] Owen Bennett, 'Searching for sustainable and progressive policy solutions for illegal content in Europe' (*Mozilla*, 11 July 2018) <https://blog.mozilla.org/netpolicy/2018/07/11/sustainable-policy-solutions-for-illegal-content/> accessed 31 October 2019

[4] European Commission, 'Impact Assessment *accompanying* the document Proposal for a Regulation of the European Parliament and of the Council on preventing the dissemination of terrorist content online' (2018) <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52018SC0408&from=EN> accessed 31 October 2019

exercise, performed periodically, generates useful feedback, which can be, in turn used to improve the system.

Research overview

Background

Section 79 of the Information Technology Act ["the IT Act"], alongside a few other empowering provisions[5] have comprised the governing law of intermediary liability in India. Of this, section 79 is of significance because of two reasons. First, it is an exempting provision, granting safe harbour to the intermediaries for third party content that they host or transmit. Second, it empowers the Central Government to make rules prescribing guidelines for intermediaries to follow if they want to qualify for exemption from liability for third-party content that they host or transmit.[6]

Aside section 79, the other pertinent provision is section 69A and the allied IT (Procedure and Safeguards for Blocking for Access of Information by Public) Rules, 2009 ["the blocking rules"]. Under them, the central government has the power to issue blocking orders to intermediaries, under grounds mentioned in section 69A. While the grounds for sending takedown orders more or less overlap for section 69A and section 79, decency and morality does not form part of the former.

The importance of Section 79, was only emphasised in 2008, through the case of *Avnish Bajaj v State*[7]. In that case, the managing director (MD) of *bazeee.com* (now *ebay.com*), was arrested for a listing on his website, which offered the sale of the infamous DPS-MMS clip. Following this, the IT Act was amended in 2008. Section 79 was modified to incorporate more details, and aggregated with the Intermediary Guidelines of 2011, the India regime of intermediary liability became stricter.

As per the new framework, the intermediary would be exempted from individual liability for the content posted by its users, i.e., would be entitled to a 'safe harbour' protection. For claiming such protection, however, the intermediary was bound to take down any offending content within 36 hours of the existence of the content being made known to them.[8] This, of course, drew considerable amount of criticism due to the ramifications of the same on free speech online. Some intermediaries, for instance, were found to over-comply with the requirements of the law as a means of avoiding liability.[9]

---

[5] Section 67C, 69, 69A, Information Technology Act ["IT Act"] 2008
[6] Section 79(2)(c), IT Act
[7] *Avnish Bajaj v State*, (2005) 3 CompLJ 364 Del
[8] Rule 3(4), Information Technology (Intermediaries guidelines) Rules, 2011 ["IT Rules 2011"]
[9] Rishabh Dara, 'Intermediary Liability in India: Chilling Effects on Free Expression on the Internet', *Centre for Internet and Society* <https://cis-india.org/internet-governance/chilling-effects-on-free-expression-on-internet> accessed on 31 October 2019

Finally in 2015, the Supreme Court read down the section, in the landmark *Shreya Singhal* [10] decision, holding that the section must be construed with two caveats in mind. Firstly, the thirty six hours period mentioned would be applicable only when the intermediary becomes aware '*from a court order or on being notified by the appropriate government or its agency that unlawful acts relatable to Article 19(2) are going to be committed then fails to expeditiously remove or disable access to such material.*'[11] Secondly, the court order or the government notification must adhere to the grounds of 'reasonable restriction' laid down in Article 19(2). Any act which is deemed unlawful beyond the applicability of Article 19(2) cannot be a part of a legal takedown order under section 79. These two caveats, in the view of the Court, were enough to save the section from being deemed unconstitutional.

Starting from 2017-18, the government has been calling for changes in the intermediary liability regime in India, citing problems of misinformation and hate speech.[12] Deliberate misinformation campaigns, through popular social media apps like Facebook and communication apps like WhatsApp, spilled into both online and offline harms.[13]. This has spurred the Indian government to take concrete steps to strengthen the regime governing the liability of intermediaries.[14]

Accordingly, in December 2018, MeitY introduced the 'The Information Technology (Intermediary Guidelines (Amendment) Rules) ["the rules"], which among other things, shortened the previous 36 hours period of compliance to 24 hours. Draft Rule 3(8), the relevant rule, currently read as:

> "*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 [...] 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. [...]*"

Objective

In the process of submitting CIS's response to the rules, we found a gap in research which made a case for supporting a mandatory 24-hour timeline for intermediaries to comply with content removal orders. Moreover, no research exists that peruses a) parallel

---

[10] *Shreya Singhal v Union of India*, AIR 2015 SC 1523
[11] Id
[12] Ministry of Electronics and Information Technology, 'Comments invited on Draft of Intermediary Guidelines 2018' <https://meity.gov.in/comments-invited-draft-intermediary-rules> accessed on 4 November 2019 [*hereinafter* "MeitY Comments"]
[13] BBC News, 'How WhatsApp helped turn an Indian village into a lynch mob' (19 July 2018) <https://www.bbc.com/news/world-asia-india-44856910> accessed on 4 November 2019; Timothy Mclaughlin, 'How WhatsApp Fuels Fake News and Violence in India' (*Wired*, 12 December 2018) <https://www.wired.com/story/how-whatsapp-fuels-fake-news-and-violence-in-india/> accessed on 4 Novermber 2019
[14] MeitY Comments (n 12)

regulations around the world with differing time-frames and b) the literature surrounding these regulations, in a bid to contextualize the current time-frame reflected in the rules.

Thus, the objective of this study is to understand the appropriate timeframe for an intermediary to respond to government requests for content removal ["turnaround time"], and how the same can be framed in law and policy.

Methodology and scope

To inform this investigation, we rely on (i) a survey of regulations that define a turnaround time; and (ii) the response from Indian stakeholders to the turnaround time defined in the draft intermediary guidelines released by the MeitY in 2018.

Survey of regulations that define a turnaround time

We first survey prominent regulations and laws around the world which have proposed different timeframes for intermediaries operating in their countries. A nuanced understanding of an appropriate turnaround time requires analysing the law on certain concomitant metrics. These include seeing whether the law introduces a mandated transparency reporting system, classifies intermediaries and differentiates on the scale of liability of each of these classes, and institutes a notice system.

Legislation on these factors, invariably affects the time period taken by the intermediary in responding to content removal requests. For instance, a removal mechanism that involves notice and appeal would take more time to process as opposed to one without such an option. Hence, the study will attempt to develop these metrics and highlight their importance in determining the appropriateness of a turnaround time.

Accordingly, in the first part, we study five key pieces of regulations and laws: the NetzDG, the European Union Regulation on Terrorist Content, the European Union Code on Countering Illegal Hate Speech Online, the Australian law criminalizing sharing of abhorrent violent criminal content online and the Harmful Digital Communications Act (HDCA).

Perspectives of Indian stakeholders

The second part of the report looks at the response of the Indian stakeholders with respect to the updated timeframe by way of both interviews and collection of information from publicly available documents. Thereafter, we map out the responses of the Indian stakeholders to the proposed timeframe in the draft rules. The stakeholders whose views are represented in this report can be divided into three categories: (a) Civil society, (b) Industry bodies; (c) Intermediaries operating in India

The public facing documents used were the various submissions on the draft amendments to the MeitY, by the above stakeholders. We found a total of 19 submissions relevant to the study that are available on the MeitY website.[15]

Apart from these submissions, we requested interviews with representatives from several intermediaries, content moderation organisations and important industry bodies operating in India; however most of them declined to comment on the issue.

Number of entities contacted	Number of entities who declined to comment/did not respond
16	13

Interviews were conducted with the three intermediaries who reverted positively to our request for comments. These are discussed in the second part of the report, though as per organisational norms, their responses have been anonymised. An illustrative list of the interview questions are provided at the end of the report, in Annexure 1.

During the survey, we also noted that the question of timeframe of takedown, leads us to assess the law on several other metrics. This include:

- Whether the law introduces any gradation in how it views intermediaries? In other words, does it introduce any way of classifying intermediaries?
- Whether the law has a notice and appeals mechanism?
- Whether the law provide carve-outs for companies who have not violated the law before?
- Whether the law provides any transparency reporting obligations, or any other way of monitoring the impact and implementation of the law?
- Whether the law proposes separate turnaround timeframes depending on the nature of the content flagged for removal?

Finally, we utilize the highlights and learnings from the laws surveyed, and utilize them to introduce some nuances into the current Indian framework.

Limitations

We recognize that our selection of relevant regulations does not reflect enough regionalistic diversity. More specifically, our study does not take into account any Asian, African or Latin American laws into account.

[15] Comments on draft of "The Information Technology [Intermediary Guidelines (Amendment) Rules] 2018 <https://meity.gov.in/writereaddata/files/public_comments_draft_intermediary_guidelines_rules_2018.pdf> accessed on 18 Novermber 2019 [*hereinafter* "Comments on IL Rules"]

Survey of relevant regulations

In this section, we give a brief overview of the various laws and discuss their scope, legislative history wherever relevant, and the core obligations the law places on the intermediaries. As part of this overview, we also identify the timeframe given to the intermediaries to respond to content removal requests. Additionally, the critical opinions around the laws are analysed within this part since some of the concerns are relevant to the Indian context.

NetzDG

Overview

The Network Enforcement Act [*Netzwerkdurchsetzungsgesetz* in German], or NetzDG in short, has its roots in, among other things, in a series of agitations and discriminatory online behaviour against minorities in late 2015-16.[16] In its final form, the law makes social media platform operators with at least two million users within Germany, responsible for removing certain categories of unlawful content (with reference to the twenty-two existing statutes in the German criminal code that include "incitement to hatred", "dissemination of depictions of violence" and so on.[17]) from their platforms. *Social media platforms are defined as "telemedia service providers" that operate with profit-making purposes and allow users to share content with other users, or make such content available to the public.*[18] Platforms which provide user-to-user communication, like private messaging or emails, and platforms providing editorialised content, are excluded from the scope of the law.[19]

The following are the two core obligations under the law;

1. **Setting up an effective and transparent complaints management infrastructure**:[20] This system is supposed to be integrated into the intermediary's normal flagging mechanism, so both users and government can use this to flag content violating the NetzDG. As part of this obligation, social media platforms must ensure they delete or block illegal content within a specified timeframe. Once a social media platform within the scope of the law receives a complaint through its complaints mechanism, it must investigate whether the content is 'illegal'. If the content is 'manifestly unlawful', then the platform has an obligation to remove the content within 24 hours. For every other case, the timeframe of compliance is 7 days. Persistent violation to adhere to this core obligation may invite a fine of up to 50

[16] Amelie Heldt, 'Reading between the lines and the numbers: an analysis of the first NetzDG reports' (2019) 8(2) Internet Policy Review <https://policyreview.info/articles/analysis/reading-between-lines-and-numbers-analysis-first-netzdg-reports> accessed on 4 November 2019
[17] Article 1, Section 1(3), NetzDG
[18] Article 1, Section 1(1), Network Enforcement Act 2017 ["NetzDG"]
[19] Id
[20] Article 1, Section 3, NetzDG

million euros.

2. **Compiling bi-annual reports on the complaints management activity:**[21] This includes mandatory transparency reporting, in case the social media platform receives more than 100 complaints per year. The law also mentions certain details which must be included in the transparency report.

Critical opinions

While the law was being developed, several critics pointed out that it may have adverse implications for free speech online. The critics included tech industry players, activists and academicians, though such criticism was offset by a large majority of German voters in favour of the law.[22]

Note that concerns of NetzDG violating German constitutional law are are out of the scope of this study. Broadly, the other concerns surrounding NetzDG relate to:

- **Over-removal**: Over-removal or over-blocking occurs when content is blocked or deleted without any reason, because there exists incentive to immediately delete content, rather than perform checks to ensure that the fundamental requirements of takedown are present.[23] Analysing whether or not a particular piece of content falls under the purview of 'illegal content' would require considerable knowledge of German law and language,[24] as well as complex case-by-case investigation. None of this, according to the critiques of the NetzDG, would be possible in such a short timeframe. Coupled with heavy fines in instances of persisting failures, this would present a strong incentive for companies to remove content without paying heed to the merits of the takedown complaints.[25] This risks the censoring of legitimate speech online as well, thus resulting in false positives.

- **Privatization of enforcement**: The law turns private companies with no democratic legitimacy into judges of the legality of the content posted by users.[26]

- **Absence of judicial oversight**: There is no requirement for the complainant to obtain a court order before requesting the platform to takedown the content.[27]

- **No commitment for appeals**: The law also does not institute any clear appeals mechanism for any user who feels content has been unjustly taken down.[28]

---

[21] Article 1, Section 2, NetzDG
[22] Heidi Tworek, Paddy Leerssen, 'An Analysis of Germany's NetzDG Law' (2019) Transatlantic Working Group <https://www.ivir.nl/publicaties/download/NetzDG_Tworek_Leerssen_April_2019.pdf> accessed on 4 November 2019
[23] Amelie Heldt (n 15)
[24] Amelie Heldt (n 15) Heidi Tworek (n 21)
[25] Id
[26] Amelie Heldt (n 15)
[27] Id
[28] Heidi Tworek (n 21)

- **Ambiguous terms**: Additionally, the term 'manifestly' lacks any clear legal specifications, thus rendering a wide range of discretion on the platform to takedown content as it sees fit. This, in turn, results in a chilling effect on free speech.[29]

NetzDG enforcement: Numbers and realities

We will briefly summarize the patterns indicated by the NetzDG transparency reports published by Google (and Youtube), Facebook, Twitter and Change.org,regarding the enforcement realities of the law. While the widespread concerns of over-censorship continued to prevail, the first round of NetzDG transparency reports seemed to indicate a contrasting reality.

The first round of transparency reporting focussed on the time period between January-June 2018. The numbers and percentage of removal that were presented in these reports can be summarized as:

Table 1: Numbers and percentages of removal under NetzDG[30]

Name of the platform	Total items reported	Reports resulted in action	Removal rate within 24 hours
Facebook	1704	362 (21.2 %)	76.4 %
Youtube	241,827	58,297 (27.1 %)	93.0 %
Google+	2769	1277 (46.1 %)	93.8 %
Twitter	264,818	28,645 (10.8 %)	97.9 %
Change.org	1257	332 (26.4 %)	92.7 %

It is worth observing that despite Facebook's prospectively large user-base, it received far fewer removal requests than Twitter and Youtube. The TransAtlantic Working Group, which made a study of the law, posits that this is because Facebook's complaints mechanism was relatively harder to access than those of Twitter and Youtube, who integrated NetzDG complaints in their own flagging interfaces.[31]

Additionally, the Working Group noted that majority of the takedowns resulting from the NetzDG complaints mechanism, for Google, Facebook and Twitter, happened under their Community Guidelines, rather than the referenced German speech laws.[32] All these three intermediaries seemed to prioritize checks with their Community Guidelines over the German speech laws.

[29] Amelie Heldt (n 15)
[30] Heidi Tworek (n 21); Kirsten Gollatz et. al, 'Removals of online hate speech in numbers' (*Digital Society Blog*, 9 August 2018) <https://www.hiig.de/en/removals-of-online-hate-speech-numbers/> access on 4 November 2019
[31] Heidi Tworek (n 21)
[32] Id

The numbers represented in these Transparency Reports have ushered in another fresh round of debates regarding the efficacy of the law. It is difficult to assess how far NetzDG actually has been able to reach its objectives, and how far of its effects had an impact on free speech online, since the transparency reporting obligations under the law is not standardized, and as pointed out, neither are the complaints mechanism system for each intermediary.[33]

That being said, it is interesting to note that all the major intermediary incumbents have reportedly been able to take down a substantial portion of illegal content within the 24 hour timeframe mandated by the law. While this is a telling point, we must also note here that the mere ability to takedown content does not equal the accuracy of the action. In other words, the concerns regarding over-blocking or censorship are not assuaged by the numbers presented by the intermediaries, since the reports, more often than not, omit important information on context and intent.[34]

Regulation of the European Parliament on Preventing the Dissemination of Terrorist Content Online

In 2018, the European Commission (EC) published a set of operational measures as a follow-up to the EC's 2017 Communication on tackling illegal content online. Among its recommendations on more generalized procedures, the measures also make a reference to terrorist content online. In relation to that, the measures recommend an one-hour takedown timeline as a *general rule*.[35] Following these measures, the EC published the complete Regulation on Preventing the Dissemination of Terrorist Content Online ("the Regulation"), which were passed in 2019.

Overview

The scope of the Regulation includes hosting service providers, who offer their services within the Union, irrespective of their size or place of establishment. Additionally, illegal terrorist content is defined as "*information which is used to incite and glorify the commission of terrorist offences, encouraging the contribution to and providing instructions for committing terrorist offences as well as promoting participation in terrorist groups.*"[36]

The Regulation envisages each Member State setting up a competent authority in their State, who would be empowered to flag problematic content online. Once they do, this authority will send out a removal order to the concerned hosting service provider, who

[33] Kirsten Gollatz (n 26)
[34] Kirsten Gollatz (n 26)
[35] European Commission Press Release, 'A Europe that protects: Commission reinforces EU response to illegal content online' <https://europa.eu/rapid/press-release_IP-18-1169_en.htm> accessed on 4 November 2019
[36] European Commission, Regulation Of The European Parliament And Of The Council on preventing the dissemination of terrorist content online [2018] COM(2018) 640 final

must delete or disable access to that content for the users residing in EU, within a one-hour timeframe.[37]

On subsequent revision, the European Parliament decided that for companies which have never received a removal order before, the competent authority should contact it first, provide information on procedures and deadlines, at least twelve hours before issuing the first content removal order.[38] The Parliament also modified the authority empowered to send out content removal order from 'competent authority' to judicial or a functionally independent administrative authority.[39]

Lastly, the proposal discusses setting up of an appeals mechanism, but does not lay down minimum standards for such an appeals system. This includes absence of a provision that enables service providers to ask for clarification in case the removal order contains missing information or technical deficiencies.[40]

Critical opinions

Since last September, when the Regulation was first made publicly available, there has been considerable amount of criticism surrounding the takedown timeframe given to the hosting service providers. Like with the NetzDG, concerns were voiced regarding the possibility of the over-removal,[41] and unduly burdening the smaller companies with onerous obligation[42], thus instituting higher barriers to market entry. Critics voicing the latter opinion also pointed out that such an obligation would strengthen the larger companies at the cost of the smaller ones.[43]

Additionally, critics also pointed out that the timeframe, coupled with disruptive sanctions, would also lead to some companies automating the takedown procedure,[44] which would complicate the process and give rise to additional free speech concerns.

---

[37] Id
[38] European Parliament News, 'Terrorist content online should be removed within one hour, says EP' (17 April 2019) <https://www.europarl.europa.eu/news/en/press-room/20190410IPR37571/terrorist-content-online-should-be-removed-within-one-hour-says-ep> accessed on 4 Novermber 2019
[39] Alexander Pirang, ' The EU Terrorist Content Regulation Is Unfinished Business for the European Parliament' (*Global Public Policy Institute*, 23 May 2019) <https://www.gppi.net/2019/05/23/unfinished-business-for-the-european-parliament-the-eu-terrorist-content-regulation> accessed on 4 Novermber 2019
[40] Joris van Hoboken et. al, 'The Proposed EU Terrorism Content Regulation: Analysis and Recommendations with Respect to Freedom of Expression Implications' (2019) Transatlantic Working Group <https://www.ivir.nl/publicaties/download/TERREG_FoE-ANALYSIS.pdf> accessed on 4 November 2019
[41] Faiza Patel, 'EU 'Terrorist Content' Proposal Sets Dire Example for Free Speech Online' (*Just Security*, 5 March 2019)< https://www.justsecurity.org/62857/eu-terrorist-content-proposal-sets-dire-free-speech-online/ ' https://blog.mozilla.org/netpolicy/2019/03/07/one-hour-takedown-deadlines-the-wrong-answer-to-europes-content-regulation-question/> accessed 4 November 2019
[42] Draft Opinion of the Committee on Culture and Education for the Committee on Civil Liberties, Justice and Home Affairs 2018/0331(COD) <http://www.europarl.europa.eu/doceo/document/CULT-PA-632087_EN.pdf?redirect> accessed on 4 November 2019
[43] Alexander Pirang (n 35)
[44] European Union Agency for Fundamental Rights, 'Proposal for a Regulation on preventing the dissemination of terrorist content online and its fundamental rights implications' <https://fra.europa.eu/sites/default/files/fra_uploads/fra-2019-opinion-online-terrorism-regulation-02-2019_en.pdf> accessed on 4 Novermber 2019

Lastly, in the Draft Opinions filed by the Committee on the Internal Market and Consumer Protection (IMCO), it was recommended that instead of keeping a stringent one-hour timeframe, the language of the Regulation should be modified to '. . . *expeditiously, taking into account the size of and resources available to the hosting service provider*. . .'.[45] Along with this, IMCO also opined that single instances of failure to adhere to this timeline should not be subject to sanction, unless such failure is part of a persisting absence of adherence to the law. Principles of proportionality must also be invoked while deciding the issue of sanctions.[46]

Criminal Code Amendment (Sharing of Abhorrent Violent Material) Bill 2019

In the wake of the Christchurch shootings, Australia passed the Criminal Code Amendment (Sharing of Abhorrent Violent Material) Bill, 2019 ["the Bill"], which puts a heavy burden on social media companies if they do not remove 'violent material' from their platforms. On instances of failure, the government has the power to fine the social media companies, or even jail individuals responsible for running these companies.

Overview

The Bill applies to providers of both content service and hosting service,[47] which would broadly include both social media platforms and ISPs. Marked as a criminal law, the Bill considers a person to be committing an offence, if the person providing the content service or the hosting service, does not ensure 'expeditious' removal of 'abhorrent violent material' from their platforms. 'Abhorrent violent material' applies to a large spectrum of audio, visual and audio-visual material that depicts the acts of terrorist attacks, murders, rapes or kidnappings.[48]

Critical opinions

One of the major criticisms of this law has been that it is reactionary, and is passed without any consultation with the relevant stakeholders[49]. The 'expeditious' standard of removal timeframe has also been a cause of concern for many. What would constitute an expeditious removal of content would undoubtedly be a subjective issue, and would differ from person to person. The Explanatory Memorandum to the law states:

---

[45] Draft Opinion of the Committee on the Internal Market and Consumer Protection for the Committee on Civil Liberties, Justice and Home Affairs 2018/0331(COD) <http://www.europarl.europa.eu/doceo/document/IMCO-PA-632028_EN.pdf?redirect> accessed on 4 November 2019

[46] Id

[47] Section 474.34, Criminal Code Amendment (Sharing of Abhorrent Violent Material) Bill 2019

[48] Section 474.31, Criminal Code Amendment (Sharing of Abhorrent Violent Material) Bill 2019

[49] 'Australia can now jail social media executives over streamed violence' (*CBS News*, 4 April 2019) <https://www.cbsnews.com/news/australia-social-media-law-violent-video-streaming-illegal-facebook-new-zealand/> accessed on 4 November 2019

> "[...] *the type and volume of the abhorrent violent material, or the capabilities of and resourcing available to the provider may be relevant factors.*"[50]

This does not however, provide any clarity regarding the exact impact of these factors on the timeframe. Additionally, in the second reading speech, which courts can utilize for interpretation of the law[51], the Attorney-General condemned the fact that the video of the Christchurch shooting was broadcast without any interference for seventeen minutes, and that it was available for almost an hour and ten minutes before the first attempts were made to take it down.[52] This speech indicated that an expeditious timeframe would be calculated in terms of hours and minutes, rather than days.[53]

Lastly, determining whether an online company has acted in an expeditious manner in removal of concerned content would therefore likely be subject to judicial scrutiny, invoking technical questions, and would lead to extended court battles.[54]

The EU Code of conduct on countering illegal hate speech online

In May 2016, Facebook, Microsoft, Twitter and Youtube ("the IT Companies"), along with the EC, agreed to bind themselves with the Code of conduct on countering illegal hate speech online ("the Code"). Adherence to the Code is entirely voluntary. Over the course of its existence, Instagram, Google+, Snapchat, Dailymotion and Jeuxvideo.com have joined.[55]

Overview

The Code's definition of illegal hate speech is borrowed from Framework Decision 2008/913/JHA of 28 November 2008 on combating certain forms and expressions of racism and xenophobia by means of criminal law and national laws transposing it as "*all conduct publicly inciting to violence or hatred directed against a group of persons or a member of such a group defined by reference to race, colour, religion, descent or national or ethnic origin.*"[56]

Under the Code, the IT Companies undertake the following obligations:

---

[50] Explanatory Memorandum, Criminal Code Amendment (Sharing of Abhorrent Violent Material) Act
[51] Section 15AB(2)(f), Acts Interpretation Act 1901
[52] Evelyn Douek, 'Australia's "Abhorrent Violent Material" Law: Shouting "Nerd Harder" And Drowning Out Speech', [forthcoming 2020] *Australian Law Journal* <: https://ssrn.com/abstract=3443220> accessed 4 November 2019
[53] Id
[54] Damien Cave, 'Australia Passes Law to Punish Social Media Companies for Violent Posts' (*The New York Times*, 3 April 2019) <https://www.nytimes.com/2019/04/03/world/australia/social-media-law.html> accessed 4 November 2019
[55] 'The EU Code of conduct on countering illegal hate speech online' <https://ec.europa.eu/info/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/countering-illegal-hate-speech-online_en> accessed 4 Novermber 2019
[56] Id

- To have in place clear and effective processes to review notifications regarding illegal hate speech on their services so they can remove or disable access to such content.
- Upon receiving a valid removal notification, the IT Company will review the requests against its community guidelines, and where necessary, national laws transposing the Framework Decision 2008/913/JHA. They are also to have in place dedicated teams reviewing requests.
- Most importantly, the IT Companies must review 'the majority' of valid notifications within less than 24 hours, and remove/disable access to the content within that timeframe, if necessary.

The Code also envisages participation of civil society organisations (CSOs) to prevent the rise of hate speech online. Accordingly, through the Code, the IT Companies promise to promote the institution of a system of notice and flagging of hateful content at scale by experts, and specially by partnerships with CSOs.

The EC's sub-group on countering hate-speech also decided to agree upon a common methodology to assess how the IT Companies fared under the Code. Accordingly, there have been four rounds of monitoring exercises on the impact of the Code, the numbers of which are presented in the next section.

Results of the monitoring exercises

The full versions of the monitoring exercises are available online.[57] For the purposes of our report, we would be presenting the data related to the average time taken by the IT Companies to assess the removal notifications.

Table 2: Results of monitoring exercises for the period 2016-19

Year	Less than 24 hours	Less than 48 hours	Less than a week	More than a week
2016	40%	43%	N/A	N/A
2017	51.4%	20.7%	14.7%	13.2%
2018	81.7%	10%	4.8%	3.5%
2019	89%	6.5%	3.9%	0.7%

As the results of the monitoring exercises show, the overall response time for the IT Companies to removal notification has lessened. Where in 2016, the IT Companies could respond to only 40% of the notifications within less than 24 hours, in 2019, the number has more than doubled. The latter data must also be considered in the context of the additional IT Companies who had joined the initiative in 2018.

[57] Id

Undoubtedly, the Code enjoys certain advantages over the other relevant laws and regulations discussed above. Broadly, they can be summarized as:

- Adherence to the Code is voluntary, and accordingly, the scope of the obligations till now have only included the market incumbents, as opposed to top-down, blanket regulation, which, more often than not, burden the smaller companies.

- Adoption of a common methodology for assessing the impact of the Code also ensured that some of the transparency reporting fallacies noted in the NetzDG methodology can be clarified here.

However, like the laws and regulations before, the bare numbers do not represent the whole picture, and accordingly, the concerns of over-blocking and privatization of enforcement remain. Additionally, the IT Companies do not undertake to provide users whose content has been taken down with a notice to that effect, nor do they undertake to provide such users with an appeal mechanism.

Harmful Digital Communications Act (HDCA)

In 2012, the Law Commission of New Zealand provided a Ministerial Briefing, Harmful Digital Communications: The Adequacy of the Current Sanctions and Remedies, which also contained a proposed Bill.[58] The initial regulatory scope of the law had been to address issues of cyber-bullying. Subsequently, this scope has broadened. In its final form, which was passed in 2015, the law targeted 'harm', interpreted to mean serious emotional distress[59], giving way to a large variety of possible conduct that could be brought under the law.

Overview

Unlike the laws reviewed previously, the HDCA envisions a notice-and-notice system of content takedown. According to section 24, an intermediary will not be liable for harmful content posted by the user, *if* it receives a notice of complaint about the content, serves the originator of the content with the notice within 48 hours, and notifies the originator that should they want, they would have 48 hours to issue a counter-notice. In case the intermediary fails to locate the originator of the content after taking reasonable steps, then the onus is on the intermediary to disable the content within 48 hours of receiving the notice.[60] The concerned section also lays down the constituents of a valid notice and counter-notice.[61] This includes the telephone number, e-mail ID and a physical address of the concerned user.

---

[58] The Law Commission, 'Regulatory Gaps and the New Media' (2010) <https://www.lawcom.govt.nz/our-projects/regulatory-gaps-and-new-media> accessed on 4 November 2019
[59] Section 4, Harmful Digital Communications Act, 2015
[60] Section 24, Harmful Digital Communications Act, 2015
[61] Id

Section 25 of the law further states that the safe harbour protection does not apply to the intermediary if it fails to provide the user with an easily accessible mechanism to issue notices[62] as per section 24.

Critical opinions

The HDCA has been praised for the considerably long period of consultation that preceded the final law.[63] The Law Commission, the legislative body responsible for the law, debated numerous considerations surrounding the law, including privacy and the issue of harmful digital communications.[64]

On the other hand, critiques of the law point out that the notice-and-notice regime closely emulates the Digital Millennium Copyright Act (DMCA) model existing in the US, except that it encompasses far many more categories of content than mere copyright breaches.[65] As the Electronic Frontier Foundation (EFF) has been mapping over the years, the DMCA model has resulted in several unintended consequences, including censorship of perfectly legitimate digital content.[66]

Additionally, unlike the DMCA, the HDCA does not institute any penalties for misrepresentation. This means that should the situation arise, a user can be made target of coordinated, malicious issuance of multiple takedown notices. Such attacks can make the user vulnerable, and accordingly make them miss the 48-hour deadline for sending the counter-notice.[67] Critiques have also pointed out that the mandate of including personal information in the notice and counter-notice makes users vulnerable to doxxing [68].

Stakeholder mapping and empirical data

Stakeholder mapping from public facing documents

As we have indicated in the introduction, the MeitY is looking significantly modify the Indian intermediary liability regime as indicated by the release of the draft rules in December 2018.

---

[62] Section 25, Harmful Digital Communications Act
[63] Daithi Mac Sithigh, 'Tackling the trolls: how New Zealand raised the bar with its new laws' (*The Conversation*, 16 July 2015) <https://theconversation.com/tackling-the-trolls-how-new-zealand-raised-the-bar-with-its-new-laws-44691> accessed on 4 November 2019
[64] Id
[65] Danny O'Brien, 'New Zealand's Harmful Digital Communications Act: Harmful to Everyone Except Online Harassers' (*Electronic Frontier Foundation*, 8 July 2015) <https://www.eff.org/deeplinks/2015/07/nz-digital-communications-act-considered-very-harmful> accessed on 4 November 2019
[66] Id
[67] Id
[68] Id

The data used in the mapping exercise indicated in the table below has been collected through information extracted from stakeholder submissions to the rules.[69]

Table 3: Consolidated stakeholder submissions to the rules

Concerns	Relevant organisations
Lack of due process and procedural safeguards: The shortened timeframe does not allow the intermediaries to scrutinize the takedown order to ensure that it is adhering to the technical and legal requirements. This in turn, does not institute due process, thus running counter to the judgment of *Shreya Singhal*. If the intermediary is unable to scrutinize the request properly before adhering to it, this may also result in third-party actions against user content. Additionally, the concerned rule also does not create any checks which would ensure that the government exercises its powers in only lawful cases.	- Asia Internet Coalition - ITU APT Foundation - Amnesty International - CCAOI - Broadband India Forum - Internet Society (Delhi Chapter) - NASSCOM - Cellular Operators Association of India
Issues with implementation: The significantly short timeframe also raise several challenges with regards to implementation. Specifically, for smaller companies with fewer employees, such a timeframe can both be burdensome, from both a financial and capability point of view. This in turn, may result in over-censorship.	- Asia Internet Coalition - Computer and Communications Industry Association - Internet Freedom Foundation - Global Network Initiative - Centre for Communication Governance (CCG)
Lack of classification of content: The scope of the rules cast a large net of content that is 'unlawful' and therefore, prohibited. This includes content that is 'defamatory' and content that relates to 'security of the State'. In view of such a varied gamut of content, the rules ought to differentiate the timeframe between content removal requests based on priority and urgency amongst requests.	- Change.org - BananaIP - Asia Cloud Computing Association - CCG - Sharechat

[69] Comments on IL RUles (n 15)

For content that relates to public wrongs, certain median timelines may be imposed, but for content that constitute a private wrong like defamation, a uniform, strict timeframe would be unreasonable.	
Necessity of having 'stop the clock' provisions in the rules: The draft rule does not envisage situations where additional time, over and above the allotted time, is required for the concerned entity to comply with its obligations under the law. Such provisions are specifically required alongside the 24-hour timeframe, for situations where the intermediary requires more information or clarity about the request in question.	- Information Technology Industry Council - US - India Business Council - AMCHAM India

Data collected through interviews

As indicated in the methodology section, we also attempted to contact a diverse range of entities to gather their inputs for the subject. Three out of sixteen intermediaries we approached responded, and their interviews are discussed below. While we observed some diversity of views in these interviews, we think that no conclusive evidence for policy recommendations can be drawn due to the small number entities willing to be interviewed. The main points discussed are nevertheless presented here to provide a full picture of the research conducted.

Intermediary #1

Intermediary #1 is incorporated in India and works around content in Indic languages. They recommended that for content that is legally charged to be waging war against India, should be removed within 3 hours from the time of a valid request. According to them, the graded timeframe would look like:

Table 4: Intermediary #1's suggestion of a graded timeframe

Timeframe of response	Nature of the content
3 hours	Relating to terrorism, or waging war against India
72 hours	Relates to risk to public order at large, such as creating enmity between groups, and rioting
7 days	All other requests

Intermediary #2

Intermediary #2 was also incorporated in India, and curated content exclusively in Hindi. Their user-base constitutes largely of students and retired professionals over the age of 65 years, belonging to tier-2 and tier-3 cities. Due to a considerably small user-base of

approximately 200,000 users, they said that they receive fewer removal requests in a month, and did not find it difficult to remove them in accordance with the demands of the government. Accordingly, they did not think that the updated timeframe would affect their business model in any way.

Intermediary #3

Intermediary #3 is a platform with a large Indian userbase. A majority of the content removal requests received by them were issued under Section 69A of the Information Technology Act, 2000 and the corresponding rules ["the blocking rules"]. Thus, a substantial part of our discussion was focused on the procedure established by the blocking rules.

In their opinion, since section 79 of the Information Technology Act, 2000 is an exemption provision as noted in *Shreya Singhal v. Union of India* therefore it should not be used to empower authorities to issue content removal directions. Finalizing the draft rules in its current form would further render the procedure established by the blocking rules redundant and will result in side-stepping the checks and balances available in the blocking rules.

They also emphasized on the necessity of a graded approach while affixing a turnaround time for specific types of content. There were several factors which come into play while determining how much time would it take for the intermediary to disable access to the content. Among other things, this includes:

- **Volume of the referenced information**: The content directed for blocking by the Government can vary on several counts. For example, while one request may contain 2 URLs to be examined, another may contain 2000 URLs. Blocking the latter would take more time than the former.
- **Type of content being referred**: The content in question may be either:
    - An entire account: Where the referenced link contains multiple speech dimensions, like an entire account or a handle or a page.
    - A specific piece of content: When the referenced link contained only one piece of information, or only one speech dimension.
- **Language**: Given the diversity of languages in which content is being uploaded, translation is another challenge to be factored in.

Takeaways and recommendations

In this part, we utilize the discourse from the previous sections to formulate some key recommendations that need to be taken into account while affixing a timeframe of compliance. These are further developments from both the metrics discussed in the last section, as well as the views of the stakeholders reflected in the mapping exercise and in the review of relevant legislation.

We recognize that implementing all the following recommendations may not be completely within the scope of the scheme of section 79. As we have pointed out, for instance, the requirement under the rules regarding the intermediary to establish a company under the Companies Act was *ultra vires* the section[70]. In light of that, the following recommendations only highlight the requirement of having a nuanced and granular legislation on turnaround timeframes.

The need to harmonize the existing legal system on content takedown

Our conversation with intermediary #3 led to an important revelation: despite the existence of the rules under section 79 facilitating content removal, government officials often choose to take the blocking procedure under section 69A for achieving the same goal.

This points to a larger issue of a confusing legal framework for content removal. While on the one hand the blocking rules give the intermediary at least 48 hours to reply and/or seek clarifications on a removal request, the new rules under section 79 will only give a 24 hours turnaround time. The confusion is further exacerbated by the fact that the grounds for removal under section 69A and section 79 often overlap, save a few provisions[71]. Additionally, the scope of section 69A is broader than section 79, inasmuch as the former targets anyone with a computer resource, while the latter applies to intermediaries seeking safe harbour.

In such circumstances, there is a need for harmonizing the two legal framework. We believe there are multiple ways of achieving the same:

- Instituting a single content takedown regime, legislated by a separate section.
- Retain both provisions, but ensure that the procedure and the safeguards attached to the takedown is uniform across the sections. This would include, among other things, introducing a shared oversight mechanism in the procedure followed under the rules, where intermediaries would be given a fixed minimum time to contest or reply to removal orders.

Mandated transparency reporting

Efficacy of a particular turnaround time

Transparency reporting, apart from ensuring accountability, is also a useful tool for understanding the impact of the law, specifically with relation to time period of response.

---

[70] Gurshabad Grover, Elonnai Hickok et. al, 'Response to the Draft of The Information Technology [Intermediary Guidelines (Amendment) Rules] 2018' (*Centre for Internet and Society*, 31 January 2019) <https://cis-india.org/internet-governance/resources/Intermediary%20Liability%20Rules%202018.pdf> accessed on 4 November 2019

[71] More specifically, the ground of 'decency and morality' does not form part of section 69A, while it forms part of section 79.

Despite the free speech concerns surrounding NetzDG, the transparency reporting obligations mandated in the law have been met with positive response. Article 19, while asking for the repeal of the law, has asked that the transparency reporting provisions be maintained[72]. Similarly, the EU Code has in-built processes for monitoring the effects of the regulation on a yearly basis. Even the Australian law envisages a process by which a review of the law would be conducted post two-year of the commencement of the law.[73]

However, neither the rules nor any provisions in the IT Act envisage any review procedure to map the effects of the law on the intermediaries or on the speech rights of the users online. We recommend that there be a legal mandate on the intermediaries to come up with regular reports detailing the way the obligations have affected moderation of content online. To ensure uniformity in these reports (and thereby avoid the issues with the NetzDG transparency reporting inadequacy), we recommend that the intermediaries operating in India, in consonance with the lawmaking authorities, develop a standard reporting framework. In one of our reports[74], where we examine the existing transparency reports of the incumbent non-Indian intermediaries, we proposed some of the following elements to be factored in while reporting government requests for content removal:

1. **Numbers:** The transparency report should contain both the number of requests received and the number of requests granted.
2. **Sources:** We believe that the transparency report must classify the government requests to those from the executive, the judiciary and from third party.
3. **Items:** The intermediary should show the number of items taken down, in addition to the number of requests acted upon since a single request may specify numerous items to be taken down. An 'item' here refers to one particular piece of user content, be it a blog post or a video.
4. **Platforms:** For further accountability, we believe an intermediary owning or operating multiple platforms, should publish platform-wise breakdown of content taken down and data produced. Alternatively, they should publish separate transparency reports for each platform.
5. **Geographical Scope:** Intermediaries should aim to remove the content only from the jurisdiction where it is deemed to violate the law. In the current study, content removed in India should be made available elsewhere as far as the law permits.

In absence of a consensus on a standard reporting practice from platforms, we recommend adherence to the Santa Clara Principles,[75] which provide guidance to online companies to ensure transparency and accountability in their reporting, as a potential framework

---

[72] Article 19, 'Germany: The Act to Improve Enforcement of the Law in Social Networks' (2017) <https://www.article19.org/wp-content/uploads/2017/09/170901-Legal-Analysis-German-NetzDG-Act.pdf> accessed on 19 November 2019

[73] Section 474.45, Criminal Code Amendment (Sharing of Abhorrent Violent Material) Bill 2019

[74] Torsha Sarkar, Suhan S and Gurshabad Grover, 'Through the looking glass: Analysing transparency reports' (*Centre for Internet and Society*, 30 October 2019) <https://cis-india.org/internet-governance/blog/torsha-sarkar-suhan-s-and-gurshabad-grover-october-30-2019-through-the-looking-glass> accessed on 4 November 2019

[75] The Santa Clara Principles on Transparency and Accountability in Content Moderation <https://santaclaraprinciples.org/> accessed on 4 November 2019

We also believe such mandate would provide the public with a better understanding of the impact of varied timeframes of response, facilitating further research into this area.

Efficacy of the monitoring mechanism

Due to the experimental nature of the provision, including a review provision in the law would ensure the efficacy of the exercise can also be periodically assessed. If the regulations surveyed in the preceding section are any indication, the issue of an appropriate turnaround time is currently in a regulatory flux, with no correct answer. In such a scenario, periodic assessments compel policymakers and stakeholders to discuss effectiveness of solutions, and the nature of the problems faced, leading to evidence-based policymaking.[76]

Institution of an user notification and appeals system

The mere fact that the larger intermediaries are able to respond to removal notifications within the specified timeframe does not imply that their actions do not have ramifications on free speech. Ability to takedown content does not imply accuracy of their actions. Additional obligations of informing users when their content has been taken down, therefore, institutes due process in the procedure. In the context of legal takedown, such notice mechanisms also empower users to draw attention to government censorship and targeting.[77]

While the proposed rules give the intermediary prerogative to suspend user accounts in instances of violation of terms of service, as well as obligates them to respond to government requests for content takedown[78], they do not mandate the intermediary to institute a notification and appeals system for users whose content has been taken down, or whose account has been suspended. We recognize that for government requests for content removal, the appeals process needs to be expedited, since routing such a process through the regular court procedure may take additional time. However, as the Santa Clara Principles, the Manila Principles and the Electronic Frontier Foundation (EFF) all require in any instance of user content being taken down or user accounts suspended, intermediaries must at least notify users of takedown requests relating to their account or content.

While some of the major intermediaries already undertake this task voluntarily, we believe both the intermediary liability rules and the blocking rules should have a provision whereby any intermediary who falls within the scope of the legislation, must notify its users whenever it removes user content pursuant to a government request, or whenever it disables user accounts.

---

[76] Surya Prakash B.S, 'Making laws with sunset clauses' (*Livemint*, 3 January 2018) <https://www.livemint.com/Opinion/svjUfdqWwbbeeVzRjFNkUK/Making-laws-with-sunset-clauses.html> accessed on 4 November 2019

[77] Gennie Gebhart, 'Who Has Your Back? Censorship Edition 2019' (*Electronic Frontier Foundation*, 12 June 2019) <https://www.eff.org/wp/who-has-your-back-2019#scope> accessed on 4 November 2019

[78] Rule 3(4), The Information Technology [Intermediaries Guidelines (Amendment) Rules] 2018

Classifications of intermediaries

A single timeframe of content takedown, for *all* intermediaries, irrespective of the resources available to them, would prove to be counter-productive. As experiences with both the NetzDG and the Code of Countering Illegal Hate Speech Online show, some of the larger online companies would mostly be able to respond within a 24 hour timeframe. However, without sufficient interaction with the rest of the intermediaries and content hosting services, it would be impossible to gauge if such a timeframe would be able to deliver its objectives without collateral damage on the way these entities do their business and without avoiding inducing a chilling effect on speech.

In view of that, the law in question must propose a demarcation of the intermediaries, which would in turn decide the parameters of their liability under the law. We believe there are several ways of doing this. In the following subsections, we discuss some key classifications that we believe would be necessary for introducing nuance into the current regulatory system.

Technical architecture of the intermediary

The IT Act, which is the parent act to the rules, takes within its scope a large number of entities as 'intermediaries' for the purposes of legislation. This includes traditional intermediaries like social media websites, e-commerce platforms and ISPs, but also includes other entities like cyber cafes[79].

On a fundamental level, a classification can be made between these entities on the basis of the technical architecture of intermediaries and the functions they provide[80] For example, an ISP's function is very different from a search engine's. The business models operated by these intermediaries also become a key factor of differentiation.[81]

User-base of an intermediary

Like NetzDG does in Germany, a primary level of classification can be done based on the size of the user-base of the intermediary. This classification however, has seem to run into its own set of problems, including confusion regarding how this user-base would be calculated.[82]

---

[79] Section 2(w), IT Act 2008

[80] Rebecca MacKinnon, Elonnai Hickok et. al, 'Fostering Freedom Online: The Role of Internet Intermediaries', UNESCO Publishing <https://unesdoc.unesco.org/ark:/48223/pf0000231162/PDF/231162eng.pdf.multi> accessed on 4 November 2019

[81] Id

[82] Subhodeep Jash, 'Outsourcing Censorship, Attacking Civil Liberties: Germany's NetzDG' (*The Governance Post*, 27 October 2017) <https://www.hertie-school.org/the-governance-post/2017/10/outsourcing-censorship-attacking-civil-liberties-germanys-netzdg/> accessed on 4 November 2019; Lucinda Southern, 'Germany's law against hate speech faces hurdles' (*Digiday*, 3 January 2018) <https://digiday.com/media/netzwerkdurchsetzungsgesetz/> accessed on 4 November 2019

Additionally, as we have noted in our previous submission for the rules, there are several challenges in determining the active user-base of an intermediary. This includes mandating the intermediaries to regularly report the number of users they have, verifying the veracity of such information,[83] and also taking into account the fact that different intermediaries have in place different metrics to determine the user bases.

Facebook, for instance, has been reporting the number of daily active users (DAUs) as well as monthly active users (MAUs) for at least the past four years, as part of their representations to the Securities and Exchange Commission (SEC).[84] The metrics they use to measure these numbers is completely different from Twitter, which has only revealed the number of DAUs early this year, and has discontinued reporting of MAUs henceforth.[85]

Annual turnover

An alternative method of demarcation of intermediaries, would be by taking into account the annual turnover of these companies. Certain entities would be held to a higher level of scrutiny if their revenues touch a particular threshold for a given period, and accordingly, they would be expected to respond to content removal orders at a faster rate than those whose turnovers fall below this threshold.

We recognize that determining a finite threshold for this would be a challenge. Several Indian laws lay down different turnover rates for the purposes of classification, and extend exemptions to specific entities. The merger control notification under the Competition Act, 2002, for instance, exempts certain entities with a finite turnover rate from the application of section 5 of the Act.[86]

Another pertinent law that envisages exemptions from specific liability based on the turnover rate is the Personal Data Protection Bill ["PDP Bill"], 2018. Section 48(2) exempts entities with a turnover less than twenty lakhs rupees in the preceding financial year (and a few other specifics), from certain obligations under the Bill.[87]

The feasibility of such a threshold is however, debatable. Some commentators have pointed out that the threshold is fairly low,[88] and such a limited exemption would

---

[83] Gurshabad Grover (n 70)
[84] United States Securities and Exchange Commission, Transition Report for Facebook Inc. <https://www.sec.gov/Archives/edgar/data/1326801/000132680119000009/fb-12312018x10k.htm> accessed on 4 November 2019
[85] Ben Lovejoy, 'Twitter share price drops 10% as it reveals daily active users for the first time' (*9TO5Mac*, 7 February 2019) <https://9to5mac.com/2019/02/07/twitter-daily-active-users/> accessed 4 November 2019
[86] Ministry of Corporate Affairs Notification <https://www.cci.gov.in/sites/default/files/notification/SO%20673(E)-674(E)-675(E).pdf>
[87] Section 48(2), Personal Data Protection Bill 2018
[88] Rishabh Bailey et. al, 'Comments on the (Draft) Personal Data Protection Bill, 2018' (NIPFP) <https://www.medianama.com/wp-content/uploads/NIPFP-Submission-India-Draft-Data-Protection-Bill-Privacy-2018.pdf> accessed on 4 November 2019; Graham Greenleaf, 'GDPR-Lite and requiring strengthening – Submission on the draft Personal Data Protection Bill to the Ministry of Electronics and Information Technology (India)' <https://www.medianama.com/wp-content/uploads/Graham-Greenleaf-Submission-India-Draft-Data-Protection-Bill-Privacy-2018.pdf> accessed on 4 November 2019

accordingly continue to impact several smaller companies adversely.[89] In such light, uniformly importing any such exemption threshold from any other law may not prove to be productive.

A risk-based classification

Some commentators to the PDP Bill have suggested that instead of a revenue-based exemption, such immunity should essentially be a function of the risk involved in the data processing. In other words, the liability of an entity should be directly proportional to the likelihood of harm caused by a specific type of processing.[90]

Such a rationale for classification also finds place in several international documents. The UN Guiding Principles for Business and Human Rights, for instance, while arguing for a size-based gradation of enterprises for assessing their liability, also implores the assessment to take into account the severity of impact of the enterprise.[91] In the context of the intermediaries, this would mean that the greater the probability of harm being caused to the objects of legal protection, the higher degree of precaution the intermediary should deploy while developing their codes of ethics and similar policies.[92]

Thus, a fundamental level of classification should also take into account the reach, scope and potential impact of the intermediary's actions. We recognize that this determination might be context-specific, and can occasionally be difficult to regulate.

Gradation of sanctions in instances of violation

The NetzDG envisages sanctions for social media platforms if they display persisting failure to adhere to the law, and not for a single instance of failure. Similarly the IMCO opinions to the EU Terrorist Regulations also pushed for institution of sanctions in instances of systematic violations and *not* in case of a single case. It also provides carve-outs for online companies who have never received a removal order before, by allowing them a response time of 12 hours, as opposed to the one-hour general timeline.

We believe that a uniform timeframe of compliance, coupled with sanctions would go on to further disrupt the competition from the smaller intermediaries. As our stakeholder mapping indicates, the threat of sanctions may also have adverse effects on free speech online, resulting in potential over-censorship of legitimate speech.

---

[89] Ikigai Law, 'Comments of certain start-ups on the Personal Data Protection Bill, 2018: Consolidated views' <https://www.ikigailaw.com/comments-of-certain-start-ups-on-the-draft-personal-data-protection-bill-2018-consolidated-views/#acceptLicense> accessed on 4 November 2019; Dvara Research, 'Comments to the Ministry of Electronics and Information Technology (MEITy) on the draft Personal Data Protection Bill 2018, dated 27 July 2018, submitted by the Committee of Experts on a Data Protection Framework for India' <https://www.dvara.com/blog/wp-content/uploads/2018/10/Response-to-draft-Personal-Data-Protection-Bill_DvaraResearch.pdf> accessed on 4 November 2019

[90] NIPFP Submissions (n 79)

[91] Principle 14, UN Guiding Principles for Business and Human Rights

[92] Council of Europe, 'Recommendation CM/Rec(2018)2 of the Committee of Ministers to member States on the roles and responsibilities of internet intermediaries'

Section 79 of the IT Act and the rules and IT Act do not specify possible sanctions that the intermediary would be subject to should it fail to remove content within the given timeframe, beyond the loss of safe harbour under section 79 of the IT Act. It would nevertheless be useful for the law to identify situations when it would be appropriate for the authorities to prosecute an intermediary, and to what extent. This is in contrast to section 69A, which mandates imprisonment for seven years in the instance of failure to comply with the directions issued under the section.[93]

In tune with the global practices discussed above, we also recommend restricting sanctions to instances of systematic violations. We recognize that for critical content (defined below), the contours of what constitutes systematic violation may differ. The regulator must accordingly take into account the nature of content which the intermediary failed to remove, while assessing their liability.

Finally, we also recommend that for ensuring proportionality of punishments, the sanctions should be a direct function of the financial capacity and resources available to the intermediary.[94]

Gradation of the timeframe based on the nature of the content

For the majority of the laws reviewed, their subject matter was restricted to a specific kind of content, and the timeframe was accordingly applicable. The rules however, choose to club a vast spectrum of speech related offences under a singular timeframe of compliance.

Accordingly, we recommend that the 24-hour timeframe be replaced with a graded system, where certain kinds of content are treated with more urgency than others. Both interviews with intermediary #1 and intermediary #3 also revealed concurrence to this idea.

In particular reference to the gradation proposed by intermediary #1, it was interesting to note that despite being a relatively smaller intermediary, they were an advocate of an even shorter timeframe of 3 hours for specific kinds of content. Such a stance contrasts the civil society concerns of a uniform timeline dampening the business of a smaller intermediary, or making compliance difficult for them.

Due to lack of a substantial amount of interviews from similarly placed intermediaries, it is difficult to gauge if there is concurrence on this point. However, in our opinion, such a view can be further supplemented by reviewing the practices of organisations offering content moderation services to intermediaries.

[93] Section 69A(3), IT Act 2008
[94] For instance, see: Article 83, General Data Protection Regulation (GDPR)

Among the review of the stakeholder submissions, CCG was the only one who proposed a separate graded timeframe for content removal. According to them, the turnaround time should be:

Table 5: CCG's suggestion of a graded timeframe

Timeframe of response	Nature of the content
24-48 hours	Relating to unlawful acts endangering 'the sovereignty and integrity of India', 'security of the State' and potentially 'public order'
At least 14 days	Relating to other grounds under Article 19(2)

Here, a balance must be struck between the free speech rights of the users, the capabilities of the intermediaries and the enforcement concerns of the government. On the basis of the laws perused, accordingly, we recommend the following graded timeframe.

For critical content

For critical content, which relates to endangering the 'sovereignty and integrity of India', 'security of the State', and 'public order', and any content that relates to prohibited materials under Section 67B of the IT Act[95], we recommend a fixed turnaround time.

The exact metrics of the timeframe are difficult to formulate at this juncture, and it is also difficult to justify a 24-hour turnaround time for this from research coming out of other jurisdictions. One important fact that sets the jurisdictions reviewed apart from India is the diversity of languages in the latter. As our interview with intermediary #3 shows, the language of the content also becomes a significant contributing factor towards the determination of the turnaround time.

[95] *Section 67B: Punishment for publishing or transmitting of material depicting children in sexually explicit act, etc. in electronic form.*
Whoever,- (a) publishes or transmits or causes to be published or transmitted material in any electronic form which depicts children engaged in sexually explicit act or conduct or
(b) creates text or digital images, collects, seeks, browses, downloads, advertises, promotes, exchanges or distributes material in any electronic form depicting children in obscene or indecent or sexually explicit manner or
(c) cultivates, entices or induces children to online relationship with one or more children for and on sexually explicit act or in a manner that may offend a reasonable adult on the computer resource or
(d) facilitates abusing children online or
(e) records in any electronic form own abuse or that of others pertaining to sexually explicit act with children, shall be punished on first conviction with imprisonment of either description for a term which may extend to five years and with a fine which may extend to ten lakh rupees and in the event of second or subsequent conviction with imprisonment of either description for a term which may extend to seven years and also with fine which may extend to ten lakh rupees.

For other content

For content falling within the scope of the other reasonable restrictions under Article 19(2), i.e., decency and morality, contempt of court, defamation or incitement to an offence[96], we recommend a more relaxed timeframe. The subjective nature of these speech elements may mean that a unilateral judgment regarding the legality of such content may not be productive. Thus, the metrics of this can be arbitrated by the intermediary with the oversight authority on a case-to-case basis, in tune with the procedure laid down under the blocking rules.

Alternatively, we recommend the institution of a notice-and-notice regime for these subjective speech elements. For content that is not manifestly illegal, such a system reduces the burden on the intermediaries, while also striking a balance between the competing interests of the users and the government.[97] However, as experiences with the HDCA show, such a system should not mandate the intermediaries to include user's personal information as part of the counter-notice issued to the complaining authority, as that may lead to concerns regarding user privacy and government surveillance[98].

Conclusion

As indicated previously, the proposed Indian regime for legal requests of content takedown, with reference to a turnaround time, envisages a uniform framework, despite the fact that such a stance ignores several regulatory nuances. To inform our research therefore into what constitutes an appropriate turnaround time, we have accordingly attempted to review evidence from different jurisdictions and stakeholder submissions to the rules. In the course of the same, we came across several other concomitant factors that the law must account for, and must legislate to incorporate key nuances. These include:

- **Harmonizing the existing legal system for content takedowns:** Both section 69A and section 79 allows the government to remove unlawful content, but under very different procedures. In such light, we recommend that the two sections be harmonized to ensure a uniform regime of content takedown.
- **Mandated transparency reporting**: We believe such an exercise would be useful in assessing the efficacy of any turnaround time that the regulator may propose. Such reporting must be standardised across intermediaries to ensure optimal accessibility of information
- **Institution of a notifications and appeals mechanism**: Informing users when their content has been taken down and allowing them a chance to contest the same, ensures due process.

---

[96] Article 19(2), Constitution of India 1950
[97] Akriti Bopanna and Gayatri Puthran, 'Comparison of the Manila Principles to Draft of The Information Technology [Intermediary Guidelines(Amendment) Rules], 2018' (*Centre for Internet and Society*, 30 September 2019) <https://cis-india.org/internet-governance/files/draft-rules-and-manila-principles-1> accessed on 4 November 2019
[98] Danny O'Brien (n 65)

- **Classification of intermediaries:** A blanket regime of compliance has the potential of harming smaller intermediaries. We believe that the law should classify between these entities and affix their liabilities accordingly. We propose the following factors that must be taken into account:
    - Technical architecture of the intermediary.
    - User-base of an intermediary.
    - Revenue generated in one focus year.
    - The level of risk associated with the intermediary's actions.
- **Gradation of sanctions:** Fear of prosecution oftentimes prompts entities towards over-compliance. We believe that the law should not lay down sanctions for single instances of failure to comply with a particular turnaround time. Instead, systematic violations should be identified for any possible sanctions. Additionally, we believe that what would be a 'systematic violation' would be the function of the nature of content in question, and the regulator should factor that in as well.
- **Gradation of timeframes based on the nature of content**: Lastly, the turnaround time must factor in the nature of content in question. Accordingly we proposed:
    - For critical content: A fixed turnaround time.
    - For other content: An expeditious turnaround time, or alternatively, a notice-and-notice regime whenever it is appropriate.

Annexure 1: Indicative list of questions asked to the intermediary

1. Have you, at any point, been consulted during the law-making process, when the law directly pertains to social media platforms? Alternatively, have you, at any point, taken part in any open consultations on the intermediary liability regime?
2. What do you think on a general note, about the current regime and about the draft amendments?
3. What is the usual process undertaken by you to respond to these orders?
4. Has there been any instance where you decided against taking the content down? Was it because of any technical limitations on your part? Or was their procedural faults on the part of the government?
5. If in certain instances, it has taken you longer to respond, has it been because of the kind of content that has been in question? If yes, then what kind of content has taken you longer to respond to? In case it has been any other reason, can you elaborate on the same?
6. On that note, are there any internal policy demarcations regarding the kind of content that becomes subject to takedown orders? For instance, do you respond to 'terrorist' content on a faster scale, than content that is 'offensive' or 'defamatory'?
7. In case you have taken longer to respond, has there been any repercussions of the same? Do you think a single instance of failure should invite sanctions, or should there be a graded system?

8. In case the government chooses to finalize the draft amendments, and update the time-frame to 24 hours, how do you think it will impact your content moderation practices? Do you think you would need to deploy more resources to adhere to the new timeline? Would that in turn, impact your business model in any way? Would there be any other changes in the organisation that you can think of?
9. Do you think a shorter time-frame would impact the decision making process of the internal team responsible for content moderation?
10. In case there is an alternative to the current model, and the law, instead of keeping one timeline, makes the time of response 'expeditious'. Do you think that would be sufficient? If you think that an expeditious timeline works, then what could be the underlying guidelines for the same?