5

Safe Harbor and Content Moderation Regulation in India

Jhalak Mrignayani Kakkar, Shashank Mohan, and Vasudev Devadasan

5.1 INTRODUCTION

Social media platforms in India are regulated under the Information Technology Act, 2000 (IT Act). When enacted, the IT Act did not (and possibly could not) envisage the rise of social media platforms and thus, the legislation makes no specific reference to them. However, the IT Act regulates "intermediaries" – defined as entities that receive, store, transmit, or provide any service with respect to third-party content, or user generated content (UGC).[1] As the activities conducted by social media platforms predominantly fall within this definition, with respect to UGC, platforms have been regulated as intermediaries under the IT Act for the last two decades.

In 2021, the Indian government issued the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules 2021 (Intermediary Guidelines).[2] These Guidelines, which constitute delegated legislation under the IT Act expressly defines a "social media intermediary" as an intermediary that primarily enables online interaction between two or more users and allows them to upload, share, and disseminate content using its services.[3] The Intermediary Guidelines further

[1] The Information Technology Act, 2000, § 2(1)(w).

[2] Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules (2021), Part III [hereinafter Intermediary Guidelines]. The Intermediary Guidelines (2021) replaced earlier guidelines which had been in effect since 2011. The 2011 guidelines did not specifically define social media intermediaries. The Intermediary Guidelines 2021 introduce numerous conditions in addition to their predecessor, which have been explained in detail in this chapter. *See* discussion *infra* Sections 5.1.3, 5.4.

[3] Intermediary Guidelines, Rule 2(w). Subsequent guidance by the Indian government clarified that entities enabling commercial transactions or providing access to the internet, search engines, email services, and online storage services would not qualify as a social media intermediary. *See* Ministry of Electronics and Information Technology, *Frequently Asked Questions (FAQs) – The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021*, 10 (2021), https://www.meity.gov.in/writereaddata/files/FAQ_Intermediary_Rules_2021.pdf (last visited Nov. 3, 2021).

differentiate between (i) "intermediaries," (ii) "social media intermediaries," and (iii) "significant social media intermediaries" (SSMIs, i.e., social media intermediaries that have more than 5 million registered Indian users)[4] – imposing additional obligations on SSMIs.[5] The Guidelines also impose certain distinct obligations on SSMIs that provide messaging services.[6] Finally, the Guidelines distinguish between foreign and domestic SSMIs by requiring foreign SSMIs to have local officers who are resident in India – officers who may also be subject to personal liability.[7]

On October 28, 2022, amendments were made to the Intermediary Guidelines introducing additional compliance obligations for intermediaries in an attempt to make the rules and regulations/privacy policies more accessible to the users. These amendments also introduced a mechanism for the establishment of government-appointed grievance redressal committees (GACs). The GACs offer an appellate procedure to aggrieved users who are not satisfied with content-related decisions made by intermediaries.[8] Moreover, on April 6, 2023 amendments were made in the Intermediary Guidelines to include "Online Gaming Intermediaries" as intermediaries to impose due diligence obligations on them and a "Fact Checking Unit" of the central government to identify fake, false, or misleading information about any central government business to help intermediaries take down such content.[9] A number of petitioners, including a stand-up comedian and various news media organizations challenged the amendment introducing the Fact Checking Unit on constitutional grounds in the High Court of Bombay (a state level constitutional appeals court). On January 31, 2024, the High Court delivered a split verdict with one judge agreeing to strike down the provision and the other declaring it as constitutional and legally sound. The case will now proceed to a third judge for final determination.[10]

As social media platforms constitute intermediaries hosting and transmitting UGC under the IT Act, they are distinguishable from entities that publish their own content. Platforms are also regulated distinctly from print and broadcast media, which are governed by the Press Council of India Act (1978) and the Cable Television Networks (Regulation) Act (1995), respectively. Crucially, unlike

[4] Ministry of Electronics and Information Technology, S.O. 942(E) (notified on Feb. 25, 2021).
[5] Intermediary Guidelines, Rules 2(1)(v), 2(1)(w), 4. *See* discussion *infra* Sections 5.1.3, 5.4.
[6] Intermediary Guidelines, Rule 4(2). *See* discussion *infra* Section 5.4.4.
[7] Intermediary Guidelines, Rule 4(1)(a). *See* discussion *infra* Section 5.3.3.
[8] Ministry of Electronics and Information Technology, The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules (2021), https://www.meity.gov.in/writereaddata/files/IT%20Intermediary%20Rules%2C%202021%20updated%20on%2028.10.2022.pdf/ (last visited Feb. 16, 2024).
[9] *Id.*
[10] Sharmeen Hakim, *Bombay High Court Delivers Split Verdict on Please Challenging IT Rules Amendment, Govt to Not Notify "Fact Check Unit" for 10 Days*, LiveLaw, https://www.livelaw.in/top-stories/bombay-high-court-judgment-it-amendment-rules-2023-social-media-fake-news-fact-checking-unit-kunal-kamra-vs-union-of-india-248120/ (last visited Feb. 16, 2024).

publishers, broadcasters, and distributors who are typically strictly liable for content they publish, where intermediaries do not have "actual knowledge" of unlawful content on their network and comply with the conditions set out under the IT Act, they are exempt from liability.[11]

Finally, it is also relevant to note that at the time of writing, the legality and constitutionality of the Intermediary Guidelines remains under dispute. Several individuals, organizations, and platforms have filed petitions challenging various provisions of the Intermediary Guidelines in High Courts across the country. The Union government has requested that all these proceedings be clubbed and heard together by the Supreme Court of India.[12] In May 2022, the Supreme Court directed High Courts to stop hearing the challenges to the Intermediary Guidelines,[13] which would suggest the challenges to the Intermediary Guidelines will be heard by the Supreme Court in an omnibus fashion.

5.1.1 *Centrality of Safe Harbor to Platform Regulation*

The Intermediary Guidelines, coupled with the rules on government blocking of content,[14] form the core regulatory structure that governs platform conduct in India. Section 79 of the IT Act offers intermediaries conditional legal immunity (or "safe harbor") for unlawful UGC on their networks. One condition for safe harbor under Section 79 is compliance with the Intermediary Guidelines, which set out various additional obligations that intermediaries must comply with to avail of this safe harbor.[15] As set out in Sections 3 and 4, through the Intermediary Guidelines, the government has imposed wide ranging obligations on platforms as condition precedent for safe harbor. However, the power of the government to prescribe *when* platforms must remove content to retain safe harbor is circumscribed by the Supreme Court decision in *Shreya Singhal v. Union of India*.[16] The Court interpreted "actual knowledge" in Section 79 to mean a court order, effectively ruling that intermediaries could not be compelled to take down content at the behest of private complainants to retain safe harbor and that platforms would only lose safe harbor if they failed to remove content after receiving a government or court order.[17]

[11] Chinmayi Arun, *Gatekeeper Liability and Article 19(1)(A) of the Constitution of India*, 7 NUJS L. Rev. 73 (2015).

[12] Sohini Chowdhury, *IT Rules 2021: Supreme Court to Hear Centre's Plea to Stay Interim Orders Passed by High Courts on July 27*, LiveLaw (2022), https://www.livelaw.in/top-stories/supreme-court-it-rules-cable-tc-amendment-rules-online-media-ott-regulation-204329 (last visited Aug. 1, 2022).

[13] Skand Bajpai v. Union of India WP (Civil) 799 of 2020, order passed on May 9, 2022 (Supreme Court of India).

[14] Information Technology (Procedure and Safeguards for Blocking for Access of Information by Public) Rules, 2009 [hereinafter Blocking Rules].

[15] *See* discussion *infra*, Sections 5.3.1.3, 5.4.

[16] (2015) 5 SCC 1 (Supreme Court of India).

[17] *Id.*, ¶ 122. *See* discussion *infra*, Section 5.3.1.2.

This has limited the government's ability to institute a traditional notice-and-take-down regime for online platforms, where platforms risk losing safe harbor if they fail remove content pursuant to user complaints.

To avail of safe harbor under Section 79, an intermediary must:

1. Either limit its functionality to providing access to a communication system over which UGC is transmitted
   OR
   must not: (i) initiate the transmission; (ii) select the receiver of the transmission; and (iii) select or modify the information contained in the transmission;[18]
2. Comply with the Intermediary Guidelines;[19]
3. Upon receiving "actual knowledge" (interpreted by *Shreya Singhal* to mean a court order), or being notified by the appropriate government or its agency, of unlawful content on its network, remove the concerned material without vitiating any evidence;[20] and
4. Not aid, abet, or induce the commission of an unlawful act on its network.[21]

Additional detail on each of these limbs is provided in Section 3.1 ("Defence to liability"). As noted previously, a key condition to avail of immunity under Section 79 of the IT Act is compliance with the Intermediary Guidelines (i.e., delegated legislation). The Ministry of Electronics and Information Technology (MEITY) has relied on the Intermediary Guidelines to regulate platform behavior, imposing obligations ranging from transparency reporting and cooperation with law enforcement, to requiring users be provided with a hearing prior to their content being taken down, under the Intermediary Guidelines,[22] with platforms in breach of these obligations at risk of losing safe harbor. The obligations imposed on platforms under the Intermediary Guidelines are discussed in Sections 3 and 4.

The corollary of this approach is that the exclusive tool to hold social media platforms accountable is through the threat of losing safe harbor, which can only be enforced through individual actions brought before a court of law for hosting unlawful content. This approach may be contrasted with jurisdictions that employ a regulator to penalize platforms for a variety of problematic behavior. For an intermediary to be penalized in India, an action must be brought against it for hosting unlawful content that proves (i) the illegality of the content hosted by the intermediary, (ii) the secondarily liability of the intermediary in hosting the illegal

[18] The Information Technology Act, 2000, § 79(2)(a)–79(2)(b).
[19] The Information Technology Act, 2000, § 79(2)(c).
[20] The Information Technology Act, 2000, § 79(3)(b).
[21] The Information Technology Act, 2000, § 79(3)(a).
[22] *See* discussion *infra*, Sections 5.3.4, 5.4.

content, and (iii) the intermediary's ineligibility for safe harbor. The efficacy of this approach is analysed in Section 5.3.

The immunity provided by Section 79 is nonetheless vital for platform operations in India because, if platforms are ineligible for such immunity, they risk incurring both civil and criminal liability for content they host. Without such immunity, the regulatory environment around platforms is not suitable for the creation of a dynamic information and communication system that platforms provide today. This is because Indian law includes a wide range of content-related offenses. These content areas are discussed in detail in Section 2. While no platform has finally and definitively been held liable for hosting unlawful UGC, the wide range of criminalized content in India may incentivise platforms to comply with Section 79 and the Intermediary Guidelines to retain safe harbor.

5.1.2 *Content Removal by Government Orders*

The Indian government is also empowered to directly block content on the internet under Section 69A of the IT Act in the interests of "the defence, security, or sovereignty and integrity of India, its friendly relations with other States, public order, or to prevent the incitement of an offence related to these categories." This provision was used between 2020 and 2022 to block over one hundred mobile applications in India, including popular platforms such as TikTok, WeChat, PUBG, and Helo.[23] The Indian government claimed these applications had been transmitting user data to foreign servers in a manner prejudicial to the integrity and defense of India.[24] Given that these applications were overwhelmingly created by Chinese developers and the restrictions were imposed contemporaneously with a border dispute between India and China, media reports suggested that the blocking of mobile applications was a strategic move by the Indian government against Chinese platforms.[25] The provision has also been used to block popular websites

[23] *Chinese Apps Banned in India: India Bans 59 Chinese Apps including TikTok, WeChat, Helo*, The Economic Times (July 29, 2020), https://economictimes.indiatimes.com/tech/software/india-bans-59-chinese-apps-including-tiktok-helo-wechat/articleshow/76694814.cms (last visited Sept. 26, 2020); PUBG Mobile, *117 Chinese Apps Banned in India: Check the Full List*, The Indian Express (Sept. 5, 2020), https://indianexpress.com/article/technology/tech-news-technology/india-bans-pubg-mobile-116-chinese-apps-full-list-6580365/ (last visited Apr. 26, 2022); *Govt Bans 54 Chinese Apps over Security Threat Concerns*, Hindustan Times (Feb. 14, 2022), https://www.hindustantimes.com/india-news/govt-to-ban-54-chinese-apps-that-pose-threat-to-india-report-101644814634095.html (last visited Apr. 26 2022).

[24] Press Information Bureau, *Government Blocks 118 Mobile Apps Which Are Prejudicial to Sovereignty and Integrity of India, Defence of India, Security of State and Public Order*, Press Information Bureau (2020), https://pib.gov.in/pib.gov.in/Pressreleaseshare.aspx?PRID=1650669 (last accessed Apr. 26, 2022).

[25] Sameer Yasir & Hari Kumar, *India Bans 118 Chinese Apps as Indian Soldier Is Killed on Disputed Border*, New York Times (Sept. 2, 2020), https://www.nytimes.com/2020/09/02/world/asia/india-bans-china-apps.html (last accessed Apr. 26, 2022).

such as GitHub (for allegedly hosting terrorism related content), tweets by journalistic organizations and Members of Parliament, and even individuals protesting government policies.[26] On May 1, 2023, following the instructions from Ministry of Home Affairs, the central government banned fourteen apps under Section 69A allegedly on the basis that those apps were used by terrorists in Jammu & Kashmir.[27]

Intermediaries who fail to comply with directions under Section 69A can be fined and imprisoned for up to seven years.[28] While the Blocking Rules generally require that the user who uploaded the disputed content or the intermediary who hosted the content be provided with a notice and hearing,[29] in emergencies, the government has the power to dispense with a notice and hearing for the blocking of content.[30] Further, while blocking orders are required to be reasoned and in writing,[31] the orders themselves are confidential.[32]

In practice, there are few publicly reported instances of the government providing an *ex-ante* hearing to a user or voluntarily disclosing the blocking order.[33] However, where a website owner challenged the blocking of his satirical website under Section 69A, the Delhi High Court directed the MEITY to disclose the blocking order and grant the website owner a *post*-decisional hearing.[34] In 2022, Twitter challenged several blocking orders issued by the Indian government on the grounds that (i) the users whose content was being blocked were not notified, (ii) the content did not satisfy the substantive thresholds for illegality set out under Section 69A, and

[26] Aroon Deep, *Twitter Censors Tweets from MP, MLA, Editor Criticising Pandemic Handling*, MediaNama (2021), https://www.medianama.com/2021/04/223-twitter-mp-minister-censor/ (last visited Apr. 26, 2022); Revathi Krishnan, *Accounts of Prasar Bharati CEO, Caravan, Actor Sushant Singh among Those "Withheld" by Twitter*, The Print (2021), https://theprint.in/india/accounts-of-prasar-bharati-ceo-caravan-actor-sushant-singh-among-those-withheld-by-twitter/596638/ (last visited Apr. 26, 2022); PK Jayadevan & Neha Alawadhi, *Government Blocks over 60 Websites including Github & Sourceforge on Anti-terror Advisory*, The Economic Times (Dec. 31, 2014), https://economictimes.indiatimes.com/tech/internet/government-blocks-over-60-websites-including-github-sourceforge-on-anti-terror-advisory/articleshow/45704384.cms?from=mdr (last visited Apr. 26, 2022).

[27] Danny D'cruze, *Govt Blocks 14 Messenger Mobile Applications for Spreading Terror*, Business Today (May 1, 2023), https://www.businesstoday.in/technology/news/story/govt-blocks-14-messenger-mobile-applications-for-spreading-terror-379496-2023-05-01#:~:text=The%20Indian%20government%20has%20blocked,the%20Information%20Technology%20Act%2C%202000/ (last visited Feb. 18, 2024).

[28] The Information Technology Act, 2000, § 69A(3).

[29] Blocking Rules, rule 8; Aarathi Ganesan, *Summary: Twitter's Writ Petition before the Karnataka High Court*, MediaNama (2022), https://www.medianama.com/2022/07/223-why-is-twitter-suing-the-indian-government/ (last visited Jan. 10, 2023).

[30] Blocking Rules, rule 9.

[31] The Information Technology Act, 2000, § 69A(1).

[32] Blocking Rules, rule 16.

[33] Apar Gupta, *But What about Section 69A?*, The Indian Express (Mar. 27, 2015), https://indianexpress.com/article/opinion/columns/but-what-about-section-69a/ (last visited Jan. 10, 2023).

[34] Tanul Thakur v. Union of India WP (Civil) 13037 of 2019, decided on May 11, 2022 (High Court of Delhi).

(iii) blocking orders against entire user accounts (as opposed to specific posts) were disproportionate.[35] On June 30, 2023, the Karnataka High Court dismissed Twitter's challenge to the government's blocking orders. It imposed exemplary costs on Twitter as it considered the case speculative litigation.[36] The court upheld the power of the government to block entire accounts instead of specific tweets and elaborated that such powers were needed as some tweets may have "great propensity to incite anti-national feelings."[37] Between 2018 and October 2023, the central government has sent 13,660 blocking orders to social media platform X (earlier known as Twitter).[38]

5.1.3 *Content Prohibited by Intermediary Guidelines*

Under Rule 3(1)(b) of the Intermediary Guidelines, platforms are required to ensure that their terms of service prohibit users from uploading or sharing a wide range of content including content that is insulting; harmful to children; obscene; infringes any trademark, patent or copyright; threatens public order or the security of India; or violates any Indian law.[39] These categories (cumulatively "Intermediary Guidelines

[35] Ganesan, *supra* note 29.

[36] X Corp v. Union of India WP 13037 of 2022, decided on June 30, 2023 (High Court of Karnataka), https://www.livelaw.in/pdf_upload/wp13710-22-30-06-2023-478944.pdf. (last visited Feb. 18, 2024).

[37] Archit Lohani, *Decoding the Karnataka High Court Ruline: Blocking Accounts vs. Tweets*, Medianama (July 29, 2023), https://www.medianama.com/2023/07/223-karnataka-high-court-ruling-blocking-twitter-accounts-tweets/ (last visited Feb. 18, 2024).

[38] *Centre Issued 36800 Blocking Orders to Social Media Platforms Since 2018 under IT Act*, Scroll.in, https://scroll.in/latest/1060336/centre-issued-36800-blocking-orders-to-social-media-platforms-since-2018-under-it-act/ (last visited Feb. 18, 2024).

[39] Under Rule 3(1)(b) platforms are expected to prohibit the following categories in their ToS:

- Content that contains a software virus or code that is designed to interrupt, destroy, or limit the functionality of a computer resource;
- belongs to another person and to which the user does not have any right;
- Content that impersonates another person;
- Content that deceives or misleads the recipient about the origin of the message, is misinformation, or is patently false, untrue, or misleading;
- Content that relates to or encourages money laundering or gambling;
- Content that infringes on any trademark, patent, copyright or other proprietary rights;
- Content that is obscene, pornographic, paedophilic, invasive of another user's privacy (including bodily privacy), insulting or harassing of other users on the basis of gender, racially or ethnically objectionable, or promotes enmity between different groups on the grounds of religion or caste with an intent to incite violence;
- Content that is harmful to children;
- Content that threatens the unity, integrity, defence, security, or sovereignty of India, friendly relations with foreign States, or public order, or insults any other nation, or causes the incitement of a serious offence or prevents the investigation of an offence; and
- Content that violates any Indian law.

Prohibited UGC") form the broad umbrella of content that platforms are expected to restrict in their ToS.

Under the Intermediary Guidelines, platforms are legally required to inform their users, at least once a year, that noncompliance with the platform's ToS may result in the removal of noncompliant content or termination of the user's access to the platform.[40] Platforms only lose safe harbor if they fail to remove content after receiving "actual knowledge" of unlawful content (interpreted by the Indian Supreme Court in *Shreya Singhal* to mean a court order) or fail to comply with a government order for removal of content.[41] However, in practice, most large social media platforms will remove most of the above-mentioned categories of content pursuant to their voluntary content moderation activities where they believe such content violates their ToS.

In October 2022, the MEITY amended the Intermediary Guidelines to stipulate that intermediaries "shall make reasonable efforts to cause" their users not to "host, display, upload, modify, publish, transmit, store, update or share" content that constitutes Intermediary Guidelines Prohibited UGC.[42] (As noted above, the Intermediary Guidelines previously merely required platforms to prohibit such content in their ToS.)

Given the recent adoption of this text, there exists some ambiguity over what an obligation to make reasonable efforts to cause users not to publish or transmit Intermediary Guidelines Prohibited UGC involves. A literal interpretation of this language may suggest that the recent amendments change the legal obligation on platforms from – a requirement to include prohibitions against Intermediary Guidelines Prohibited UGC in their ToS – to an obligation to *prevent* users from uploading Intermediary Guidelines Prohibited UGC onto their networks. Such an interpretation may effectively create a strict liability standard for platforms because the hosting of unlawful content by a platform would be a violation of its obligation to prevent users from uploading unlawful content, leading to a breach of the Intermediary Guidelines and consequently a loss of safe harbor. However, this obligation to prevent users is qualified by the expression make "reasonable efforts."

Further, such an interpretation would conflict with Section 79 and other provisions of the Intermediary Guidelines. Section 79(1) of the IT Act expressly provides intermediaries immunity for hosting unlawful content. This immunity would be rendered ineffective if platforms lost this immunity simply upon a user uploading unlawful content onto their network. As Section 79(1) constitutes primary legislation, and the recent amendments amend delegated legislation (the Intermediary Guidelines), the Amendments cannot override the statutory scheme set out in

[40] Intermediary Guidelines, Rule 3(1)(c).

[41] The Information Technology Act, 2000, § 79(3); Intermediary Guidelines, Rule 3(1)(d); Shreya Singhal v. Union of India (2015) 5 SCC 1 (Supreme Court of India), ¶ 122. *See* discussion *infra*, Section 5.3.1.2.

[42] Intermediary Guidelines, Rule 3(1)(b).

Section 79. Similarly, Rules 3(1)(d) and 3(1)(g) of the Intermediary Guidelines expressly state that platforms are only required to remove unlawful content pursuant to a government or court order, or in the case of nonconsensual intimate images, pursuant to a user complaint.[43] Thus, despite the language introduced by the recent amendments suggesting that platforms have to *prevent* users from uploading unlawful content, a holistic reading of Section 79 and the Intermediary Guidelines would suggest that platforms are not required to ensure an absolute prohibition against Intermediary Guidelines Prohibited UGC on their networks but rather simply demonstrate that they have taken certain affirmative steps toward restricting such content.

5.2 PLATFORM RESPONSIBILITY FOR VARIOUS SUBJECT AREAS

A wide range of content is unlawful under Indian law. This includes online content (primarily regulated by offenses in the IT Act), and general application statutes such as the Indian Penal Code, 1860 (IPC),[44] which regulate content whether found on an online or offline medium. Given the wide range of unlawful content in India, social media platforms may be secondarily liable for UGC on their networks that violate Indian law unless they secure safe harbor under Section 79 of the IT Act. This is because civil or criminal proceedings may be initiated against a platform for hosting unlawful UGC unless the platform can demonstrate it qualifies for immunity under Section 79.[45] Section 79 immunity is applicable against both civil and criminal proceedings that may be brought against platforms.

However, platforms can avoid secondary liability for unlawful content by complying with Section 79 and taking down content upon receiving a court or government order.[46] The obligations of platforms to take down content do not change based on the subject matter of the content hosted except in the cases of (i) nonconsensual intimate content (which must be taken down within 24 hours of receiving a complaint)[47] and (ii) rape and child-sex-abuse material (which SSMIs must "endeavour" to proactively identify using automated tools).[48] Outside of these two categories, intermediaries, including social media platforms, are only required to take down

[43] Intermediary Guidelines, Rule 3(2).

[44] Recently, the Indian Parliament overhauled India's criminal laws by replacing the colonial era laws on criminal activity, criminal procedure, and the law of evidence. The new set of laws are yet to be notified and made operational. *See* Mayank Kumar, *It's Back to the Classroom for Delhi Police Officers to Learn New Criminal Laws, Unlearn IPC, CrPC*, The Print (Jan. 26, 2024), https://theprint.in/india/its-back-to-the-classroom-for-delhi-police-officers-to-learn-new-criminal-laws-unlearn-ipc-crpc/1933652/ (last visited Feb. 18, 2024).

[45] Google India Pvt Ltd v. Visaka Industries Ltd (2020) 4 SCC 162 (Supreme Court of India), ¶53; The Information Technology Act, 2000, § 81. There are certain minor carve outs with respect to copyright and patent actions that are not relevant to the present chapter.

[46] *See* discussion *infra*, Section 5.3.1.2.

[47] Intermediary Guidelines, Rule 3(2). *See* discussion *infra*, Section 5.2.4.

[48] Intermediary Guidelines, Rule 4(4). *See* discussion *infra*, Section 5.4.1.

content pursuant to a court or government order.[49] The remainder of this section lists content that is unlawful in India and then sets out the data protection obligations imposed on intermediaries.

5.2.2 *Hateful, Inciteful, and Defamatory Speech*

The IPC criminalizes:[50]

- content promoting enmity between – different religious, racial, linguistic groups,[51] caste or communities, or any two classes of people;[52]
- content intended to outrage religious feelings or beliefs;[53]
- content prejudicial to "national integration";[54] and
- content that is likely to cause "fear or alarm to the public" or incite individuals to breach the public peace.[55]

Indian law recognizes both civil and criminal defamation.[56] Content that intentionally insults, intimidates, or humiliates a member of a Scheduled Caste or a Scheduled Tribe (identified in the Constitution and various statutes), including the use of abuses involving caste names, is also criminalized in India.[57] Section 66A of the IT Act proscribed "grossly offensive" or "annoying" expression online; however, this provision was struck down by the Supreme Court of India in 2015 as an unconstitutionally vague and overbroad restriction on free expression.[58] The Supreme Court has also intervened in the case of Section 124A of the IPC, which criminalizes seditious speech (defined as speech that causes "disaffection towards the government"). In May 2022, while hearing a constitutional challenge to Section 124A, the Supreme Court ruled that Indian authorities should desist from instituting fresh cases during the pendency of the challenge.[59]

[49] Shreya Singhal v. Union of India (2015) 5 SCC 1 (Supreme Court of India), ¶ 122.
[50] Recently, the Indian Parliament overhauled India's criminal laws by replacing the colonial era laws on criminal activity, criminal procedure, and the law of evidence. The new set of laws are yet to be notified and made operational. Kumar, *supra* note 44.
[51] The Indian Penal Code, 1860, § 153A.
[52] The Indian Penal Code, 1860, § 505(2).
[53] The Indian Penal Code, 1860, §§ 298, 295A.
[54] The Indian Penal Code, 1860, § 153B.
[55] The Indian Penal Code, 1860, § 505(1).
[56] The Indian Penal Code, 1860, § 499; Subramaniam Swamy v. Union of India (2016) 7 SCC 221 (Supreme Court of India), ¶¶66–68.
[57] The Scheduled Castes and the Scheduled Tribes (Prevention of Atrocities) Act, 1989, §§ 3(1)(r)–3(1)(s).
[58] Shreya Singhal v. Union of India (2015) 5 SCC 1 (Supreme Court of India).
[59] S G Vombatkere v. Union of India WP (Civil) 682 of 2021, order passed on May 11, 2022 (Supreme Court of India).

5.2.3 *Platform Conduct during Elections*

Under Section 171G of the IPC, any person who publishes a statement they know or believe to be false with the intention of affecting the outcome of an election may be fined. Further, content that is "patently false or misleading in nature" or situations where a person "knowingly and intentionally communicates any misinformation" falls within the ambit of Intermediary Guidelines Prohibited UGC and platforms must both prohibit such content in their ToS and make reasonable efforts to cause users not to publish and share such content.[60] In April 2023, the government made amendments to the Intermediary Guidelines to insert a clause that obligates intermediaries to make reasonable efforts to not host content that is fake or false and is in respect to any business of the central government. Under this clause, the government will notify a fact-checking unit that will identify and communicate fake, false, or misleading information to be acted upon by intermediaries.[61]

More importantly, Indian elections have a high volume of misinformation being disseminated over private messaging platforms such as WhatsApp.[62] In an attempt to curb this misinformation, the Intermediary Guidelines require messaging platforms to trace the "originator" of messages.[63] This obligation is discussed further in Section 5.4.

While the Election Commission of India's Model Code of Conduct does prescribe certain restrictions on election-related speech,[64] these restrictions are applicable against electoral candidates, and platforms are not held secondarily liable for violations by candidates. Violations of the Model Code of Conduct are typically addressed through non-monetary penalties imposed directly on the candidate (e.g., suspension of campaigning). Similarly, while the use of social media by electoral candidates and political parties is scrutinized by the Election Commission of India, platforms do not have any election-specific obligations under Indian law.

However, in 2019, major online platforms such as Facebook, Google, WhatsApp, and ShareChat (through the Internet & Mobile Association of India) adopted a voluntary Code of Ethics that platforms agreed to adhere to during state and national elections in India.[65] The Code of Ethics has two key commitments. First, the platforms agreed to enforce the "cooling off period" mandated by Section 126 of

[60] Intermediary Guidelines, Rule 3(1)(b).
[61] Intermediary Guidelines, Rule 3(1)(b)(v).
[62] Vidya Narayanan et al., *News and Information over Facebook and WhatsApp during the Indian Election Campaign* (2019), https://demtech.oii.ox.ac.uk/wp-content/uploads/sites/93/2019/05/India-memo.pdf (last visited Feb. 18, 2024).
[63] Intermediary Guidelines, Rule 4(2). *See* discussion *infra* Section 5.4.4.
[64] Model Code of Conduct, Election Commission of India, https://eci.gov.in/mcc/ (last visited Jan. 10, 2023).
[65] Press Information Bureau, *"Voluntary Code of Ethics" by Social Media Platforms to Be Observed in the General Election to the Haryana & Maharashtra Legislative Assemblies and All Future Elections* (2019), https://pib.gov.in/pib.gov.in/Pressreleaseshare.aspx?PRID=1586297 (last visited Jan. 10, 2023).

the Representation of the People Act, 1951;[66] which prohibits the display of any election related content forty-eight hours prior to polling.[67] This is operationalized by allowing the Election Commission to directly notify platforms of election-related content during the cooling off period, and platforms have committed to take down the flagged content within three hours.[68] The Commission reported that during the 2019 national elections, 909 posts were taken down pursuant to this mechanism, suggesting that the Commission is ultimately only able to flag a small amount of content.[69]

The second key commitment found in the voluntary Code of Ethics is that platforms will only host political advertisements that have been pre-screened in accordance with the Election Commission's regulations.[70] Such pre-screening of political advertisements was previously applicable to television and has been extended to social media through the adoption of this voluntary Code of Ethics. Under the Code, platforms are also required to tag or label political advertisements so that viewers can distinguish between such advertisements from other content on the site.[71]

5.2.4 *Terrorism-related Content*

Section 66F of the IT Act criminalizes "cyber terrorism." This offense primarily pertains to *conduct* involving the unauthorized access to a computer network or the denial of access to a computer network that is likely to cause death, injury, or disrupt essential services, including critical information infrastructure.[72] However, the provision has sporadically been used against *content* on social media platforms, primarily against content that allegedly incites communal violence.[73] Where the provision is used against content, platforms may be held secondarily liable for cyberterrorism subject to their defense of safe harbor.

The Indian government remains conscious of the use of the internet to promote and facilitate terrorism, primarily responding to such situations by directly blocking

[66] Internet and Mobile Association of India, *Voluntary Code of Ethics – reg.* (Sept. 23, 2019), https://static.pib.gov.in/WriteReadData/userfiles/IAMAI-ECI%20VCE.pdf [hereinafter IAMAI Code of Ethics].

[67] The Representation of the People Act, 1951, § 126(1)(b).

[68] IAMAI Code of Ethics.

[69] Nalin Mehta, *Digital Politics in India's 2019 General Elections*, 54 Economic and Political Weekly (2019), https://www.epw.in/engage/article/digital-politics-indias-2019-general-elections (last visited Jan. 10, 2023).

[70] IAMAI Code of Ethics.

[71] IAMAI Code of Ethics.

[72] The Information Technology Act, 2000, § 66F.

[73] *Three Kashmiri Students Arrested in Agra for Celebrating Pakistan's Cricket Win against India*, Scroll.in, October 28, 2021, https://scroll.in/latest/1009069/three-kashmiri-students-arrested-in-agra-for-celebrating-pakistans-cricket-win-against-india (last visited Jan. 10, 2023); Mukesh Kumar, *Hisar Journalists Say Junk FIR as Cops Look for Colleague*, The Times of India (Apr. 12, 2021), https://timesofindia.indiatimes.com/city/chandigarh/journalists-say-junk-fir-as-cops-look-for-colleague/articleshow/82021430.cms?utm_source=contentofinterest&utm_medium=text&utm_campaign=cppst (last visited Jan. 10, 2023).

content under Section 69A of the IT Act. For example, in 2015, the government blocked thirty-two websites in India, including vinmeo.com, dailymotion.com, and github.com, until they removed content that Indian authorities alleged was ISIS propaganda.[74] The government has blocked YouTube channels, Facebook accounts, and Twitter accounts for allegedly engaging in coordinated disinformation campaigns that threaten national security.[75] These blocked accounts included accounts operated by organizations made illegal under India's primary anti-terrorism statute, The Unlawful Activities (Prevention) Act, 1967.[76] The Central Government has issued directions to 635 URLs from December 2021 till July 2023 for publishing fake news that was against national sovereignty.[77]

5.2.5 *Intimidation, Trafficking, Nonconsensual Intimate Content, Child Pornography, and Sexually Explicit Material*

The publishing of content depicting the private area of a person "under circumstances violating their privacy" is a criminal offense under the IT Act.[78] Under Rule 3(2) of the Intermediary Guidelines, any user can lodge a complaint with an intermediary against content that depicts the user in a state of nudity or committing a sexual act, including content that has been digitally altered to depict the user as such.[79] The intermediary must remove the complained-against content within twenty-four hours and implement a distinct mechanism for such complaints or risk losing safe harbor vis-à-vis this content.[80] In the case of SSMIs, the user must be allowed to track the status of their complaint by being assigned a unique ticket number for their complaint.[81] It is also relevant to note that the IPC criminalizes the publication of content that discloses the identity of victims of sexual violence or rape absent express authorization.[82]

[74] Kim Arora, *Government Blocks 32 Websites to Check ISIS Propaganda*, The Times of India (Jan. 1, 2015), https://timesofindia.indiatimes.com/tech-news/government-blocks-32-websites-to-check-isis-propaganda/articleshow/45712815.cms (last visited Jan. 10, 2023).

[75] Sarvesh Mathi, *MIB Blocks Twenty-Two YouTube Channels for Spreading Fake News*, MediaNama (2022), https://www.medianama.com/2022/04/223-mib-blocks-youtube-channels/ (last visited Jan. 10, 2023).

[76] *Id.*

[77] *Centre Issued Directives to Block 635 URLs since December 2021 for Spreading Fake News: Anurag Thakur*, The Hindu (July 27, 2023), https://www.thehindu.com/news/national/centre-issued-directives-to-block-635-urls-since-december-2021-for-spreading-fake-news-anurag-thakur/article67127248.ece/ (last visited Feb. 18, 2024).

[78] The Information Technology Act, 2000, § 66E.

[79] *See also* Vasudev Devadasan et al., *CCG Working Paper: Tackling the Dissemination and Redistribution of NCII* (2022), https://ccgdelhi.s3.ap-south-1.amazonaws.com/uploads/ccg-ncii-wp-16dec22-fn-332.pdf. Critically analyzing the overbroad nature of Rule 3(2) of the Intermediary Guidelines.

[80] Intermediary Guidelines, Rule 3(2)(b).

[81] Intermediary Guidelines, Rule 4(6).

[82] The Indian Penal Code, 1860, § 228A.

While India does punish extortion,[83] criminal intimidation,[84] online stalking,[85] trafficking,[86] and identity theft,[87] these offenses primarily apply to the conduct of individuals using the internet and are thus unlikely to give rise to content-related liability for platforms. While the draft Trafficking of Persons (Prevention, Protection and Rehabilitation) Bill (2021) punishes the publication of content that promotes trafficking,[88] the draft legislation has yet to be introduced into Parliament. However, India does criminalize the publication of (i) "obscene material" (content that is lascivious, appeals to the prurient interest, or tends to deprave or corrupt persons)[89] and (ii) sexually explicit material.[90] Thus, platforms could, in principle, be prosecuted for hosting obscene or sexually explicit material, with the ultimate imposition of liability being subject to the platforms' claim to safe harbor under Section 79 of the IT Act.

Finally, the possession or storage of child pornography is criminalized in India.[91] Thus, platforms may be prosecuted for hosting child pornography. Further, under the Intermediary Guidelines, SSMIs have a distinct obligation to "endeavor to deploy" automated tools to proactively identify rape and child sexual abuse material.[92] This obligation is discussed in Section 4.1 of the IT Act ("Obligation to detect certain content").

5.2.6 *Content Removals Pursuant to Court or Government Orders*

One of the preconditions to safe harbor under Section 79 of the IT Act is that platforms remove content upon receiving court or government orders.[93] Court or government orders directing content removal are not limited to a specific subject area. Courts may require intermediaries to takedown specific content pursuant to injunctions in defamation[94] or intellectual property suits,[95] the right to be

[83] The Indian Penal Code, 1860, § 383.
[84] The Indian Penal Code, 1860, § 505.
[85] The Indian Penal Code, 1860, § 354D.
[86] The Indian Penal Code, 1860, § 370.
[87] The Information Technology Act, 2000, § 66C.
[88] Ministry of Women and Child Development, *Draft Trafficking in Persons (Prevention, Care and Rehabilitation) Bill, 2021* (June 2021), § 29, https://wcd.nic.in/sites/default/files/DRAFT%20TRAFFICKING%20IN%20PERSONS%20%28PREVENTION%2C%20CARE%20AND%20REHABILITATION%29%20BILL%202021%20%281%29.pdf (last visited Feb. 18, 2024).
[89] The Information Technology Act, 2000, § 67.
[90] The Information Technology Act, 2000, § 67A.
[91] The Protection of Children from Sexual Offences Act, 2012, § 15; the Information Technology Act, 2000, § 67B.
[92] Intermediary Guidelines, Rule 4(4). *See* discussion *infra*, Section 5.4.1.
[93] The Information Technology Act, 2000, § 79(3); Intermediary Guidelines, Rule 3(1)(d).
[94] Subodh Gupta v. Herdsceneand CS (OS) 483 of 2019 decided on September 18, 2019 (High Court of Delhi); Zulfiqar Ahmad Khan v. Quintillion Business Media CS (OS) 642 of 2018, decided on December 14, 2018 (High Court of Delhi).
[95] Jagran Prakashan Ltd. v. Telegram FZ LLC CS (Comm) 146 of 2020, decided on May 29, 2020 (High Court of Delhi); Dept. of Electronics and Information Technology v. Star India Pvt. Ltd.

forgotten,[96] to remove nonconsensual intimate images,[97] or impose broader obligations to coordinate with government authorities to take down certain classes of content pursuant to public interest litigation.[98] Similarly, government orders have been issued against a wide range of content including (as noted above) Chinese mobile applications alleged to have national security implications[99] and the Twitter accounts of media organizations.[100]

5.2.7 *Data Protection Obligations*

As the Indian Supreme Court has ruled that the right to privacy is a fundamental right guaranteed by the Indian Constitution,[101] India has recently passed the Digital Personal Data Protection Act (DPDP) in 2023. The Personal Data Protection Bill was earlier introduced into India's Parliament in 2019 and scrutinized by a Joint Parliamentary Committee, which released its report in December 2021.[102] However, the Bill was subsequently withdrawn in August 2022[103] and replaced by the Digital Personal Data Protection Bill (2022).[104] Subsequently, by way of a fresh bill introduced by the government in Parliament, both houses passed India's first comprehensive data protection legislation in August 2023.[105] The DPDP creates obligations for data fiduciaries[106] and significant fiduciaries, provides rights for data principals,[107]

FAO (OS) 57 of 2015, decided on July 29, 2016 (High Court of Delhi); Snapdeal Pvt. Ltd. v. GoDaddy LLC CS (Comm) 176 of 2021, decided on April 18, 2022 (High Court of Delhi).

[96] Jorawer Singh Mundy v. Union of India WP (Civil) 3918 of 2021, decided on April 17, 2021 (High Court of Delhi).

[97] X v. Union of India WP (Cri) 1082 of 2020, decided on April 20, 2021 (High Court of Delhi).

[98] Sabu Mathew George v. Union of India (2017) 2 SCC 514 (Supreme Court of India) (advertising for pre-natal sex determination procedures); Registrar (Judicial) v. Union Ministry of Communications 2017 SCC Online 25298 Mad. (High Court of Madras) (content related to video games alleged to promote suicide); *In re*: Prajwala Letter dated 18.2.2015 SMW (Cri) 3 of 2015 (Supreme Court of India) (rape videos).

[99] *Chinese Apps Banned in India: India Bans 59 Chinese Apps including TikTok, WeChat, Helo*, *supra* note 23; Yasir and Kumar, *supra* note 25.

[100] Krishnan, *supra* note 26.

[101] K S Puttaswamy v. Union of India (2017) 10 SCC 1 (Supreme Court of India).

[102] *The Personal Data Protection Bill, 2019*, PRS Legislative Research, https://prsindia.org/billtrack/the-personal-data-protection-bill-2019 (last visited Jan. 10, 2023).

[103] *Govt Withdraws Data Protection Bill, 2021*, The Economic Times (Aug. 4, 2022), https://economictimes.indiatimes.com/tech/technology/govt-withdraws-data-protection-bill-2021/articleshow/93334281.cms (last visited Jan. 10, 2023).

[104] Ministry of Electronics and Information Technology, *Explanatory Note: The Digital Personal Data Protection Bill, 2022* (2022), https://www.meity.gov.in/writereaddata/files/Explanatory%20Note-%20The%20Digital%20Personal%20Data%20Protection%20Bill%2C%202022.pdf (last visited Jan. 10, 2023).

[105] Although the law has been notified by way of the president's assent, the law is currently not in operation as different clauses will be operationalized in a phased manner on future dates.

[106] The terminology used for data controllers or data processing entities in the Indian DPDP.

[107] The terminology used for data subjects or individuals to whom data relates to in the Indian DPDP.

and establishes a specialized adjudicator for resolving disputes related to data protection. It also provides for penalties that can be imposed by the specialized adjudicator for violations of the law. The law is currently not operational and is expected to be implemented in a phased manner in the next few months.[108] Platforms continue to have certain data protection obligations under the IT Act and the Information Technology (Reasonable security practices and procedures and sensitive personal data or information) Rules (2011) (Personal Data Rules). Section 43A of the IT Act requires corporate bodies possessing or handling "sensitive personal data" to implement reasonable security practices. The DPDP repeals section 43A of the IT Act and will become the primary legislation for data protection in India, once it is made operational by the government. The DPDP Act provides a national framework for processing personal data, replacing limited categories of "sensitive data" covered under the Personal Data Rules.

The Personal Data Rules define "sensitive personal data" as including passwords, financial information, sexual orientation, medical records, and biometric information. Entities that collect, store, or handle sensitive personal data must (i) collect such information for a lawful purpose; (ii) disclose to users the fact that information is collected, the purpose for which it is collected, and the intended recipients of the information; (iii) only retain sensitive personal data for the time it is necessary for the purpose collected; (iv) allow users to correct incorrect or deficient information upon request; and (v) provide a grievance redressal mechanism.[109] However, these obligations do not apply to entities that collect personal data "under a contractual obligation with another Indian or foreign company,"[110] and thus, are only applicable to entities that directly collect data from users.[111] If platforms fail to comply with the Personal Data Rules, they may be liable to compensate users for any losses stemming from the disclosure of sensitive personal data.[112]

5.3 ENFORCEMENT OF PLATFORM RESPONSIBILITY

Compliance with the Intermediary Guidelines constitutes a precondition for safe harbor under Section 79 of the IT Act. Therefore, the threat of losing safe harbor

[108] Digital Personal Data Protection Act, 2023, section 1(2).

[109] Information Technology (Reasonable security practices and procedures and sensitive personal data or information) Rules, 2011; Collection & Processing in India – DLA Piper Global Data Protection Laws of the World, https://www.dlapiperdataprotection.com/index.html?t=collection-and-processing&c=IN (last visited Jan. 10, 2023).

[110] Ministry of Communications & Information Technology, *Clarification on Information Technology (Reasonable Security Practices and Procedures and Sensitive Personal Data or Information) Rules, 2011 under Section 43A of the Information Technology Act, 2000* (Aug. 24, 2011), https://www.meity.gov.in/writereaddata/files/PressNote_25811.pdf.

[111] Aditi Chaturvedi, *GDPR and India* (2017). https://cis-india.org/internet-governance/files/gdpr-and-india (last visited Jan. 10, 2023).

[112] The Information Technology Act, 2000, § 43A.

under Section 79, which may lead to platforms being held liable for unlawful UGC on their networks, is the primary method of enforcing platform compliance with the various obligations outlined in the Intermediary Guidelines.

5.3.2 *Defense to Liability*

As noted in Section 5.1, to avail of safe harbor, an intermediary must (i) not initiate the transmission, select the receiver of the transmission, or modify the information contained in the transmission; (ii) comply with the Intermediary Guidelines; (iii) remove content upon receiving "actual knowledge"; and (iv) not aid or abet the commission of an unlawful act on its network.[113]

5.3.3 *Neutrality and Moderation*

The requirement that platforms must not initiate the transmission, select the receiver of the transmission, or select or modify the information in the transmission is analogous to the requirements of neutrality in Article 12 of the European E-Commerce Directive (Mere conduit).[114] Section 79 does not have an express equivalent to Article 14 of the Directive (Hosting), wherein even platforms that are not mere conduits can avail of safe harbor provided they remove content upon receiving actual knowledge. Rather, the text of Section 79 requires intermediaries to be *both* mere conduits *and* remove content upon receiving actual knowledge. However, as noted above, no platform has been denied safe harbor due to its interference with content, and commentators have argued that even hosting platforms should be able to avail of safe harbor under Section 79.[115]

Furthermore, the Intermediary Guidelines, introduced in 2021, clearly state that the removal of any Intermediary Guidelines Prohibited UGC will not amount to a breach of the neutrality required of Section 79.[116] The Guidelines thus recognize and promote voluntary content moderation by platforms. It remains unclear whether the use of recommender systems would violate the conditions of neutrality required by Section 79. On the one hand, recommender systems may amount to selecting the contents of a transmission. However, no court has specifically returned a finding that a platform's recommender system violates the neutrality requirements of Section 79. Similarly, the Indian government has neither suggested that such systems may lead to the loss of safe harbor or attempted to regulate them through the Intermediary Guidelines.

[113] *See supra*, Section 5.1.1.

[114] The Digital Services Act of the European Union contains the exact same language as in the E-Commerce Directive.

[115] Chinmayi Arun & Sarvjeet Singh, *NoC Online Intermediaries Case Studies Series: Online Intermediaries in India*, 11 (2015), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2566952.

[116] Intermediary Guidelines, Rule 3(1)(d) (third proviso).

5.3.4 Notice and Actual Knowledge

Neither Section 79 nor the IT Act defines the term "actual knowledge." Under the previous iteration of the Intermediary Guidelines (adopted in 2011), "actual knowledge" was set out to mean a complaint by another internet user, effectively setting up a traditional notice and takedown regime where platforms were required to remove content pursuant to private complaints.[117] However, in 2015, the Supreme Court of India in *Shreya Singhal v. Union of India* interpreted "actual knowledge" to mean a court order, effectively ruling that intermediaries would not lose safe harbor unless they failed to comply with a removal order by a court or authorized government agency.[118] This shifted the burden of determining illegality from intermediaries to courts and the government and increased the protection afforded to intermediaries as they were no longer legally required to remove content pursuant to private complaints,[119] although they remained free to do so in accordance with their ToS (i.e., voluntary content moderation).

In 2021, the Indian government codified the interpretation in *Shreya Singhal*, noting that platforms are *only* required to take down content pursuant to a court or government order.[120] However, pursuant to Rule 3(2) of the Intermediary Guidelines and the decisions of courts, platforms are nonetheless deemed to have "actual knowledge" and required to remove content pursuant to a private notice in the case of copyright infringing content[121] and nonconsensual intimate images.[122] As discussed above, the legal position again evolved in October 2022, when the MEITY amended the Intermediary Guidelines to stipulate that platforms must make reasonable efforts to cause their users not to host or transmit Intermediary Guidelines Prohibited UGC. The impact of this recent change has been discussed Section 5.1.3.

The October 2022 amendments to the Intermediary Guidelines also stipulate that, where a complaint pertains to a request to remove Intermediary Guidelines Prohibited UGC, the complaint shall be "acted on" and "redressed" within seventy-two hours.[123] The Supreme Court in *Shreya Singhal* expressly disapproved of this approach, noting that platforms receive a high volume of user complaints,

[117] Arun & Singh, *supra* note 115. This notice and takedown regime neither specified a counter-notice system or a put-back (reinstatement) requirement, thus risking legal content being removed at the behest of private complainants.

[118] Shreya Singhal v. Union of India (2015) 5 SCC 1 (Supreme Court of India).

[119] Kyung-Sin Park, *From Liability Trap to the World's Safest Harbor: Lessons from China, India, Japan, South Korea, Indonesia, and Malaysia*, Oxford Handbook of Online Intermediary Liability 250 (Giancarlo Frosio, ed. 2020).

[120] Intermediary Guidelines, Rule 3(1)(d).

[121] The Copyright Act, 1957 § 52(1)(c); The Copyright Rules, 2013, Rule 75; Myspace Inc. v. Super Cassettes Industries Ltd. 2016 SCC Online Del. 6382 (High Court of Delhi); Aradhya Sethia, *The Troubled Waters of Copyright Safe Harbors in India*, 12 Journal of Intellectual Property Law & Practice 398–407 (2017).

[122] Intermediary Guidelines, Rule 3(2).

[123] Proposed Amendments, Rule 3(2)(i).

and this would effectively lead to platforms deciding which complaints were legitimate and which were not, effectively determining what speech was legal and what speech is not.[124] The amendments state that platforms may institute "appropriate safeguards" to avoid abusive complaints by users.[125] However, short time frames to decide complaints against content has been proven to result in platform overcompliance with removal requests.[126]

5.3.5 *Additional Conditions for Safe Harbor in Intermediary Guidelines*

The Intermediary Guidelines also stipulate other conditions platforms must comply with to secure safe harbor, including (i) data retention obligations;[127] (ii) cooperation with law enforcement;[128] reporting of cyber security incidents;[129] and, in the case of SSMIs, (iii) appointing local compliance and grievance officers;[130] (iv) providing users with notice prior to taking down their content pursuant to ToS violations;[131] (v) publishing transparency reports;[132] (vi) endeavouring to proactively detect rape and child-sex abuse material;[133] and, for SSMIs providing messaging services, (vii) identification of the first originator of messages.[134]

5.3.6 *Efficacy of Enforcement*

The IT Act and the Intermediary Guidelines rely on the risk of losing safe harbor as the primary regulatory tool to govern platform behavior, imposing varied obligations (see Sections 3.1.3 and 4) on platforms as prerequisites to safe harbor. However, given that the loss of safe harbor is determined on a case-by-case basis, and the lengthy nature of litigation in India, no platform has definitively been held liable for hosting unlawful content. For example, in 2008, criminal defamation proceedings were instituted against Google for content on its Google Groups platform. Google sought to have the criminal complaint summarily quashed. The issue of whether the charges against Google should be summarily quashed or decided by trial took over a decade to decide, with the Supreme Court ultimately ruling that a trial should be conducted.[135]

[124] Shreya Singhal v. Union of India (2015) 5 SCC 1 (Supreme Court of India).
[125] Proposed Amendments, Rule 3(2)(i).
[126] Arun & Singh, *supra* note 115; Rishabh Dara, *Intermediary Liability in India: Chilling Effects on Free Expression on the Internet*, SSRN Journal (2011), http://www.ssrn.com/abstract=2038214 (last visited Feb. 18, 2023).
[127] Intermediary Guidelines, Rules 3(1)(g)–3(1)(h).
[128] Intermediary Guidelines, Rule 3(1)(j).
[129] Intermediary Guidelines, Rule 3(1)(l).
[130] Intermediary Guidelines, Rule 4(1)(a)–(c).
[131] Intermediary Guidelines, Rule 4(8). *See* discussion *infra* Section 5.4.2.
[132] Intermediary Guidelines, Rule 4(1)(d). *See* discussion *infra* Section 5.4.3.
[133] Intermediary Guidelines, Rule 4(4). *See* discussion *infra* Section 5.4.1.
[134] Intermediary Guidelines, Rule 4(2). *See* discussion *infra* Section 5.4.4.
[135] Google India Pvt. Ltd. v. Visaka Industries (2020) 4 SCC 162 (Supreme Court of India).

This dispute highlights how the nature of litigation in India coupled with the legal resources of platforms may render intermediary liability (i.e., the risk of liability enforced through private lawsuits) a weak regulatory tool to regulate platform behavior. However, there is some evidence to suggest that the government may believe that a loss of safe harbor either for content or noncompliance with the due diligence rules under the Intermediary Guidelines opens up the platform to liability for all content on the platform,[136] with the MEITY having issued Twitter multiple warnings to "comply with the Intermediary Guidelines or be liable to punishment under the IT Act."[137] However, such an understanding would be contrary to both the principles of secondary liability and the text of the IT Act.

Finally, it is also relevant to note that the IT Act applies to "any offence committed outside India."[138] Additionally, the IPC also applies to any offenses that "target computer resources located in India."[139] Thus, both statutes envisage extra-territorial application in certain situations. However, as the primary mechanism to regulate platform conduct is currently Section 79 and the Intermediary Guidelines, which are in the form of prerequisites to safe harbor against lawsuits initiated against platforms in India, India's regime of platform regulation relies on platforms being sued for hosting or transmitting unlawful content and being subject to the jurisdiction of Indian courts when this occurs.

5.3.7 *Additional Enforcement Methods*

In addition to the loss of safe harbor, there exist three methods through which Indian authorities ensure that platforms comply with specific obligations. First, noncompliance with a government direction for content removal under Section 69A of the IT Act is punishable with a prison term of up to seven years and a fine.[140] Similarly, if a platform does not comply with an order of a court, contempt proceedings may be initiated against it.[141] Finally, under the Intermediary Guidelines, SSMIs are required to appoint a Chief Compliance Officer who is a resident in India.[142] This Officer may be held personally liable in any proceedings relating to unlawful UGC on the platform's network if the Officer fails to ensure the

[136] Surabhi Agarwal, *Twitter Has Failed to Comply with Intermediary Guidelines: Ravi Shankar Prasad*, Times of India (June 17, 2021), https://economictimes.indiatimes.com/tech/technology/twitter-has-failed-to-comply-with-intermediary-guidelines-ravi-shankar-prasad/articleshow/83566900.cms?from=mdr (last visited February 18, 2024).

[137] Aashish Aryan & Surabhi Agarwal, *Twitter India Given "Last Chance" to Follow IT Rules*, ETTelecom.com (June 29, 2022), https://telecom.economictimes.indiatimes.com/news/twitter-india-given-last-chance-to-follow-it-rules/92532133 (last visited February 18, 2023).

[138] The Information Technology Act, 2000, § 1(2).

[139] The Indian Penal Code, 1860, § 4(3).

[140] The Information Technology Act, 2000, § 69(3).

[141] Facebook Inc v. Swami Ramdev FAO (OS) 212 of 2019, decided on January 28, 2020 (High Court of Delhi).

[142] Intermediary Guidelines, Rule 4(1)(a).

platform acts with "due diligence" in complying with the IT Act and the Intermediary Guidelines.[143] However, no liability will be imposed on the Compliance Officer without the Compliance Officer being granted a hearing.[144]

5.4 DETECTION AND MODERATION OF UNLAWFUL UGC

The Intermediary Guidelines, compliance with which is necessary for platforms to avail of safe harbor under Section 79, impose certain obligations on SSMIs with respect to content moderation. These obligations are not imposed on ordinary intermediaries (that do not perform social media functions or have less than 5 million Indian users).

5.4.1 *Obligations to Detect Certain Content Using Automated Tools*

SSMIs are required to "endeavour to deploy technology-based measures" to "proactively identify" content that (i) depicts rape or child sexual abuse material or (ii) is *identical* to content that either a court or government order directed be removed.[145] SSMIs are required to disable access to these two categories of content and inform users trying to access this content why the content has been blocked.[146] This best-efforts mandate to use automated tools to detect and remove content is subject to certain safeguards: (a) The action taken by the SSMI must be proportionate to the free speech and privacy interests of internet users;[147] (b) the automated tools used by the SSMI must be subject to "appropriate human oversight" and periodic review of these automated tools;[148] and finally, (c) the automated tools used by the SSMIs are to be evaluated to ensure "accuracy and fairness," guard against "the propensity of bias and discrimination," and determine their impact on privacy and security.[149] While the inclusion of these safeguards is commendable, in the absence of a designated regulator with meaningful oversight and enforcement powers, it is hard to determine whether these safeguards are complied with in practice.

5.4.2 *Responsibilities When Moderating*

Where an SSMI seeks to remove any Intermediary Guidelines Prohibited UGC voluntarily from its platform, Rule 4(8) of the Intermediary Guidelines requires the SSMI to provide the user who uploaded the relevant content a notice explaining the

[143] Intermediary Guidelines, Rule 4(1)(a).
[144] Intermediary Guidelines, Rule 4(1)(a).
[145] Intermediary Guidelines, Rule 4(4).
[146] Intermediary Guidelines, Rule 4(4).
[147] Intermediary Guidelines, Rule 4(4) (first proviso).
[148] Intermediary Guidelines, Rule 4(4) (second proviso).
[149] Intermediary Guidelines, Rule 4(4) (third proviso).

grounds for removal *before* the SSMI removes the content.[150] The user must also be provided with an "adequate and reasonable opportunity to dispute" the removal of their content and seek reinstatement if the content has already been removed.[151] Such disputes must be decided within fifteen days.[152] The Resident Grievance Officer of the SSMI is expected to oversee the dispute settlement mechanism under Rule 4(8).[153]

Despite the Intermediary Guidelines being in operation for more than a year, there is no evidence that SSMIs are complying with this notice and hearing requirement. One potential reason for this could be that the consequence of noncompliance with Rule 4(8), as with any provision of the Intermediary Guidelines, is a loss of safe harbor. In other words, failure to provide notice and hearing under Rule 4(8) could lead to a platform losing its immunity for hosting unlawful content. However, when a platform voluntarily removes content, it is *not* hosting this content and has removed unlawful content *prior* to when it is legally required to do so (i.e., prior to a court or government order). Therefore, it cannot be held secondarily liable for unlawful content and has few incentives to comply with the conditions necessary to avail of safe harbor. Thus, the loss of safe harbor flowing from a breach of Rule 4(8) may be inconsequential to an SSMI where it has already voluntarily decided to not host content.

In October 2022, the MEITY amended the Intermediary Guidelines to allow users to appeal against platform decisions to government-appointed Grievance Appellate Committee(s) (GACs).[154] On January 27, 2023, MEITY notified the establishment of three GACs in India,[155] which run completely online[156] and are face-less in their operation. Beyond the constitution of the three GACs, there isn't any other information in the public domain about their functioning or the decisions made by them.[157] According to Rule 3A(3), a user may appeal against any decision taken by a platform's Grievance Officer,[158] suggesting that a user can both appeal a platform's decision to remove content but also a platform's decision to *not* remove content in response to a user complaint. Appeals must be initiated within thirty days of being notified of the platform's decision,[159] and the GACs shall "endeavor" to decide the appeal within thirty days[160] pursuant to an online dispute resolution

[150] Intermediary Guidelines, Rule 4(8)(a).
[151] Intermediary Guidelines, Rule 4(8)(b).
[152] Intermediary Guidelines, Rule 4(8)(b).
[153] Intermediary Guidelines, Rule 4(8)(c).
[154] Intermediary Guidelines, Rule 3A.
[155] Establishment of Grievance Appellate Committees, Ministry of Electronics and Information Technology, Government of India, https://www.meity.gov.in/writereaddata/files/243258.pdf.
[156] Grievance Appellate Committee, Government of India, https://gac.gov.in/.
[157] On logging-in with an India based mobile number, the online portal does display the number of total appeals files and total appeals disposed of.
[158] Intermediary Guidelines, Rule 3A(3).
[159] Intermediary Guidelines, Rule 3A(3).
[160] Intermediary Guidelines, Rule 3A(4).

mechanism.[161] Each GAC shall consist of three members; two members shall be independent (but appointed by the Indian government), and one member shall be *ex-officio* (their membership of the GAC will be automatic by virtue of the office they hold).[162] GACs may seek assistance from any person having the requisite qualifications or experience in the subject matter being adjudicated.[163]

The creation of the GACs raises several concerns. First, it is unclear how the independence of GAC members will be secured. For example, selection by an independent body, disclosure of conflicts of interests, security of tenure and salary, and oath of office are some traditional methods to secure independence, but the Intermediary Guidelines do not provide for any of these safeguards in relation to the GACs.[164] Such independence is vital to protect the rule of law as the Indian government, or its instrumentalities, may be litigants before the GACs. Second, while the Intermediary Guidelines does contemplate more than one GAC, it is unclear how the GACs will deal with a large volume of appeals. Platforms make millions of moderation decisions every day and even if a small fraction of these are appealed to the GACs, it may create significant state capacity issues. Third, the Intermediary Guidelines do not expressly provide for basic due process safeguards with respect to the operation of the GACs, such as a notification to the person whose content is under dispute[165] or a written, reasoned order that is publicly available. Although there are three GACs currently in operation in India, beyond their constitution, there isn't any other information available about them in the public domain.

Finally, the Intermediary Guidelines (since October 2022) also require intermediaries to respect the constitutional rights of Indian citizens.[166] While an individual has sued to enforce his constitutional free speech rights against a platform's moderation decision (citing the platform's power over public speech), this case is still pending before the Delhi High Court.[167] Under current constitutional doctrine, Indian citizens may not enforce their constitutional free speech rights against private social media platforms.[168]

[161] Intermediary Guidelines, Rule 3A(6).
[162] Intermediary Guidelines, Rule 3A(2).
[163] Intermediary Guidelines, Rule 3A(5).
[164] Vasudev Devadasan & Bilal Mohamed, *Comments to the MEITY on the Proposed Draft for Amendment in Part-I and Part-II of the information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021* (2008), https://ccgdelhi.s3.ap-south-1.amazonaws.com/uploads/ccgnlud-comments-draftamendments-itrules2021–6jul22–301.pdf.
[165] *E.g.*, where User 1 complaints against User 2's content, the platform refuses to remove the content, and User 1 appeals to the GAC, the Intermediary Guidelines do not expressly require that User 1 be notified of the proceeding despite it being User 1's content (and consequently free speech rights) in dispute.
[166] Intermediary Guidelines, Rule 3(1)(n).
[167] Shravya Reddy, *Does Twitter Perform Public Functions? The Sanjay Hegde Case*, Bar and Bench (Nov. 13, 2020), https://www.barandbench.com/columns/does-twitter-perform-public-functions-the-sanjay-hegde-case (last visited Jan. 10, 2023).
[168] *See* Constitution of India (1950), art. 12; Ananth Padmanabhan, Rights: Breadth, Scope, and Applicability, Oxford Handbook on the Indian Constitution (Sujit Choudhry, Madhav Khosla

5.4.3 *Additional Obligations on SSMIs*

SSMIs are required to publish reports documenting their voluntary content moderation activities and responses to user complaints.[169] However, an analysis of these reports suggests they reveal more about the *scale* of platform moderation in India than they do about the *quality* of moderation.[170] SSMIs are also required to provide a user with a "demonstrable and visible mark of verification" (akin to Twitter's "blue-tick") if the user voluntarily verifies their account using "any appropriate mechanism" including an Indian mobile number.[171] Finally, as noted above, SSMIs are also required to appoint a Resident Grievance Officer and a Chief Compliance Officer, who are residents in India,[172] and a nodal contact person to facilitate coordination with law enforcement.[173] However, only the Chief Compliance Officer may be held personally liable.[174]

5.4.4 *Obligations on Messaging Platforms*

Rule 4(2) of the Intermediary Guidelines requires SSMIs that provide services "primarily in the nature of messaging" to "enable the identification of the first originator" of content on their platforms when directed by a court or an order passed under Section 69 of the IT Act ("power to issue directions for interception, monitoring, or decryption").[175] Where the first originator of unlawful content is located outside India, whomsoever is the first originator within India shall be deemed to be the first originator with respect to the content in question.[176]

An order directing the identification of an originator under Rule 4(2) may be passed for the purposes of (i) prevention, detection, investigation, prosecution or punishment of an offense and (ii) where such offense is related to the sovereignty, integrity, or security of the Indian State, its relation with foreign States, public order, or any offense relating to rape or sexually explicit material punishable by a prison term of five or more years.[177] Rule 4(2) further states that an identification order shall not be passed where a less intrusive means of identifying the first originator is effective[178] and that the SSMI shall not be required to disclose the contents of any

& Pratap Bhanu Mehta, eds. 2016). *Cf.* Kaushal Kishore v. State of Uttar Pradesh WP (Criminal) 113 of 2016, decided on January 3, 2023 (Supreme Court of India).

[169] Intermediary Guidelines, Rule 4(1)(d).

[170] Vasudev Devadasan, *Compliance Reports by Social Media Platforms Are Unhelpful*, MediaNama (Apr. 18, 2022), https://www.medianama.com/2022/04/223-transparency-reports-social-media-platforms-unhelpful/ (last visited Feb. 18, 2023).

[171] Intermediary Guidelines, Rule 4(7).

[172] Intermediary Guidelines, Rules 4(1)(a), 4(1)(c).

[173] Intermediary Guidelines, Rule 4(1)(b).

[174] Intermediary Guidelines, Rule 4(1)(a).

[175] Intermediary Guidelines, Rule 4(2).

[176] Intermediary Guidelines, Rule 4(2) (fourth proviso).

[177] Intermediary Guidelines, Rule 4(2) (first proviso).

[178] Intermediary Guidelines, Rule 4(2) (second proviso).

message or any other information regarding the content originator or any information related to its other users.[179]

Critics of the Rule have pointed out that messaging platforms providing end-to-end-encrypted services cannot trace originators on their platform[180] and that this is beyond the scope of technical assistance platforms are required to provide law enforcement under Indian law.[181] Commentators have also argued that both the methods proposed for the implementation of this requirement (assigning hash values to every unique message and affixing encrypted originator information to messages)[182] are easily circumvented, require significant technical changes to the architecture of messaging services, offer limited investigatory or evidentiary value, and will likely undermine the privacy and security of all users to catch a few bad actors.[183] Facebook and WhatsApp have challenged the legality and constitutionality of Rule 4(2) in the Delhi High Court.[184] As discussed in Section 5.1, the central government has requested these challenges be transferred to the Supreme Court and heard alongside other challenges to the Intermediary Guidelines. Recently, the Indian government suggested that it may make use of Rule 4(2) to ask messaging platforms to identify the first originator of messages carrying deepfakes of Indian politicians. Government officials suggested that such videos could harm electoral integrity in India.[185] In another situation, the High Court of Tripura, the highest state level constitutional court, stayed the application of the Rule for identifying the originator behind the fake message related to the resignation of the Chief Minister of the state of Tripura.[186]

[179] Intermediary Guidelines, Rule 4(2) (third proviso).

[180] Aditi Agarwal, *Traceability and End-to-End Encryption Cannot Co-exist on Digital Messaging Platforms: Experts*, Forbes India (2021), https://www.forbesindia.com/article/take-one-big-story-of-the-day/traceability-and-endtoend-encryption-cannot-coexist-on-digital-messaging-platforms-experts/66969/1 (last visited Jan. 10, 2023).

[181] Vrinda Bhandari, Rishab Bailey & Faiza Rahman, *Backdoors to Encryption: Analysing an Intermediary's Duty to Provide "Technical Assistance,"* SSRN Journal (2021), https://www.ssrn.com/abstract=3805980 (last visited May 1, 2023).

[182] Aditi Agarwal, *supra* note 164; Aditi Agarwal, *Can Traceability and End-to-End Encryption Co-exist? Here's the Legal View*, Forbes India (2021), https://www.forbesindia.com/article/take-one-big-story-of-the-day/can-traceability-and-endtoend-encryption-coexist-heres-the-legal-view/67001/1 (last visited May 1, 2023).

[183] Gurshabad Grover, Tanaya Rajwade & Divyank Katira, *The Ministry and the Trace: Subverting End-To-End Encryption*, 14 NUJS Law Review (2021); Greg Nojeim & Namrata Maheshwari, *Encryption in India: Preserving the Online Engine of Privacy, Free Expression, Security, and Economic Growth*, 17 Indian Journal of Law and Technology 1 (2021).

[184] Facebook Inc. v. Union of India WP (C) 7281 of 2021 (High Court of Delhi); WhatsApp LLC v. Union of India WP (C) 7284 of 2021 (High Court of Delhi).

[185] Aarathi Ganesan, *Govt May Send WhatsApp Traceability Notice over Deepfakes of Politicians Circulating: Report* (Oct. 16, 2023). https://www.medianama.com/2023/10/223-govt-whatsapp-traceability-notice-deepfakes-politicians/ (last visited Feb. 18, 2024).

[186] *Id.*