

REIMAGINING REGULATION OF SPEECH ON SOCIAL MEDIA PLATFORMS IN INDIA

Merrin Muhammed Ashraf

LLM in Law and Development, Azim Premji University, Bengaluru

Abstract

The driving objective of India's current regulatory response to speech on social media platforms ('SMPs') is to curb harmful and unlawful speech online. But its ability to achieve this is seriously undermined due to a lack adequate understanding of the free speech problems posed by the evolving and complex nature and functions of these platforms. Further, an exclusive focus on curbing harmful and unlawful speech has led to the framing of the regulatory question as one concerned with striking a balance between curbing undesirable speech and protecting freedom of speech. Such a framing gives enormous power to the state and the platform to decide what legitimate speech is and what is not. This thesis posits that there is a need to re-orient the goal of India's regulatory approach to speech on SMPs from one that is solely focused on fixing liability for harmful content to one that seeks to make SMPs a healthy and vibrant public sphere. When this becomes the regulatory goal, the real balancing act to be done is between empowering the state and platforms to regulate speech on the one hand, and placing restrictions on their power, on the other.

Key words: social media platforms, intermediary liability, freedom of speech, digital public sphere, content moderation

I. Introducing the Free Speech Problem in the Digital Public Sphere in India

Freedom of speech is of significant value for a democracy, which is a form of government that depends on information and communication.¹ Today, social media platforms ('SMPs') such as Twitter and Facebook have become an important space for people to express and exchange their ideas, views, and opinion. They also serve as crucial sites where public discourse starts and takes shape, where citizens, activists, and journalists publicly call into question state actions and failures, demand accountability, and organize protests and movements.² But SMPs also have the potential to serve as ideal places for the proliferation of harmful and unlawful content, and this is what has attracted most of the regulatory attention to the operation of these

¹ Benjamin R Barber, *Three Scenarios for the Future of Technology and Strong Democracy*, 113(4) Political Science Quarterly 573, 582 (1998-1999).

² Rutvi Zamre, *From streets to tweets: Surveying the impact of online activism*, ORF (Sept. 8, 2020), <https://www.orfonline.org/expert-speak/streets-tweets-surveying-impact-online-activism/>.

platforms. Countries across the globe, including India, have come up with legislations to check such content online.³ Platforms themselves take active measures to remove offensive, vile and illegal content to ‘present their best face’⁴ to users, advertisers and to the public at large.⁵ But, some of the events and developments that happened in India in the year 2021 put into doubt the real interest of the government and the platforms to regulate speech on SMPs.

In February 2021, a tussle broke out between the Government of India and Twitter when the latter refused to fully comply with the former’s order to take down thousands of individual accounts that posted content criticizing the government’s response to the farmers’ protests which were ongoing then.⁶ Twitter took a firm stand by refusing to take down these accounts which were alleged by the government to spread misinformation. It issued a statement terming the Government order as inconsistent with the Indian law and against its principle of ‘defending protected speech and freedom of expression’.⁷ However, it was soon coerced, under the threat of prison terms for its employees, into fully complying with the order.⁸

Unfortunately, Twitter’s initial response is only a rare instance of resistance by SMPs to a doubtful government request for content takedown. In April 2021, SMPs like Facebook and Twitter unhesitatingly complied with the government’s request to remove tweets that criticized the Indian government’s handling of the Covid pandemic.⁹ Likewise, there have been numerous other instances when SMPs have been complicit in the State’s attempts to censor speech online, the legitimacy of which is contested.¹⁰

Later in October 2021, leaked documents obtained by the Associated Press revealed Facebook’s biased approach towards the removal of hate speech, misinformation, and fake news, especially speech with anti-Muslim content, and that its algorithms and default settings

³ *Welcome to World Intermediary Liability Map*, WILMAP, <https://wilmap.stanford.edu/>, (last visited Apr 22, 2022).

⁴ TARLETON GILLESPIE, *CUSTODIANS OF THE INTERNET: PLATFORMS, CONTENT MODERATION, AND THE HIDDEN DECISIONS THAT SHAPE SOCIAL MEDIA*, 5 (Yale University Press, 2018).

⁵ *Id.*

⁶ Shoaib Daniyal, *Why is the Indian Government at War with Twitter*, Scroll.in (Jul. 8, 2021), <https://scroll.in/article/999171/why-is-the-government-of-india-at-war-with-twitter> .

⁷ *Updates on our response to blocking orders from the Indian Government*, Twitter Safety (Feb. 10, 2021) https://blog.twitter.com/en_in/topics/company/2020/twitters-response-indian-government .

⁸ Daniyal, *supra* note 6.

⁹ *Twitter removes over 50 posts on govt. orders; majority related to COVID-19*, THE HINDU (Apr 25, 2021), <https://www.thehindu.com/news/national/twitter-removes-over-50-posts-on-govt-orders-majority-related-to-covid-19/article34404805.ece>.

¹⁰ Paroma Soni, *Online censorship is growing in Modi’s India*, (Dec 14, 2021), <https://www.cjr.org/investigation/modi-censorship-india-twitter.php> .

played a role in amplifying such content.¹¹ This report revealed the interest or willingness of the platform to sometimes keep up and even spur harmful content.

These events and developments present us with a regulatory quagmire. On the one hand, the regulatory steps taken by the state and the platform are not succeeding in curbing harmful and unlawful content on SMPs, as is evident from the increased reports of online gender-based violence,¹² child pornography,¹³ hate-speech,¹⁴ disinformation, fake news,¹⁵ etc. On the other hand, the events discussed above indicate that they also have an interest in using their regulatory capacity to restrict legitimate forms of expression, and sometimes even to promote harmful speech. This anomalous situation indicates signs of failure of India's regulatory response to speech on SMPs. The central concern of this paper is to identify the reason for such failure, which could then enable meaningful legislative and policy deliberations to address the free speech problems raised by SMPs. This inquiry becomes significant in the backdrop of India being categorized as a 'partly free country' by the Freedom on Net ranking 2022, which termed the legal obligation of SMPs to remove content as per the direction of the government as problematic.¹⁶ This indicates that India's regulatory efforts to curb harmful speech online are in reality making SMPs less free for the users to express themselves, share information, and critique and hold power to account. While laws like sedition¹⁷ and UAPA¹⁸ present coercive and violent means to muzzle critical voices, cyber law, being the free speech law today,¹⁹

¹¹ Associated Press, *Facebook dithered in curbing divisive user content in India*, THE HINDU (Oct 24, 2021), <https://www.thehindu.com/news/national/facebook-dithered-in-curbing-divisive-user-content-in-india/article37149506.ece> .

¹² Anita Gurumurthy & Amshuman Dasarathy, *Profitable Provocations: A Study of Abuse and Misogynistic Trolling on Twitter Directed at Indian Women in Public-political Life*, IT for Change (2020) <https://itforchange.net/sites/default/files/2132/ITfC-Twitter-Report-Profitable-Provocations.pdf> (last visited Oct. 30, 2022).

¹³ *Child pornography: Maliwal summons Twitter India Policy head, Delhi Police*, Economic Times (Sept. 20, 2022), <https://economictimes.indiatimes.com/news/india/child-pornography-maliwal-summons-twitter-india-policy-head-delhi-police/articleshow/94324203.cms?from=mdr> .

¹⁴ Shweta Senger, *Hate Speech On Facebook Up 38%, Violent Content To 86% On Instagram In April, Says Meta*, India Times (Jun. 2, 2022), <https://www.indiatimes.com/news/india/hate-speech-on-facebook-up-38-violent-content-to-86-on-instagram-in-april-says-meta-571189.html> .

¹⁵ Chandrima Banerjee, *How social media gets away with spreading lies*, Times of India (May 5, 2022), <https://timesofindia.indiatimes.com/india/how-social-media-is-getting-away-with-disinformation/articleshow/91318044.cms> .

¹⁶ *Freedom on the Net 2021: The Global Drive to Control Big Tech*, Freedom House, <https://freedomhouse.org/report/freedom-net/2021/global-drive-control-big-tech> (last visited Jun 24, 2022).

¹⁷ The Indian Penal Code, No. 45 Acts of Parliament, 1860, S. 124A.

¹⁸ The Unlawful Activities (Prevention) Act, No. 37 Acts of Parliament, 1967.

¹⁹ Anupam Chander and Uyn P. Li, *Free Speech*, 100 IOWA Law Review 501, 501 (2015).

presents a subtle but equally nefarious means of wiping out speech that discomfits the state.²⁰ Therefore, we must uncover the regime of censorship that is normalized through the regulation of SMPs in the name of safety and security, but which at the same time fails to ensure online safety for the vulnerable and marginalized.²¹

In thinking about the reasons for the failure of India's regulatory response to speech on SMPs, I found Martin Moore's article on '*Tech Giants and Civic Functions*'²² helpful. According to him, the solutions to the problems posed by technology giants are destined to fail if we do not adequately define the problem that we are trying to address and if we do not have a vision of where we want our digital society to end up. Therefore, a useful starting point to identify the reasons for the failure of India's regulatory response is to inquire how it addresses these two considerations by examining the objective and rationale behind regulatory efforts.

In order to identify and critique the objective and rationale behind India's regulatory efforts toward speech on SMPs, I relied on primary as well as secondary resources. Primary resources include the Information Technology Act, Information Technology (Procedure and Safeguards for Blocking for Access of Information by Public) Rules, 2009, Information Technology (Intermediary Liability Guidelines) Rules, 2011, 2021 Rules, and Supreme Court case laws. Secondary resources include journal articles, journalistic pieces, news reports, prior empirical studies, and analysis of the transparency reports published by SMPs in India. Further, to gain an internal perspective of how India's intermediary liability law impacts the platforms' operations and how they view their role in mediating the speech of their users, I conducted interviews with public policy representatives of two significant SMPs having a presence in India. I also interviewed three legal researchers working on this subject to gauge their perspective on SMP regulation and its impact on the right to freedom of speech of users. The identity of all the interviewees is anonymized on request. The public policy representatives of the two SMPs will be referred to as SMP-1 and SMP-2. Researcher 1 is a policy researcher

²⁰ See, Ronald Deibert and Rafal Rohozinski, *Beyond Denial: Introducing Next-Generation Information Access Controls*, in *ACCESS CONTROLLED: THE SHAPING OF POWER, RIGHTS, AND RULE IN CYBERSPACE* (Ronald Deibert (eds), MIT Press, 2010).

²¹ Damni Kain et al, *Online Caste-Hate Speech: Pervasive Discrimination and Humiliation on Social Media* Centre for Internet and Society (2021), https://cis-india.org/internet-governance/blog/online_caste-hate_speech.pdf (last visited Oct. 30, 2022); Anita Gurumurthy, *Born digital, Born free? A socio-legal study on young women's experiences of online violence in South India*, IT for Change (2019), https://itforchange.net/sites/default/files/1618/Born-Digital_Born-Free_SynthesisReport.pdf (last visited Oct. 30, 2022); Quratulain Rehbar & Pallavi Pundir, *Muslim Women Were 'Auctioned Like Cattle' on a Hate Site*, Vice (Jul. 13, 2022), <https://www.vice.com/en/article/88ndgx/muslim-women-islamophobia-hate-app-india>.

²² Martin Moore, *Tech Giants and Civic Power*, Centre for the Study of Media, Communication and Power, 23 (April 2016), <https://www.kcl.ac.uk/policy-institute/assets/cmcp/tech-giants-and-civic-power.pdf>.

with a leading civil society organization and is working in the areas of social media regulation, intermediary liability, and digital expression. Researcher 2 is a policy researcher working on digital rights generally. Researcher 3 is a full-time counsel at a leading policy research organization working on digital rights in India.

This paper proceeds as follows:

In order to set a context and background for the subsequent appraisal of the Indian state's existing regulatory response to speech on SMPs and its impact on freedom of speech, the Second Chapter is devoted to explaining the different functions performed by SMPs that have implications for the realization of free speech values in a democracy. The Chapter also examines the unique challenges posed to the different dimensions of freedom of speech by SMPs and the speech environment that they create. I use the term SMPs to refer to platforms like Facebook and Twitter which, for their profit-making enterprise, allow their users to communicate with each other and also to create, upload, share and disseminate information like text, pictures, videos, and audio in a *shared public space*. This includes video-sharing platforms like YouTube, but excludes email service, closed messaging platforms like WhatsApp, and business and employment-oriented platforms. I derived this definition from a combined reading of the definition of 'social media intermediary' under Indian law,²³ and that of 'social networks' under Germany's Network Enforcement Act, or NetzDG law.²⁴

The Third Chapter will critically examine the regulatory response to speech on SMPs in India which is mainly embodied in Sections 69A and 79 of the Information Technology Act ('IT Act') and their allied Rules. These provisions aim to co-opt SMPs in the state's effort to regulate online speech by requiring them to observe certain due-diligence requirements concerning their content moderation practices, failing which they will be held liable for the user-generated content. This regulatory response is commonly referred to as the intermediary liability regime, which is one of the dominant modes of fixing liability on internet intermediaries like SMPs that is followed in many jurisdictions. The most recent regulatory development is the enactment of Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021 ('2021 Rules')²⁵ on 25th February, 2021 and its subsequent

²³ The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, G.S.R. 139(E), 2(1) (w).

²⁴ Netzwerkdurchsetzungsgesetz (NetzDG Act), 1, para. 1, sentence 1.

²⁵ The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, G.S.R. 139(E)

amendment in 2022 which revamped the intermediary liability law substantially and have raised concerns from various quarters about its potential to undermine freedom of speech.²⁶ This Chapter aims to question and re-examine the objectives and rationale behind the intermediary liability regime as it stands today.

In the Fourth Chapter, I will argue for the need to re-imagine India's regulatory approach to speech on SMPs, in light of the shortcomings of the current approach and the threat it poses to freedom of speech. This Chapter aims to provide some directions for future reform that is aimed at stemming the censorial excess of the state on SMPs and limiting the harms of private content moderation to ensure that these platforms become healthy and vibrant digital public spheres.

The final Chapter will conclude the discussion and summarize the findings and suggestions.

II. Role of Social Media Platforms in a Democracy and Implications for Freedom of Speech

Freedom of speech is recognized as a fundamental civil liberty by international human rights instruments²⁷ and many national constitutions.²⁸ Among the various justifications for valuing free speech, ranging from self-fulfillment to pluralism, its importance for advancing democracy runs as a common thread and is accepted as a stand-alone justification.²⁹ Scholars like Alexander Meiklejohn, Robert Post, and Jürgen Habermas have written at length about the value of free speech in a democracy to ensure that citizens have access to information and channels for communication to participate in democratic self-governance.³⁰ Jack Balkin, writing in the age of social media, goes beyond the democratic self-governance point, and argues that the purpose of free speech in a democracy is to 'promote a democratic culture'³¹

²⁶ *IT Rules 2021 Add To Fears Over Online Speech, Privacy; Critics Believe It May Lead To 'Outright Censorship'*, Firstpost (Jul. 16, 2021), <https://www.firstpost.com/tech/news-analysis/it-rules-2021-add-to-fears-over-online-speech-privacy-critics-believe-it-may-lead-to-outright-censorship-9810571.html> ; Apar Gupta, *The IT Rules fail the test of constitutionality*, The Indian Express (Oct. 1, 2021), <https://indianexpress.com/article/opinion/columns/the-it-rules-fail-the-the-test-of-constitutionality-7542621/> ; Mandates of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, OL IND 8/2021 (June 11, 2021), <https://drive.google.com/file/d/1c42o2qg97427sywdMZFR9v0Tt0eQzVB3/view> .

²⁷ International Covenant on Civil and Political Rights, art 19, Dec. 16, 1966, 999 U.N.T.S. 171; UN General Assembly, G.A. Res. 217A (III), Universal Declaration of Human Rights, art 19, Dec. 10, 1948.

²⁸ GAUTAM BHATIA, OFFEND, SHOCK, OR DISTURB: FREE SPEECH UNDER THE INDIAN CONSTITUTION, 4(Oxford University Press, 2018).

²⁹ ERIC BARENDT, FREEDOM OF SPEECH, 6-23 (1985).

³⁰ ALEXANDER MEIKLEJOHN, FREE SPEECH AND ITS RELATION TO SELF-GOVERNMENT (1948); Robert Post, *Participatory Democracy and Free Speech*, 97(3) Virginia Law Review 477 (2011); Jürgen Habermas, *Discourse Ethics: Notes on a Program of Philosophical Justification* in MORAL CONSCIOUSNESS AND COMMUNICATIVE ACTION, 43–116. (Cambridge, Massachusetts: MIT Press, 1990).

³¹ Jack Balkin, *Digital Speech and Democratic Culture: A Theory of Freedom of Expression for the Internet Society*, 79 N.Y.U. L. REV. 1, 31 (2004).

i.e. ‘a culture in which individuals have a fair opportunity to participate in the forms of meaning making that constitutes them as individuals’.³² In the democratic culture argument, it is not only speech and information relating to politics, public issues, or matters of governing importance that are valued, but every possible subject and modes of expression that ‘reflects popular taste, popular culture, and popular enthusiasm’.³³

In the light of these scholarly arguments on the importance of free speech for democracy, it is no wonder that the internet, which revolutionized the conditions for the exercise of free speech by decentralizing the communication networks, has been touted as a democratizing force.³⁴ One of the greatest manifestations of this democratizing potential of the internet has been said to be the advent of SMPs that enable individuals to overcome the traditional barriers to access speech avenues, and make public discourse and interaction possible at an unprecedented scale.³⁵ However, they also present new and subtle ways of limiting and controlling the speech on their platforms. The affordances of SMPs, and the offline-online continuum of societal structures and prejudices exacerbate existing marginalities and discrimination by providing a conducive environment for abuse, harassment and hate speech, on the grounds of gender, race, caste, religion, etc.³⁶ The existence of such online violence limits the access to and participation in the digital public sphere constituted by SMPs for the affected individuals and groups.³⁷

Part 1 of this chapter situates SMPs in the context of democracy and highlights the important functions that it performs which can promote or impede democracy, depending on how they are performed. Part 2 of this chapter identifies the different dimensions of the right to freedom of speech that the Indian Constitution guarantees and examines whether and how these dimensions are affected when speech happens on SMPs.

II.a Social Media Platforms as Digital Public Sphere

Jack Balkin in one of his recent writings, equates SMPs to a public sphere crucial for a democracy.³⁸ He defines the public sphere as a ‘space in which people express opinions and

³² Id.

³³ Id, at 31.

³⁴ EMILY B LAIDLAW, *REGULATING SPEECH IN CYBERSPACE, GATEKEEPERS, HUMAN RIGHTS AND CORPORATE RESPONSIBILITY* 17 (Cambridge, 2015).

³⁵ Id, at 71.

³⁶ Maya Mirchandani, *Digital hatred, real violence: Majoritarian radicalisation and social media in India*, ORF (2018), https://www.orfonline.org/wp-content/uploads/2018/08/ORF_OccasionalPaper_167_DigitalHatred.pdf, (last visited Oct. 30, 2022).

³⁷ Gurumurthy, *supra* note 21, at 16.

³⁸ Jack Balkin, *How to Regulate (And Not Regulate) Social Media*, 1 *Journal of Free Speech Law* 71 (2021).

exchange views that judge what is going on in society.’³⁹ He identifies three important public functions that SMPs already perform that can significantly influence the realization of free speech values in a democracy.

First, SMPs *facilitate public participation* in arts, culture, and politics.⁴⁰ Unlike the public sphere created by traditional media, people are not mere receivers of content on SMPs. SMPs are primarily driven by user-generated content that they host. Therefore, the public actively takes part in conversations on arts, culture, and politics on SMPs, as much as they remain spectators to such conversations.

Second, SMPs *organize public conversation*.⁴¹ SMPs not only allow people to express themselves and share information but also enable them to find an audience for their speech by connecting them to others. This is important as speech is of little value unless it reaches the intended audience.

Third, social media *curate public opinion*.⁴² SMPs decide which content to remain on their platform and which not, down-rank some, amplify certain other content, and personalize the content that a user receives. These activities determine ‘the speed of propagation and the reach of content’,⁴³ which in turn influences the formation of public opinion.⁴⁴

Balkin argues that if these three functions are performed well, it will help to realize the free speech values in a digital public sphere, viz. ‘*political democracy, cultural democracy, and growth and spread of knowledge*’.⁴⁵ Political democracy entails an informed public and their participation in the formation of public opinion on issues of public concern.⁴⁶ Cultural democracy entails free and widespread participation of the public in the ‘construction and development of the cultures that constitute their identities and affect their lives’.⁴⁷ Growth and spread of knowledge is concerned with the development and spread of knowledge through participation and conversation involving ‘a host of disciplines, institutions, and public-

³⁹ Id, at 2.

⁴⁰ Id, at 5.

⁴¹ Id.

⁴² Id.

⁴³ Id.

⁴⁴ Id.

⁴⁵ Id, at 7.

⁴⁶ Id.

⁴⁷ Id.

regarding professions'.⁴⁸ Balkin argues that if these three values are realized, then the digital public sphere constituted by SMPs is a '*healthy and vibrant public sphere*'.⁴⁹

One may agree or disagree with Balkin on what constitutes a 'well-functioning digital public sphere'⁵⁰ and whether it would help to realize political democracy, cultural democracy, and the growth and spread of knowledge. Nevertheless, the value of his works lies in identifying the different functions that SMPs perform today in the digital public sphere, which have implications for the freedom of speech of users, and consequently for the functioning of a democracy.

II.b Dimensions of freedom of speech on social media platforms

Article 19(1)(a) of the Constitution of India guarantees freedom of speech and expression to all its citizens.⁵¹ At the same time, Article 19(2) empowers the state to reasonably restrict our speech on the grounds of 'sovereignty and integrity of India, security of the State, friendly relations with foreign States, public order, decency or morality, contempt of Court, defamation or incitement to an offence'.⁵² In *Shreya Singhal v. Union of India*,⁵³ the Court extended the guarantee of freedom of speech of citizens to the online space, subject to these reasonable restrictions.

Judicial pronouncements on Article 19(1)(a) demonstrate that freedom of speech and expression goes beyond the mere *right to speak*.⁵⁴ In *Sakal Papers (P) Ltd v. Union of India*,⁵⁵ it was held that freedom of circulation is an essential part of freedom of publication and that without *the right to circulate*, the right to free speech and expression would have little value.⁵⁶ Further, the Supreme Court in *Secretary, Ministry of Information and Broadcasting, Government of India & Ors. v. Cricket Association of Bengal & Anr*,⁵⁷ held that the *right to receive information* is also an integral part of freedom of speech under Article 19(1)(a).⁵⁸ These three dimensions of freedom of speech are recognized by international human rights treaties as

⁴⁸ Id.

⁴⁹ Id.

⁵⁰ Id, at 14.

⁵¹ INDIA CONST. art 19, cl 1(a).

⁵² INDIA CONST. art 19, cl 2.

⁵³ (2013) 12 SCC 73.

⁵⁴ MADHAVI GORADIA DIVAN, FACETS OF MEDIA LAW, 7 (2nd ed. Eastern Book Company, 2018).

⁵⁵ AIR 1962 SC 305.

⁵⁶ Id, 22-25.

⁵⁷ (1995) 2 SCC 161.

⁵⁸ Id, 181.

well.⁵⁹ Further, all theories of democratic justification of free speech, whether it is Meikeljohn's and Post's democratic self-governance or Habermas' deliberative democracy or Balkin's democratic culture, also consider freedom of speech as encompassing all these three dimensions.

In the following sections, I seek to show that while SMPs enhance the opportunities for speech, the three important functions they perform, as identified by Balkin – 'facilitating public participation, organizing public conversation, and curating public opinion' - have the potential to place serious restraints on the different dimensions of freedom of speech

(i) Right to speak or express

SMPs significantly facilitate the exercise of the right to speak, as anyone with an internet connection can express their ideas and views in the form of text, videos, images, GIFs, and even in the form of likes and dislikes. Though speakers on SMPs are free from traditional gatekeepers like newspaper editors, their speech has to conform to the platform's community guidelines, which are privately framed rules that the users have to agree to before signing up on these platforms.⁶⁰ These rules guide users on what speech is permissible and not permissible on their platform. If the speech is found in violation of community guidelines, platforms remove them.

Apart from the reactive removal of content that violates their community guidelines, platforms also proactively monitor content using technological tools and human reviewers whose decisions do not always go according to the published guidelines.⁶¹ Further, the right to speak of the users is affected by platform design, like the 280-word character limit of Twitter.⁶² Thus today, SMPs significantly influence not only what people are allowed to speak on their platform, but also the form and substance that their speech takes.⁶³

This kind of speech regulation by platforms is today essential 'to protect one user from another, or one group from its antagonists, and to remove the offensive, vile, or illegal content from the

⁵⁹ International Covenant on Civil and Political Rights, art 19, Dec. 16, 1966, 999 U.N.T.S. 171; UN General Assembly, G.A. Res. 217A (III), Universal Declaration of Human Rights, art 19, Dec. 10, 1948.

⁶⁰ Gillespie, *supra* note 4, 45.

⁶¹ Catherine Buni and Soraya Chemali, *The Secret Rules of the Internet*, THE VERGE, <https://www.theverge.com/2016/4/13/11387934/internet-moderator-history-youtube-facebook-reddit-censorship-free-speech>.

⁶² Sarah Myers West, *Raging Against the Machine: Network Gatekeeping and Collective Action on Social Media Platforms*, 5(3) Media and Communication 28, 28 (2017).

⁶³ See, REBECCA MACKINNON, *CONSENT OF THE NETWORKED: THE WORLDWIDE STRUGGLE FOR INTERNET FREEDOM* (New York, Basic Books, 2012).

digital public sphere'.⁶⁴ However, the lack of regional and cultural representation in the framing of community guidelines and the content moderation systems, and the partisan enforcement of platform rules, has many a time led to the removal of legitimate, and often dissident speech and forms of expression, particularly by users in the marginalized communities.⁶⁵ Further, the failure of platforms to effectively remove harmful and abusive content, and their complicity in amplifying them due to their fixation on sensationalism and virality, forces the individual and groups who are targets of online abuse to go into silence or even withdraw themselves from the platforms.⁶⁶

(ii) Right to disseminate or circulate one's view

Though SMPs have made the avenues of speech more accessible, this accessibility does not necessarily mean that one's speech gets enough audience to effectively participate in the 'marketplace of ideas'⁶⁷ due to the content curation function performed by SMPs. SMPs down-rank or suppress undesirable content without removing them, and also amplify certain other content using sophisticated algorithms.⁶⁸ Such decisions are often influenced by the platforms' economic incentive to make available content that captures the attention of the users, irrespective of the public value of such content. As a result, the reach of their speech is not dependent on how reasoned their opinion is, but on whether the speech has enough economic value for the platforms for its algorithms to recommend it to a wider audience.⁶⁹

(iii) Right to receive information

Compared to traditional media, SMPs greatly enhance the ability of individuals to access diverse information and views. However, the realization of this potential is significantly affected by the organization of public conversation functions performed by SMPs. This function is necessitated by the abundance of speech, which resulted from the lowering of access to speech avenues, which is beyond the individual capacity of a person to consume.⁷⁰ This

⁶⁴ Gillespie, *supra* note 4, 5.

⁶⁵ Ángel Díaz and Laura Hecht-Felella, *Double Standards in Social Media Content Moderation*, Brennan Centre for Justice (2021), <https://www.brennancenter.org/our-work/research-reports/double-standards-social-media-content-moderation> (last visited Oct. 30, 2022).

⁶⁶ Damni Kain et al, *supra* note 21, 17; Gurumurthy, *supra* note 21,16.

⁶⁷ Secretary, Ministry of Information and Broadcasting, Government of India & Ors. v. Cricket Association of Bengal & Anr, (1995) 2 SCC 161.

⁶⁸ Kai Reimer and Sandra Peter, *Algorithmic audiencing: Why We Need to Rethink Free Speech on Social Media*, 00(0) *Journal of Information Technology* 1 (2021).

⁶⁹ See, *Id.* See also, AMBER SINHA, *THE NETWORKED PUBLIC, HOW SOCIAL MEDIA IS CHANGING DEMOCRACY*, 174 (Rupa Publications, 2019).

⁷⁰ See, Tim Wu, *Is the First Amendment Obsolete*, 117 *MICH. L. REV.* 547 (2018).

necessitates some kind of filtering so that an individual can access what is most relevant for her. While manual filtering by the individual is available, the time and effort tax required for this induces most of them to submit to the platform's filtering mechanism.⁷¹ This mechanism is driven by algorithms that track and anticipates users' preferences, personality traits, habits, lifestyle choices, movements, and identity categorizations 'in order to algorithmically intervene in users' daily experience' by curating the digital content.⁷² Hence, each user experiences the public sphere in a personalized way. This limits the exposure diversity of individuals to only content that conforms to their existing worldview, prejudices and preferences.⁷³ If this results from self-personalization, it could be argued to be an exercise of individual autonomy. But algorithmic personalization can seriously restrict the user's right to access information and diverse views.⁷⁴ Further, as pointed out earlier, down-ranking of content by platforms, thereby limiting its reach, also affect the ability of a user to access that content amidst the vast sea of content.

The above discussion reveals that if SMPs have dismantled the power-holdings of traditional gatekeepers over individuals' speech, they have become the new gatekeepers. As Prof Jeffrey Rosen presciently observed many years ago, SMPs 'have more power in determining who can speak and who can be heard around the globe than any Supreme Court justice, any king or any president'.⁷⁵

In light of the understanding gained in this Chapter about the unique challenges posed by SMPs for the realization of different dimensions of free speech, the next Chapter traces the trajectory of India's regulatory response to speech on SMPs with a view to locate and critique its objective and rationale.

III. India's Tryst with Speech on Social Media Platforms: A Critical Review of its Regulatory Trajectory

At the peak of the Covid-19 pandemic in India, major SMPs, particularly Twitter, were used by people to seek and mobilize help for those who desperately needed oxygen cylinders,

⁷¹ CASS R. SUNSTEIN, CHOOSING NOT TO CHOOSE: UNDERSTANDING THE VALUE OF CHOICE 34, 35 (2015).

⁷² TANYA KANT, ALGORITHMIC PERSONALIZATION, IDENTITY AND EVERYDAY LIFE, 10 (Oxford University Press, 2020).

⁷³ CASS R. SUNSTEIN, #REPUBLIC: DIVIDED DEMOCRACY IN THE AGE OF SOCIAL MEDIA (2017).

⁷⁴ See, Kant, supra note 72, 10.

⁷⁵ Miguel Helft, *Facebook Wrestles with Free Speech and Civility*, New York Times, (Dec 12, 2010), <https://www.npr.org/transcripts/142714568> .

hospital beds, and other essential medical supplies to save the lives of their loved ones.⁷⁶ Apart from crowdsourcing help, SMPs also became an important site for the public to call out the government for its utter failure and mismanagement in its handling of the pandemic. In the face of government manipulation of vital Covid statistics and under-reporting of death, SMPs became a vital communication channel for many organizations to provide information to the public and to demand transparency and accountability from the state authorities.⁷⁷

But soon, amidst the devastating second wave of the pandemic, reports emerged about a major clampdown by SMPs targeting speech critical of the government, with Twitter taking down at least 52 such tweets of elected officials, journalists, and editors.⁷⁸ Twitter responded by saying that the tweets were removed in compliance with government demand and *in accordance with local laws*.⁷⁹ What are these laws? Why were these laws not invoked when SMPs failed to remove and even actively promoted divisive content and fake news, which led to riots and lynching.⁸⁰ If SMPs are a significant avenue for speech, ‘by giving “voice to the voiceless” and a means to escape state censorship’,⁸¹ what is the underlying rationale behind the laws that seek to regulate speech on SMPs?

This Chapter traces the trajectory of India’s regulatory response to speech on SMPs and seeks to show that its dominant objective is to curb harmful and unlawful speech. However, the regulatory response is focused on only one aspect of the functioning of SMPs i.e. facilitating public participation and overlooks the other two functions and the complex architecture of these platforms. This not only limits its ability to effectively realize its objective of curbing harmful and unlawful content but also fails to address the threats to the freedom of speech of the users from the functioning of SMPs.

⁷⁶ Francesca Recchia and Suchitra Vijayan, *In India, social media is a lifeline. It’s being silenced*, Washington Post (May 2, 2021), <https://www.washingtonpost.com/opinions/2021/05/06/india-social-media-covid-19/> .

⁷⁷ *Id.*

⁷⁸ Arup Deep, *Twitter Takes Down Tweets From MP, MLA, Editor Criticising Handling Of Pandemic Upon Government Request*, Medianama (Apr 24, 2021), <https://www.medianama.com/2021/04/223-twitter-mp-minister-censor/> .

⁷⁹ Yashraj Sharma, *Twitter Groups Offer India a Covid-19 Lifeline*, Wired (May 6, 2021), <https://www.wired.com/story/twitter-groups-offer-india-a-covid-19-lifeline/> .

⁸⁰ The Associated Press, *Facebook dithered in curbing divisive user content in India*, NPR (Oct 23, 2021), <https://www.npr.org/2021/10/23/1048746697/facebook-misinformation-india> .

⁸¹ *Ajit Mohan and Ors v. Legislative Assembly, National Capital Territory of Delhi and Ors*, AIR 2021 SC 3346, ¶ 2.

III.a Neutral Conduits to Censorship Proxies

Indian law subsumes SMPs under the broad category of internet intermediaries for the purpose of regulation. Section 2 (w) of the IT Act defines internet intermediaries to include ‘telecom service providers, network service providers, internet service providers, web-hosting service providers, search engines, online payment sites, online-auction sites, online-market places, and even cyber cafes’.⁸² A defining characteristic that binds all these entities is that their primary function is limited to providing *access* to a communication system over which information made available by third parties is transmitted or temporarily stored or hosted.⁸³ From this, it can be inferred that Indian law perceived SMPs as mere platforms that host user-generated content with little or no control over the content.⁸⁴

As mentioned in the introductory Chapter, the intermediary liability regime of India is majorly embodied in Sections 79 and 69A of the IT Act and their allied Rules. A brief description and analysis of these provisions follow:

Section 79 of IT Act and the Intermediary Liability Rules, 2011

Section 79 of the IT Act as it currently stands evolved in response to the Bazee.com case.⁸⁵ Obscene MMS video clips involving two minors were put up on sale on the website Bazee.com, which led to criminal prosecution of its Managing Director. The Delhi High court refused to quash the criminal proceedings against him on the ground that the website failed to employ filters to screen obscene content and to have a policy in place to prevent the listing, display, and sale of such content.⁸⁶ This incident brought to light the hardship that will be imposed on the website owners if they are held strictly liable for the third-party content on their platform over which they have little control. Subsequently, S. 79 was amended to accord intermediaries with conditional immunity from liability.⁸⁷ This means that internet intermediaries like SMPs were immunized from liability for any content hosted on their platform so long as they observed some due diligence requirements, and ‘did not initiate the transmission of the content, select the receiver of transmission or modify the content of the transmission’.⁸⁸ Other conditions for availing the protection includes removing the impugned content from their platform upon

⁸² Information Technology Act, No. 21 Acts of Parliament, 2000, S. 2(w).

⁸³ Information Technology Act, No. 21 Acts of Parliament, 2000, S. 79(2)(a).

⁸⁴ Parliamentary Standing Committee, *Report on the Information Technology (Amendment) Act, 2006* (2007).

⁸⁵ Avnish Bajaj v. State, 150 (2008) DLT 76.

⁸⁶ Id.

⁸⁷ Pritika Rai Advani, *Intermediary Liability in India*, EPW (Dec 14, 2013), <https://www.jstor.org/stable/24479053>.

⁸⁸ Information Technology Act, No. 21 Acts of Parliament, 2000, S. 79(2)(b).

receipt of knowledge of the illegality of the content and ‘not abetting, conspiring, aiding or inducing the commission of the unlawful act’.⁸⁹

While examining the amendment of Section 79, the Department of Information Technology, Government of India was quoted in the Parliamentary Standing Committee as saying:

‘...any of the service providers may not be knowing exactly what their subscribers are doing. For what they are not knowing, they should not be penalized. This is the provision being followed worldwide.’⁹⁰

The circumstances that led to the amendment of Section 79 involving the Bazeem.com fiasco, and the above observation in the Standing Committee indicate that the primary motivation behind the amendment of Section 79 was to avoid another instance of an internet intermediary being held liable for an illegal content on their platform, provided they observe some due diligence.

What are these due diligence requirements? This was not immediately clear after the amendment and it was only in 2011 that the government came up with the Information Technology (Intermediary Liability Guidelines) Rules, 2011 (‘2011 Rules’)⁹¹ which specified the due diligence requirements to be followed by the intermediaries. It included requiring intermediaries to put up their terms and conditions on their website which should inform the users not to publish certain types of content.⁹² This includes content that is grossly harmful, harassing, blasphemous, obscene, invasive of another's privacy, hateful, infringes on intellectual property rights, etc.⁹³ The Rules also require the platforms to remove or disable access to an unlawful content within 36 hours upon receiving the knowledge of the same, either by itself or upon being notified by an affected party or by the court or by the government.⁹⁴

Unlike S. 230 of the Communication Decency Act of the United States which allows the intermediaries to moderate content themselves without losing immunity from liability,⁹⁵ S. 79 and the Rules, until recently, did not allow the platforms to moderate content on their own. Their responsiveness to unlawful and harmful content was envisaged to be only in the form of fair warnings to users, and prompt removal upon receiving notice of such content on their

⁸⁹ Information Technology Act, No. 21 Acts of Parliament, 2000, S. 79(3)

⁹⁰ Parliamentary Standing Committee, *Report on the Information Technology (Amendment) Act, 2006* (2007).

⁹¹ Information Technology (Intermediary Guidelines) Rules, 2011, G.S.R. 314(E).

⁹² Id., 3(1).

⁹³ Id.

⁹⁴ Id., 3(4).

⁹⁵ Communications Decency Act, 47 U.S. Code 230.

platform. This was in consonance with the Indian law's imagination of SMPs as neutral or dumb conduits of speech that cannot exercise any sort of control over the content.

Pursuant to the enactment of these Rules, SMPs were barraged with takedown requests from both the government and individual users. An empirical study revealed that SMPs tended to err on the side of caution and take down content even when the requests were frivolous and faulty, fearing the loss of immunity under Section 79.⁹⁶ Further, the government repeatedly used this provision to stifle speech that was critical of them. The Google Transparency Reports for January –June, 2011 revealed that out of the 358 items that the government requested it to remove, 255 items were categorized as 'government criticism'.⁹⁷

S. 69A and the Blocking Rules

Section 69A empowers the Central Government to issue directions to intermediaries to block information from public access on the grounds of 'sovereignty and integrity of India, defence of India, security of the State, friendly relations with foreign States or public order or for preventing incitement to the commission of certain cognizable offence'.⁹⁸ The section has to be read along with the Information Technology (Procedure and Safeguards for Blocking for Access of Information by Public) Rules, 2009 ('Blocking Rules 2009').⁹⁹ The Rules empowers certain officers of the Central Government to receive complaints and to decide whether a particular information on the internet is prejudicial to security of state, public order, etc. and if so found, to issue blocking directions.¹⁰⁰ If the information to be blocked is hosted on the platform of an intermediary, the blocking order is sent to the intermediary for compliance.¹⁰¹ Before a blocking decision is passed, there is no mandatory requirement to give an opportunity of hearing to the originator of the content or to the intermediary.¹⁰² Moreover, requests for blocking and decisions thereon are to be kept confidential.¹⁰³ If the intermediary fails to comply with the blocking order, appropriate action may be taken against them.¹⁰⁴ Applied to SMPs,

⁹⁶ Rishabh Dara, *Intermediary Liability in India: Chilling Effects on Free Expression on the Internet*, Centre for Internet and Society, (Apr. 27, 2012), <https://cis-india.org/internet-governance/chilling-effects-on-free-expression-on-internet>.

⁹⁷ *Google Transparency Report India*, January –June 2011, <http://www.google.com/transparencyreport/governmentrequests/IN/?p=2on-o6> (last visited Jun 24, 2022).

⁹⁸ Information Technology Act, No. 21 Acts of Parliament, 2000, S. 69A.

⁹⁹ Information Technology (Procedure and Safeguards for Blocking for Access of Information by Public) Rules, 2009, G.S.R. 781(E).

¹⁰⁰ See, Id, 3-8.

¹⁰¹ Id, 8(6).

¹⁰² See, Id, 3-8

¹⁰³ Id, 16.

¹⁰⁴ Id, 12.

Section 69A has the same impact as that of Section 79 as it demands the platforms to takedown content on the threat of legal liability, that too on same grounds save a few provisions. SMPs have reported that though they receive content takedown requests from the government both under Sections 69A and 79, the majority of the requests are under the former.¹⁰⁵

The free speech concerns posed by the blocking power of the state under Section 69A become stark with the widespread resort to this section in recent times. In response to a question in the Parliament, the Minister of State of Electronics and IT, Rajiv Chandrasekhar, said that the government ordered the blocking of 9849 social media accounts/URLs in the year 2020 under Section 69A, which is 171% more than the preceding year.¹⁰⁶ Further, it has been revealed from the Lumen database, which is an American collaborative archive that ‘collects and analyzes legal complaints and requests for removal of online materials’,¹⁰⁷ that during the period between 2017 and 2020, Twitter received a blocking order under Section 69A every once in three months.¹⁰⁸ Similarly, Section 69A was used to take down posts that were critical of the government’s actions in Kashmir, its handling of the Covid-19 pandemic and farmers’ protests, and cow vigilantism in India.¹⁰⁹

Even though Section 69A provides for confidentiality of blocking orders, Twitter used to disclose the letters it received from the government containing blocking directions.¹¹⁰ But this seemed to have come to an end later as the database stopped reflecting the complete list of tweets and hashtags that the government wanted platforms to remove.¹¹¹ By late 2020, the database also stopped reflecting the specific sections under which content was removed.¹¹² This opacity around government directions for blocking and takedown of user content and lack of any recourse against decision of the government leaves little opportunity for the users to challenge restrictions on their speech online. The platforms are also not in a position under the law to contest or refuse dubious requests for taking down content, whether from the government or private individuals, due to the threat of legal liability and prison terms for its

¹⁰⁵ Torsha Sarkar & Gurshabad Grover, *Content Takedown and Users Rights*, Centre for Internet and Society (Feb 14, 2020), <https://cis-india.org/internet-governance/blog/content-takedown-and-users-rights-1> .

¹⁰⁶ *Govt blocked 9,849 social media URLs in 2020*, THE ECONOMIC TIMES, (Dec 08, 2021), <https://economictimes.indiatimes.com/tech/technology/govt-blocked-9849-social-media-urls-in-2020/articleshow/88170865.cms?from=mdr>.

¹⁰⁷ Lumen Database, <https://www.lumendatabase.org/> .

¹⁰⁸ Anuj Srivas, *Understanding the Nuances to Twitter’s Standoff With the Modi Government*, The Wire (Feb 12, 2021), <https://thewire.in/tech/twitter-modi-government-block-section-69-a> .

¹⁰⁹ Id.

¹¹⁰ Id.

¹¹¹ Id.

¹¹² Id.

employees.¹¹³ The public policy team of SMP-1 and SMP-2 confirmed in their interview that the platforms will not, nor are they permitted by the law to turn down or contest government requests for takedown or blocking of user-content. This enables the state to use platforms as proxies for censorship,¹¹⁴ and it leaves the freedom of speech of users in a very precarious position. As one of my interviewees, Researcher -1, observed, since India's intermediary liability regime was framed from the perspective of intermediaries, it did not take into account the implications of such law for the free speech of the users.

Shreya Singhal judgment

The constitutionality of Sections 79 and 69A and their allied Rules were challenged in *Shreya Singhal v. Union of India*¹¹⁵ before the Supreme Court. Section 79 was challenged on the ground that it allowed private entities like intermediaries to exercise their judgment about the unlawfulness of a speech.¹¹⁶ The court agreed with the petitioner that internet intermediaries, who are private actors, cannot be permitted to decide on their own whether an information should be taken down or not, especially without any in-built safeguards.¹¹⁷ But instead of striking down the section as unconstitutional, the Court read down the phrase 'actual knowledge' in the section and the 2011 Rules to mean that intermediaries are obliged to take down content only upon receiving knowledge via Court order *or* government notification.¹¹⁸ But the content takedown order or request has to conform to the grounds mentioned in Article 19(2).¹¹⁹ With this ruling, intermediaries are no longer liable to act upon content takedown requests by private individuals. Though this helped to mitigate the platform's power to adjudicate on the legality of speech, what it legitimized is equally a pernicious threat to free speech - i.e. empowering the government to demand platforms to take down content without adequate safeguards against misuse.

A similar endorsement of unchecked censorship power of the state can be seen in the Court's ruling on the constitutionality of Section 69A. Section 69A was challenged on the ground that

¹¹³ Daniyal, *supra* note 6; Soumyarendra Barik, *Facebook may Proactively Censor Content to protect Itself from Legal or Regulatory Actions*, Medianama (Sept 2, 2020), <https://www.medianama.com/2020/09/223-facebook-censor-content-avoid-legal-regulatory-challenge/>.

¹¹⁴ Seth F Kreimer, *Censorship by Proxy: the First Amendment, Internet Intermediaries, and the Problem of the Weakest Link* (University of Pennsylvania, Faculty Scholarship Paper 127, 2006), http://scholarship.law.upenn.edu/faculty_scholarship/127.

¹¹⁵ (2013) 12 S.C.C. 73.

¹¹⁶ *Id.*, 114.

¹¹⁷ *Id.*, 116-117.

¹¹⁸ *Id.*, 117.

¹¹⁹ *Id.*, 117.

it contained no procedural safeguards against the state's power to block access to information online.¹²⁰ The lack of opportunity for pre-decisional hearing and the confidentiality provision adversely affects the fundamental right to freedom of speech of the citizens. But none of these concerns found favor with the court and it upheld the constitutionality of Section 69A in its entirety.¹²¹

It has been critiqued that Sections 79 and 69A have been designed for controlling, rather than promoting free and open discussion in the digital public sphere.¹²² Therefore, scholars have voiced the concern that the Supreme Court, by refusing to strike down these sections, facilitated the bringing into existence of a permanent state of exception in the digital public sphere where the state has unchecked power to determine what speech is permissible and what not.¹²³ This gravely frustrates public discourse and democratic outcomes.

b. Intermediary Liability Rules, 2021: Private actors back in the censorship game!

The Supreme Court in *Shreya Singhal* case was unequivocal on one point - private actors like SMPs cannot be adjudicators of the legality of speech.¹²⁴ The Court's reason for this was that it would be impractical to ask the platforms to judge the legality of each piece of content, given the sheer scale of online content.¹²⁵ However, later it became increasingly evident that platforms are not incapable of actively reviewing the content on their platform, and that they in fact engage in self-censorship of speech on their platform for their own economic interests.¹²⁶ Further, the magnitude and virality of harmful speech on SMPs in the form of child pornography, rape videos, fake news, and misinformation became so huge, sometimes even resulting in offline harm.¹²⁷ It was recognized that waiting for government notification or court order to remove such speech from the platforms is unwise. As a result, for quite some time now, there has been push from both the government as well as the judiciary to impose greater

¹²⁰ Id, 108.

¹²¹ Id, 111.

¹²² Smarika Kumar, *Governing Speech on the Internet: From the Free Marketplace Policy to a Controlled 'Public Sphere'*, Centre for Internet and Society, (Aug 28, 2015), https://cis-india.org/raw/blog_governing-speech-on-the-internet.

¹²³ Id.

¹²⁴ *Shreya Singhal v. Union of India*, (2013) 12 S.C.C. 73. 117.

¹²⁵ Id, 117.

¹²⁶ See, Rishab Bailey et al, *Internet intermediaries and online harms: Regulatory Responses in India* 44 (Data Governance Network, Working Paper No. 06, 2020), <https://datagovernance.org/report/internet-intermediaries-and-online-harms-regulatory-responses-in-india>.

¹²⁷ Id.

obligations on the internet intermediaries to curb harmful content online, including deploying of automatic filtering tools to proactively monitor and remove such content.¹²⁸

This drive towards greater regulation of internet intermediaries like SMPs was finally given legal articulation by enacting the 2021 Rules, particularly Part II of the Rules. It replaced the 2011 Rules. As per government's counter-affidavit filed before the Madras High Court where the constitutionality of the 2021 Rules was challenged, they have been enacted in recognition of the significant growth of online SMPs and their influencing capabilities, and to deal with content such as child pornography, rape and gang-rape imageries, and other harmful messages that may incite violence or commission of crime or is prejudicial to the sovereignty and integrity of India, etc.¹²⁹

In a departure from the 2011 Rules, the 2021 Rules acknowledges the capability of SMPs to self-moderate content on their platform and allows them to do so without losing their immunity from liability.¹³⁰ Therefore, apart from requiring the platforms to comply with takedown orders from court and the government, it allows the platform to, voluntarily or upon user request, takedown content which they decide to be not conforming to their community standards.¹³¹ The Rules also impose some additional obligations on SMPs to improve their responsiveness to unlawful content on their platform, such as requiring them to remove artificially morphed image or content that exposes the private area of an individual or depicts full or partial nudity or shows such individual engaged in a sexual act within 24 hours upon receipt of a complaint from an individual.¹³²

Further, certain category of social media intermediaries, categorized as significant social media intermediaries (SSMIs) due to their large number of users, are encouraged to deploy automatic filtering tools to proactively identify content that depicts 'rape, child sexual abuse, and content identical to ones that have been removed previously on the grounds of sovereignty and integrity of India; security of the State; friendly relations with foreign States; public order; decency or morality; in relation to contempt of court; defamation; incitement to an offence relating to the

¹²⁸ Id.

¹²⁹ T.M. Krishna v. Union of India, W.P. Nos. 12515 of 2021, Madras High Court, Counter-affidavit filed by the Ministry of Electronics and Information Technology, 4-6 (Aug 25, 2021), https://drive.google.com/file/d/18NQGJ-9CIt8A2_xkM72ow0ytxcRgAn_1/view .

¹³⁰ The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, G.S.R. 139(E), 3(1)(b).

¹³¹ Id.

¹³² The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, G.S.R. 139(E), 3(2)(b).

above, or any information which is prohibited under any law'.¹³³ The Rules acknowledge the limitations of technological tools to accurately distinguish between problematic and unproblematic content, and therefore, require intermediaries using automated tools to have periodic review of such tools, and appropriate human oversight.¹³⁴ However, no indication is given on the factors that need to be looked out for in such a review or the requirement for any external audit.

These increased obligations that are imposed on intermediaries, in a way, legitimizes the enormous power that they already have over the speech of their users. Therefore, it is welcome to note that the Rules contain some safeguards to limit the unrestrained exercise of this power, especially by the SSIMs. SSIMs are now required to publish periodic monthly reports containing details of complaints received, action taken thereon, and also the number and details of communication links or parts of information that have been removed in pursuance of proactive monitoring of content.¹³⁵ This is a significant measure to ensure transparency and accountability in the functioning of the platform so that public and policy makers get more insights into their content moderation practices, which is very opaque at present. SSIMs in India like Google (including YouTube), Koo, Meta (Facebook and Instagram), ShareChat, Snap, Twitter have started publishing monthly compliance reports.¹³⁶ However, Internet Freedom Foundation, which studies and deconstructs these monthly reports for the public, has observed that there is no uniformity in the format adopted by each SSIM for reporting, and there is lack of granularity in disclosing details of the content taken down.¹³⁷ As a result, it is doubtful how much transparency actually resulted from the periodic reports published so far as per the requirements of the law.

Apart from transparency requirements, the Rules also impose important procedural limitations on the moderation power of the platform, like sending notice to the person whose content is sought to be removed and by giving her an opportunity to contest the decision.¹³⁸ While this is

¹³³ The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, G.S.R. 139(E), 4(4).

¹³⁴ Id, proviso 2 to 4(4).

¹³⁵ Id, 4(1)(d).

¹³⁶ #SocialMediaComplianceWatch: analysis of Social Media Compliance Reports for the month of March 2022, INTERNET FREEDOM FOUNDATION, (May 27, 2022), <https://internetfreedom.in/compliance-report-march-2022/>.

¹³⁷ Id.

¹³⁸ The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, G.S.R. 139(E), 4(8).

a significant improvement to limit the platform power, it is not clear what recourse a user will have if the platform fails to comply with these procedural rules.

The most concerning aspect of the 2021 Rules is that it continues on the path of its predecessor Rules by not incorporating any safeguards against the unlimited powers of the state to require the platforms to remove any content within 36 hours.¹³⁹ If a content is sought to be removed by the government, the creator of the content neither gets a notice of the same nor the detailed grounds under which her content is removed.¹⁴⁰ She is also deprived of any recourse under the Rules to contest such government-directed take-down of content. This goes against the ruling of the Supreme Court in *Anuradha Bhasin and Another v. Union of India*,¹⁴¹ that any order that affects the liberty of people must be notified directly and reliably.¹⁴² It is also important to note that the transparency reports to be published by SSIMs are not required to disclose the details of takedown requests received from the government, the grounds under which such requests were issued, and the actions taken by the platforms on them.¹⁴³

Further, in a very recent development, the Ministry of Electronics and Information Technology notified a set of amendments to the 2021 IT Rules.¹⁴⁴ One of the important amendment relates to constituting a Grievance Appellate Committee for the users to appeal against the grievance redressal process of the SMPs.¹⁴⁵ This is a positive move to limit the power of SMPs by providing users with recourse against wrongful removal of legitimate content or non-removal of harmful content by these platforms. However, the composition of the Appellate Committee, which is proposed to consist of only Union Government appointees,¹⁴⁶ has raised concerns that it is yet another attempt to extend the government's centralized censorship powers over the digital public sphere.¹⁴⁷ The Internet Freedom Foundation writes that the provision for constitution of GAC 'will make bureaucrats arbiters of our online free speech'.¹⁴⁸ When this is

¹³⁹ Id, 3(1)(d).

¹⁴⁰ Id.

¹⁴¹ AIR 2020 SC 1308.

¹⁴² Id, 96.

¹⁴³ The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, G.S.R. 139(E), 4(1)(d).

¹⁴⁴ The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2022, G.S.R. 794(E).

¹⁴⁵ Id, 4.

¹⁴⁶ Id.

¹⁴⁷ Arathi Ganeshan, *How Will The Proposed Amendments To The IT Rules Affect Free Speech And Intermediaries?*, Medianama (Jun 22, 2022), <https://www.medianama.com/2022/06/223-it-rules-amendments-india-free-speech-big-tech/>.

¹⁴⁸ *IFF's Statement on the notified IT Rules Amendment, 2022*, Internet Freedom Foundation (Oct. 28, 2022), <https://internetfreedom.in/iff-statement-on-the-notified-it-rules-amendment-2022/>.

the case, mere inclusion of a clause in the amendment requiring the internet intermediaries to respect the rights of citizens under the Constitution,¹⁴⁹ seems perfunctory and an eye-wash.

III. Critique

If the initial objective behind amending Section 79 of the IT Act in the aftermath of the Bazeed.com case was to protect the internet intermediaries from liability for user-generated content on their platform, the introduction of Section 69A and the Blocking Rules, and subsequent enactment of the 2011 Rules is motivated by the objective to curb unlawful and harmful content online. In the later years, this objective gained prominence with the increase in the magnitude and virality of offensive speech on SMPs. This is also the government's stated objective behind enacting the latest 2021 Rules.

The intermediary liability regime of India follows the gatekeeper liability model. Situated broadly within regulatory studies, the concept of gatekeepers has come to be defined as 'non-state actors with the capacity to alter the behavior of others in circumstances where the state has limited capacity to do'.¹⁵⁰ Gatekeeper liability emerges when the capacity of the state to regulate a specific conduct, as in the case of speech on SMPs is limited, but the capacity of a third party gatekeeper might be better.¹⁵¹ In such a situation, the gatekeeper will be liable for failing to prevent the misconduct. One of the major criticisms against the gatekeeper liability model of regulating online speech is that it 'sees the gatekeeper as somehow uninvolved' and as an entity that will not benefit from misconduct.¹⁵²

Today, it is almost nonsensical to say that SMPs are uninvolved actors in the online speech environment. As Balkin identified, these platforms not only facilitate users' speech but also organize and curate their speech and conversations, thereby exercising significant influence over the information flow and dissemination in the society.¹⁵³ In the performance of these functions, their economic interest, and profit motive dominate over the norms of free speech values.¹⁵⁴ In such a scenario, the gatekeeping function that they perform itself can result in

¹⁴⁹ The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2022, G.S.R. 794(E), 3.

¹⁵⁰ Laidlaw, *supra* note 34, at 39. Also see, Jonathan Zittrain, *A History of Online Gatekeeping*, 19(2) *Harvard Journal of Law and Technology* 253 (2006).

¹⁵¹ *Id.*

¹⁵² Emily Laidlaw, *A framework for identifying Internet information gatekeepers* 24(3) *International Review of Law, Computers & Technology* 267 (2010).

¹⁵³ Balkin, *supra* note 38, at 5.

¹⁵⁴ Chinmayi Arun, *Making choices: Social Media Platforms and Freedom of Expression Norms, in REGARDLESS OF FRONTIERS? FREEDOM OF EXPRESSION AND INFORMATION IN THE 21ST CENTURY*, 3 (Lee Bollinger & Agnes Callamard eds., Columbia University Press, 2021).

harm like amplification of fake news, disinformation, and other divisive content, rather than preventing harmful speech. For instance, one of the key takeaways from the revelations of the Facebook whistleblower, Frances Haugen, was that the platform's prioritization of astronomical profit over people has, in many instances, made it abandon or delay initiatives to reduce misinformation and radicalization.¹⁵⁵ When the gatekeepers themselves are involved actors, it is necessary that the law provides safeguards against the harms resulting from their functioning.

But, the intermediary liability regime only requires the SMPs to prohibit its users from engaging in certain kinds of speech and to take steps to remove them. It regulates neither the organization of public conversation function nor the content curation function performed by the platforms, which are mostly driven by their profit motive and business interest. This is not to say that the content removal function performed by platforms is not driven by such interests. But by ignoring the other two functions performed by SMPs, India's intermediary liability regime is structurally weak to address the proliferation and amplification of harmful and unlawful speech resulting from platform operations and, therefore may fail in achieving its objective.

Another drawback of the traditional gatekeeper liability model is that it is concerned only with preventing gated misconduct and ignores gated rights completely.¹⁵⁶ This means that since the primary objective of India's regulatory response to speech on SMPs is to curb harmful and unlawful content on SMPs, its intermediary liability regime considers the online community on these platforms as sources of misconduct and ignore the fact that they have rights, which cannot be taken away without fair and reasonable procedural safeguards. Since the intermediary liability regime makes the platform liable for any failure to remove unlawful content, as observed earlier, they are incentivized to err on the side of caution and comply with even doubtful content takedown requests, whether from the users or from the government, thereby infringing on a user's right to freedom of speech. While the 2021 Rules provide a notice and appeal provision before the SMPs takedown content voluntarily or upon user requests, no such procedural safeguards are provided when content is taken down upon government requests. This enables the state to use the platforms as their proxy for censorship and bypass the constitutional restraints on their power to restrict the freedom of speech of the users.¹⁵⁷ The

¹⁵⁵ Christiano Lima, *A Whistleblowers Power: Key Takeaways from the Facebook Papers*, Washington Post (Oct 26, 2021)<https://www.washingtonpost.com/technology/2021/10/25/what-are-the-facebook-papers/> .

¹⁵⁶ Laidlaw, *supra* note 152.

¹⁵⁷ Kreimer, *supra* note 114.

recent instances of major SMPs taking down content that was critical of the government actions in relation to farmers protest, Covid pandemic and Kashmir situation clearly demonstrate the threat to freedom of speech and democratic discourse posed by the intermediary liability regime.¹⁵⁸ Furthermore, the Rules also fail to address the impact of the platform functions like the organization of public conversation and content curation on dimensions of freedom of speech other than the right to speak i.e. to the right to circulate or rights to receive information as indicated in Chapter 2.

To sum up, a critical examination of India's intermediary liability regime reveals that the law neither addresses the concerns of scholars like Lessig and Reidenberg that the power of technology may turn out to be a tool of tyranny and control,¹⁵⁹ nor addresses the concerns of critical theorists of cyberspace that the state will make use of technologies to evade the constitutional restraints on the exercise of legal power.¹⁶⁰ The casualty of this failure is the diminution of freedom of speech online and opportunity for participation in democratic self-governance and 'production and distribution of culture'.¹⁶¹

IV. Re-imagining India's regulatory Response to Speech on Social Media Platforms

Returning to the two considerations put forward by Martin Moore – do we have a vision of where we want our SMPs to end up, and have we adequately defined the problem that we are trying to address?¹⁶² The discussion in the previous Chapter intended to indicate that India's regulatory response to speech on SMPs has been framed without a full comprehension of the complex architecture and functioning of SMPs. As a result, it overlooks the different ways in which the functioning of SMPs impacts the different dimensions of freedom of speech, particularly the right to disseminate or circulate one's speech, and the right to receive information.

As regards a vision for SMPs, it is to be noted that India's regulatory efforts are primarily driven by concerns about their growing influence and capabilities to be ideal places for

¹⁵⁸ Karnika Kohli and Anuj Srivas, *Modi Sarkar Doesn't Want You to Read Lutyens' Gossip, Tweets on #KashmirUnrest*, THE WIRE, (Sept 6, 2017), <https://thewire.in/politics/censorship-twitter-blocked-narendra-modi-freedom-of-expression>.

¹⁵⁹ LAWRENCE LESSIG, CODE Version 2.0, 4 (2006); Joel R. Reidenberg, *Lex Informatica: The Formulation of Information Policy Rules through Technology*, 76 TEX. L. REV. 553 (1998).

¹⁶⁰ James Boyle, *Foucault in Cyberspace: Surveillance, Sovereignty and Hardwired Censors*, 66 University of Cincinnati Law Review 177 (1997).

¹⁶¹ See Balkin, supra note 31.

¹⁶² Moore, supra note 22.

harboring unlawful and harmful content.¹⁶³ It does not give enough consideration to the digital public sphere that SMPs constitute, and the public functions that they perform which has consequences for the free speech values and democracy. Therefore, I argue that instead of the narrow focus on curbing harmful or unlawful content, the objective of India's regulatory response should be to make SMPs a healthy and vibrant digital public sphere as put forward by Balkin, as discussed in the first chapter.¹⁶⁴

Balkin's vision of SMPs is an appropriate objective to pursue due to three reasons. First, such a vision acknowledges that SMPs are not mere neutral platforms that host user content, and that they perform important functions with respect to such content, like the organization of public conversation and curation of public opinion. Consequently, second, it takes into account not just the impact of the functions of SMPs on the right to speak of users but also the right to disseminate, and the right to access information. Third, it acknowledges the immense potential of SMPs to serve as sites of public discourse and for realizing 'political democracy, cultural democracy, and the growth and spread of knowledge'.¹⁶⁵

In order to effectively realize this shift in goal orientation, we need a principled approach rather than a reactive approach to the free speech issues online, given the evolving nature of SMPs and the complexities they may produce in the future. In the following paragraphs, I provide some broad directions that such an approach can take:

a. Change the incentives for SMPs

To recall, Balkin argues that the digital public sphere constituted by SMPs will be healthy and vibrant only if the three functions of SMPs – 'facilitating public participation, organizing public conversation, and curating public opinion' - are performed by them in such a way that it helps to realize the free speech values, viz. political democracy, cultural democracy, and growth and spread of knowledge.¹⁶⁶ True to the concerns of philosophers of technology like Lessig and Reidenberg,¹⁶⁷ we have had many instances that prove that left to themselves, SMPs may not be the paradise of free speech or create conditions for a vibrant digital public sphere. At the height of the Syrian war, many videos and posts that showed the world the horrors and

¹⁶³ T.M. Krishna v. Union of India, W.P. Nos. 12515 of 2021, Madras High Court, Counter-affidavit filed by the Ministry of Electronics and Information Technology, 4-6 (Aug 25, 2021), https://drive.google.com/file/d/18NQGJ-9CIt8A2_xkM72ow0ytxcRgAn_1/view .

¹⁶⁴ Balkin, *supra* note 38, at 7.

¹⁶⁵ *Id.*

¹⁶⁶ *Id.*

¹⁶⁷ *Supra* note 159.

brutalities faced by the people were removed by major SMPs in compliance with their community rules that prohibited gory images, despite the public value involved in the content.¹⁶⁸ If this was an instance of overzealous censorship, in Myanmar, Facebook was accused of inconsistent and weak enforcement of its rules in curbing divisive content, which is said to have played a major role in the ethnic cleansing of Rohingyas that took place there.¹⁶⁹

As Koops observed, technology and markets are not usually the ‘allies of weak parties’.¹⁷⁰ This is because the norms that guide the performance of SMPs’ functions are often influenced by economic and profit incentives of the platform and are not aligned to the public interest or free speech values.¹⁷¹ This was confirmed when in my interview with SMP-1, their public police representative made the statement that being private actors, they are not obliged to honor the free speech right of its users. This alludes to one of the major failings of the intermediary liability regime of India as it incentivizes SMPs only to save themselves from sanctions, even at the cost of the freedom of speech of their users. However, as SMPs have grown so big to be entwined with the institutions of public discourse, they ought to have some responsibility to the public that they crucially affect.¹⁷²

If the goal of regulatory efforts is to make SMPs a healthy and vibrant public sphere, there should be a fundamental reorientation of the incentives that the law creates for the SMPs. The regulatory efforts should aim at creating incentives for SMPs¹⁷³ to not infringe upon the free speech right of the users, and promote its platforms as a site for political democracy, cultural democracy, and the growth and spread of knowledge. In regulatory studies, this approach to regulation has been termed meta-regulation. It refers to the ways by which external regulators such as the state ‘deliberately – rather than unintentionally – seek to induce targets to develop their own internal self-regulatory responses to public problems’.¹⁷⁴

¹⁶⁸ DAVID KAYE, *SPEECH POLICE: THE GLOBAL STRUGGLE TO GOVERN THE INTERNET*, 17 (Columbia Global Reports, New York, 2019).

¹⁶⁹ Id. Also see, Paul Mozur, *A Genocide Incited on Facebook, With Posts From Myanmar’s Military*, Washington Post (Oct. 15, 2018), <https://www.nytimes.com/2018/10/15/technology/myanmar-facebook-genocide.html>.

¹⁷⁰ E.J Koops, *Law, technology, and shifting power relations*, 25(2) Berkeley Technology Law Journal 973, 1028 (2010).

¹⁷¹ Arun, supra note 154.

¹⁷² Gillespie, supra note 4, at 207-209.

¹⁷³ Martin Husovec, *Why There Is No Due Process Online?* Balkinization (Jun. 7, 2019), <https://balkin.blogspot.com/search?q=husovec>.

¹⁷⁴ Peter Grabosky, *Meta Regulation*, in *REGULATORY THEORY: FOUNDATIONS AND APPLICATIONS*, 153 (Peter Drahos (ed), 2017).

One way to reorient the incentives is to reconsider the current way of imposing penalties or initiating prosecution for each and every failure of SMPs to take down content. Manila Principles on Intermediary Liability recommend that intermediaries must be compelled to take down content only in compliance with a court order.¹⁷⁵ But time- and content-sensitivity in some cases may make it unwise to wait for a judicial order. In such cases, platforms may be required to act in response to government notification or user flagging, with the caveat that the failure to act likewise be penalized only in case of egregious or proven systematic non-compliance.¹⁷⁶ This would help to ensure that the platforms do not over comply with any and every request for content takedown, especially from the government.

We need to recognize that the technological affordances of SMPs are neither neutral nor inevitable, but are careful design choices. The law's role should be to incentivize the platforms to re-align those design choices to protect their users from abuse and harm and at the same time from violation of their right to free speech. In this regard, it is worthwhile to refer to the United Nations Guiding Principles on Business and Human Rights, laid down by the Office of the United Nations High Commissioner for Human Rights, which call upon States to protect their citizens from the human rights impacts of business operations.¹⁷⁷ These Principles also provide a set of norms and concrete actionable steps for the States and businesses to prevent, mitigate and remediate such impacts. Though it is a soft law and is non-binding, these principles can provide useful guidance on how the States, through effective policies, legislations, regulations, and adjudication, can cause/incentivize SMPs to respect the human rights of their users in their operations, and provide access to remedy for any rights violations.¹⁷⁸

b. Target all the gatekeeping functions of SMPs

By facilitating participation, organizing public conversation, and curating public opinion, SMPs have become one of the important gatekeepers of public discourse and the flow of information in society today. They perform the gatekeeping functions through their content moderation practices, which take various forms - removal of speech or the speaker, curation of content, and down-ranking or amplifying content.¹⁷⁹ All three forms can significantly impact

¹⁷⁵ *Manila Principles on Intermediary Liability*, Principle II, https://www.eff.org/files/2015/10/31/manila_principles_1.0.pdf (last visited Jun 11, 2022).

¹⁷⁶ See, Patrick Zurth, *The German NetzDG as Role Model or Cautionary Tale? Implications for the Debate on Social Media Liability*, 31 *Fordham Intell. Prop. Media & Ent. L.J.* 1084, 1118 (2021).

¹⁷⁷ Office of the High Commissioner for Human Rights. *Guiding Principles for Business and Human Rights*, Resolution 17/4 of 16 June 2011, https://www.ohchr.org/sites/default/files/Documents/Publications/GuidingPrinciplesBusinessHR_EN.pdf.

¹⁷⁸ *Id.*, Principles 1-10.

¹⁷⁹ See Reimer and Peter, *supra* note 68.

the different aspects of the freedom of speech of its users. As observed earlier, the regulatory approach so far primarily targets the first form of content moderation i.e. removal of speech or speaker from the platform for posting prohibited content. It is one way the state can induce SMPs to uphold what it considers the public values in the functioning of the platform. If this is done right with the right motivation, it can contribute to making the platforms a vibrant public sphere.

However, it is also important to regulate the other forms of content moderation like curation of content, and down-ranking or amplifying content, which are mostly algorithmically driven. These functions of SMPs influence the flow, speed, and reach of information. There is growing evidence that these are influenced by the economic and profit interests of the platforms which make them prioritize sensationalism over reasoned debate, and familiarity over diversity in recommending content for the users.¹⁸⁰ This can result in the amplification of harmful and divisive content that often translates into offline harm.¹⁸¹ It also negatively affects the freedom of speech, as the subversion of public discourse through the elimination of some voices and promoting certain other voices may result in one's speech not reaching enough audience, and may also impinge on the audience's right to receive diverse information and views.

Hence, any regulatory effort that does not take into account content moderation practices of SMPs other than the removal of content will be missing an important threat to the free speech of users today, and to a healthy and vibrant public sphere that craves free play of diverse and plural voices. Here, it is worthwhile to look at the proposed Digital Services Act of the European Union, which requires that the recommender systems employed by the SMPs be subjected to risk assessment, independent audits, and risk mitigations measures to stem the dissemination of illegal content, minimize adverse effects for the exercise and realization of the fundamental right to freedom of expression and information, and on civic discourse, electoral processes, etc.¹⁸²

Further, studies have found that one of the main ways by which a content goes viral is through the reshare/repost/retweet functionality,¹⁸³ which are affordances provided by SMPs. Research also suggests that 'retweet or reshare cascades can be predicted with some degree of

¹⁸⁰ CASS R. SUNSTEIN, *REPUBLIC.COM* (2011).

¹⁸¹ Arthur Bainbridge, *Facebook Fuelling Violence and Instability Across the Globe*, OWP (Nov 25, 2021), <https://theowp.org/reports/facebook-fuelling-violence-and-instability-across-the-globe/>.

¹⁸² Regulation of the European Parliament and of the Council on a Single Market for Digital Services (Digital Services Act) and amending Directive 2000/31/EC, art.26-29.

¹⁸³ Anita Gurumurthy and Amshuman Dasarathy, *supra* note 12, at 67.

accuracy'.¹⁸⁴ Therefore, the viral spread of offensive content can be prevented if SMPs invest in developing such capacity. Researchers also suggest other preemptory measures platforms can take such as slowing the pace of interactions on their platforms, and encouraging users to reflect on and self-moderate the content that they post.¹⁸⁵ These studies indicate that, apart from improving the capability of SMPs to identify and remove unlawful and harmful content, equal attention should be paid to improving the content curation and ranking function that they perform. And, the law can play an important role in pushing SMPs to improve their technological affordances, so as not to infringe on the rights and freedom of its users.

c. Govern the Speech Governors

'This is not a social media intermediary versus the government problem. This is a people versus social media intermediary and people versus government problem.'

-Tanmay Singh, Internet Freedom Foundation¹⁸⁶

When the objective of regulatory efforts is solely focused on curbing harmful and unlawful content online, the question of devising an appropriate regulatory response to speech on SMPs tends to be reduced to a matter of striking a balance between curbing harmful content online and protecting the freedom of expression. Our experience so far with the intermediary liability regime of India should tell us that this is a wrong framing to look at the regulatory question.

First, there is a blurring of lines between free speech and harmful speech in the digital context, as whether a speech falls into either of the categories depends a lot on the subjective satisfaction of the state and the platform. The difficulty in accurately identifying harmful speech was shared by the public policy team of SMP 2 when they observed that there is a need for 'bringing in standard and precise definitions of terms such as "free speech", "hate speech", etc.'

Second, today it is seen that platforms, incentivized by the laws governing them, prioritize their platform's safety over their users' freedom of expression. For instance, Gillespie observed that the community guidelines of Facebook now place more emphasis on ensuring a safe experience for its users than on giving full force to freedom of expression, which was its proclaimed aim

¹⁸⁴ Cheng, J., et al. *Can cascades be predicted?* WWW '14: Proceedings of the 23rd international conference on World Wide Web, pp. 925-936 (2014, April) dl.acm.org .

¹⁸⁵ Id. Also see, Martin Graff, *Twitter is taking a step towards 'slowcial media' by asking users to rethink abusive messages*, Scroll.in (May 22, 2021), <https://scroll.in/article/995457/twitter-is-making-a-step-towards-slowcial-media-by-asking-users-to-rethink-abusive-messages>.

¹⁸⁶ Ganeshan, supra note 147.

in the initial years.¹⁸⁷ My interviews with the Public Policy personnel of two major SMPs also revealed the same. The representative of SMP-1 remarked that the removal of legitimate speech is an acceptable casualty if it is for the attainment of a greater good. The Public Policy Team of SMP- 2, in their written response, stated that their content moderation practice is headed by a Trust and Safety team. They further observed that:

‘As intermediaries, social media platforms should be responsible towards their communities at large by being compliant with the laws of the land and by building transparency, trust, and safety that enable users to engage in wholesome discussions and leverage technology in a creative manner.’

This prioritization of safety over freedom of speech, especially in the absence of procedural safeguards, can be an easy way for the state and platforms to silence dissenting and minority voices online. If the aim of regulatory efforts is to create a healthy and vibrant digital public sphere, curbing harmful content and safeguarding freedom of speech should be considered not as contradictory goals but as complementary goals. The real balance to be struck is between enabling the state and platform to regulate speech on SMPs, and placing restrictions on their power. This will ensure that while states and platforms are able to take measures to restrict harmful and unlawful content, the exercise of this power is subjected to procedural safeguards that will provide room to dispute and question their judgment that infringes on the rights of the people.

Researcher -1 observed in her interview that neither the government nor the platforms come to the negotiating table with good intentions. As noted earlier, the regulatory efforts of platforms are driven mainly by their economic incentives rather than in the public interest or in the interest of human rights norms like free speech. The spotlight then naturally turns to the state, responsible for protecting individual rights and liberties. However, it is time that we seriously acknowledge the fact or the possibility that nation-states make use of the privately-owned infrastructure to ‘surveil, police, and control speakers’ indirectly.¹⁸⁸ Repressive legislations like sedition are now not the only means for the state to silence criticism and dissent when it can operate under the umbrage of platform power. For instance, the Indian government asked

¹⁸⁷ Tarleton Gillespie, *Facebook’s improved “Community Standards” still can’t resolve the central paradox*, Social Media Collective (Mar 18, 2015), <https://socialmediacollective.org/2015/03/18/facebooks-improved-community-standards-still-cant-resolve-the-central-paradox/>. Also see, Tim Wu, *Will Artificial Intelligence Eat the Law? The Rise of Hybrid Social-Ordering Systems*, 119 Columbia Law Review 2001, 2009 (2019).

¹⁸⁸ Kreimer, *supra* note 114.

Twitter to not allow any image-based searches for Rohingya Muslims, and to proactively block certain hashtags that emerged in response to the government's action in Kashmir.¹⁸⁹ It has been observed that in doing so, the intention of the government was not merely to curb the harmful speech and threats of terror, but also to impede the spread of news about the human rights violations occurring therein.¹⁹⁰ Thus, the state's quest for information control over the cyberspace is beyond denial today.¹⁹¹

Therefore, we are in a situation where the two main governors of speech - SMPs and state - cannot be trusted to act in public interest or to ensure 'a healthy and vibrant digital public sphere'. While SMPs empower the people to challenge the authority, it simultaneously increases their own power over the speech of the users.¹⁹² This power of SMPs over the content and flow of speech is more than what the States possess, and therefore, rather than limiting it, they use it to bolster their own power over the speech of its citizens. Therefore, the governors of speech need to be governed.

An important non-negotiable step towards holding them accountable is to ask for more and more transparency regarding their interests and actions in relation to speech on SMPs. While India has taken the right step by requiring the platforms to publish periodic transparency reports, more transparency must be demanded from them than what is currently demanded. There should be transparency regarding government takedown requests received by them and the action taken thereon, the content moderation practices of platforms, and the algorithms they employ for content curation.¹⁹³ The government must also be required to publish transparency reports that disclose the content takedown requests it sends and the grounds for such requests.¹⁹⁴ The information revealed from these transparency reports of the platform and the government will enable the citizens, civil society, and academics to call out the state and the platforms for their failure to act in accordance with the free speech values and to demand more accountability from them.

The above broad principles are the least that needs to be adopted to reorient the trajectory of India's intermediary liability regime and set it on the path of realizing 'a healthy and vibrant

¹⁸⁹ Kohli and Srivas, *supra* note 158.

¹⁹⁰ *Id.*

¹⁹¹ Deibert and Rohozinski, *supra* note 20, at 23.

¹⁹² Moore, *supra* note 22.

¹⁹³ *Santa Clara Principles 2.0 on Transparency and Accountability in Content Moderation*, Principles 3 and 5, <https://santaclaraprinciples.org/> (last visited Jun 11, 2022).

¹⁹⁴ *Manila Principles on Intermediary Liability*, Principle VI https://www.eff.org/files/2015/10/31/manila_principles_1.0.pdf (last visited Jun 11, 2022).

digital public sphere¹⁹⁵ on SMPs that preserves and promotes freedom of speech and democratic values. As Researcher-2 observed, there is a need for an in-built safeguard in the intermediary liability regime against censorship of the free speech of users of SMPs.

V. Conclusion

This paper intended to examine the reasons for the failure of India's regulatory response to speech on SMPs. First, I explained the different functions performed by SMPs as identified by Balkin and analyzed the impact of these functions on the three important dimensions of the freedom of speech of users. Thereafter, I traced the trajectory of India's regulatory response to speech on SMPs with a view to identifying and critiquing its objective and rationale. I identified that the primary objective of India's regulatory efforts is to curb harmful and unlawful content online. But since such efforts have been made without adequately comprehending the complex architecture and functions performed by the SMPs, and the economic interests that drive such functions, they are weakened in their ability to thwart the flow of harmful and unlawful content, especially that which results from the functions of the platforms themselves. The regulatory efforts also remain agnostic to the power and interests of the state and the platform to exercise control over the public discourse and flow of information in society. Drawing inspiration from Jack Balkin, I argued that the regulatory efforts should aim to make SMPs a healthy and vibrant public sphere. In order to realize this, it is necessary that the law incentivizes the platforms to perform their functions in alignment with the free speech values, regulates all forms of content moderation practices of SMPs, and incorporates safeguards against the misuse of the regulatory power of the state and the platform over the speech of the users.

¹⁹⁵ Balkin, *supra* note 38, at 7.