Conceptualising India's Safe Harbour in the era of Platform Governance

- *Vasudev Devadasan**

ABSTRACT: The push for greater regulation of online platforms has led to calls to re-evaluate the statutory immunities granted to online intermediaries for hosting unlawful third-party content (i.e., safe harbour). This paper argues that greater accountability for online platforms need not interfere with existing (and indeed strengthened) safe harbour protections. However, to achieve this outcome, legislators must recognise the difference in enforcement approaches between secondary liability and platform governance regimes. This paper argues the types of obligations that can be imposed as pre-conditions to safe harbour are different from those that can be imposed as direct statutory obligations. This is because secondary liability is enforced through individual suits against individual pieces of content while direct statutory obligations are continual and apply to all content on and all procedures of a platform. Relying on the principles of secondary liability, this paper outlines the types of obligations that can be meaningfully enforced (and those that cannot) using the tool of secondary liability. Using India's Intermediary Guidelines as a case-study, this paper highlights the importance of matching the type of obligation to the appropriate enforcement mechanism (e.g., transparency mandates cannot be imposed as pre-conditions to safe harbour) and the risks to both free speech and platform accountability of failing to do so. Recognising the difference between how secondary liability and direct statutory duties operate should lead to a reassessment of arguments that call for a dilution of safe harbour in the name of greater platform accountability. The paper concludes that intelligent statutory design can distinguish between the obligations that can be imposed as pre-conditions to safe harbour and those that ought to be direct statutory duties, allowing legislators to achieve greater transparency and accountability from platforms while retaining (and even strengthening) existing safe harbour protections.

I. INTRODUCTION

The regulation of online intermediaries is witnessing a shift towards the adoption of comprehensive statutory frameworks aimed at ensuring transparency, accountability, and responsibility in the operations of online intermediaries. This has been described as "platform governance"[1] and seeks to imbue governance relationships between internet users, platforms, and States. India is a part to this broader regulatory shift, with the Government's Digital India Bill reportedly set to regulate a wide array of platform behaviour ranging from content moderation, adjudicatory mechanisms, anti-competitive practices, cyber-crime, the use of personal data by platforms, and artificial intelligence.[2] As part of this regulatory overhaul, Indian lawmakers are also re-evaluating the contours of the statutory immunities currently provided to online intermediaries for transmitting or hosting unlawful content (i.e., safe harbour),[3] with the Minister of State for Information Technology posing the question, "Should there be a safe harbour at all for all intermediaries?"[4]

Safe harbours for online intermediaries have been essential in facilitating platform ecosystems populated by user-generated content, giving us the modern internet as we know it.[5] They also

* Vasudev Devadasan is a Consultant at the Centre for Research and Planning, Supreme Court of India. The views and opinions expressed in this article are solely those of the author and do not reflect the views of the Centre for Research and Planning. The author is grateful to Faiza Rahman, Namrata Maheshwari, and Sachin Dhawan for their insightful feedback and comments on drafts of this article.

[1] Robert Gorwa, 'What Is Platform Governance?' (2019) 22(6) Information, Communication & Society 854.

[2] Aarathi Ganesan, '"Should There Be Safe Harbour At All?": 30 Talking Points from the Digital India Act Consultation' (*MediaNama*, 10 March 2023) <https://www.medianama.com/2023/03/223-30-talking-points-digital-india-act-consultation/> accessed 4 April 2023.

[3] Soumyarendra Barik, 'Govt Rethinking "Safe Harbour" in Digital India Bill: How This Could Change Internet Landscape' *The Indian Express* (10 March 2023) <https://indianexpress.com/article/explained/explained-sci-tech/digital-india-bill-new-law-internet-explained-8488748/> accessed 4 April 2023; Ganesan (n 3).

[4] Editorial, 'Safe Harbour at Risk: The Hindu Editorial on the Impact of the Proposed Digital India Act, 2023' *The Hindu* (12 March 2023) <https://www.thehindu.com/opinion/editorial/safe-harbour-at-risk-the-hindu-editorial-on-the-impact-of-the-proposed-digital-india-act-2023/article66611045.ece> accessed 4 April 2023.

[5] Kate Klonick, 'The New Governors: The People, Rules, and Processes Governing Online Speech' (2018) 131 Harvard Law Review 1598, 1603–04.

constitute vital protections for free expression, protecting online content from collateral censorship by private and State actors, who seek to impose liability on the owners of digital infrastructure.[6] India's safe harbour, found in Section 79 of the Information Technology Act, 2000 ("**IT Act**"), and its interpretation in *Shreya Singhal vs. Union of India* has been heralded as offering comparatively robust protections to online intermediaries, and in turn safeguarding online free speech for citizens.[7] However, as the Minister's statement evidences, a wholesale revision of the IT Act at a time which calls for greater platform accountability could lead to an undermining of safe harbour protections.

This article argues that the greater regulation of online platforms need not interfere with existing (and indeed strengthened) safe harbour protections in the Indian context. However, this requires recognising that the enforcement mechanisms of the two regulatory approaches are fundamentally distinct and designing legislation accordingly. Intermediary liability is grounded in the principle of secondary liability based on the function or role of the intermediary vis-à-vis the allegedly unlawful content and is enforced *ex-post* through individual legal actions against content (i.e., lawsuits or prosecutions). In contrast, platform governance seeks to regulate the design, processes, and structures of platforms in the pursuit of greater transparency, accountability, and online safety, often through *ex-ante* measures. The enforcement of platform governance obligations does not rely on the hosting of unlawful content, but rather the violation of certain normative standards of platform design, processes, and implementation.

This article argues that intermediary liability should and can co-exist within a platform governance statute or a broader legislative mandate if lawmakers *appropriately distinguish* between the two regulatory approaches. Crucially, lawmakers should avoid relying solely on liability to achieve platform governance's goals. Simply put, an intermediary could be granted safe harbour in a lawsuit for content based on the function it performs vis-à-vis a specific piece of content, while still being held accountable for breaches of its statutory obligations that regulate its design, processes, or structures. Recognising this distinction should cause lawmakers to abandon their disproportionate focus on modifying safe harbour and address the

[6] Jack M. Balkin, 'Old-School/New-School Speech Regulation' (2014) 127 Harvard Law Review 2296, 2313.

[7] Kyung-Sin Park, 'From Liability Trap to the World's Safest Harbour: Lessons from China, India, Japan, South Korea, Indonesia, and Malaysia' in Giancarlo Frosio (ed), *Oxford Handbook of Online Intermediary Liability* (Oxford University Press 2020) 259. A key reason for this observation is that in India, intermediaries are not deemed to have actual knowledge of unlawful content (and hence not legally required to remove content) until served with a court or government order.

truly thorny issues of designing normative standards for platform design, processes, and administration which can be imposed as direct statutory obligations.

In addition to distinguishing between intermediary liability and platform governance approaches, conversations about safe harbour must also factor in the government's power to directly block content. It may be argued that safe harbour ought to be narrowed because platforms are unresponsive to government demands for removal of content. However, in the Indian context, Section 69A of the IT Act empowers the Indian Government to direct intermediaries to remove (or disable access to) content under threat of imprisonment, and safe harbour does not protect intermediaries from the consequences of non-compliance. Indeed, the Government regularly relies on Section 69A to block content.[8] While examining Section 69A is beyond the scope of this paper, it is important to recognise that strong safe harbour protections operate independently of the Government's power to block content, undercutting arguments that narrowing safe harbour is essential to make platforms responsive to the government.

Overview

Understanding the distinction in enforcement mechanisms between intermediary liability and platform governance first requires a proper appreciation of the nature of secondary liability, and the rise of secondary liability for intermediaries as regulatory tool. Therefore, the first half of this article is dedicated to articulating how safe harbour, properly conceived, should operate. The article thus begins (in **Section 2**) by examining the notion of secondary liability for online intermediaries, along with its implications for unlawful content and free expression. **Section 3** analyses Sections 2(1)(w) and 79 of the IT Act, examining how intermediary liability is grounded in the principle of secondary liability and enforced through individual lawsuits. It argues that questions of 'what an intermediary is?' and 'when is it liable?' can only be determined in a suit for unlawful content. (Note: the article refers to 'suits' as a heuristic for legal actions seeking to impose civil or criminal liability on intermediaries. In practice, an intermediary may be subject to civil suits or criminal prosecution.)

[8] Revathi Krishnan and Regina Mihindukulasuriya, 'Accounts of Prasar Bharati CEO, Caravan, Actor Sushant Singh among Those "withheld" by Twitter' (*The Print*, 1 February 2021) <https://theprint.in/india/accounts-of-prasar-bharati-ceo-caravan-actor-sushant-singh-among-those-withheld-by-twitter/596638/> accessed 3 March 2021; Aroon Deep, 'Twitter Takes Down Tweets from MP, MLA, Editor Criticising Pandemic Handling Upon Government Request' (*MediaNama*, 24 April 2021) <https://www.medianama.com/2021/04/223-twitter-mp-minister-censor/> accessed 4 November 2022.

The article then turns to India's Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021 ("**Intermediary Guidelines**"),[9] arguing that the Guidelines conflate the regulatory approaches of intermediary liability and platform governance and highlighting the dangers of this approach. **Sections 4 and 5** make the case for simplifying India's approach to safe harbour based on a principled approach to gatekeeping. **Section 6** sets out a tentative roadmap for how to separate the spheres of intermediary liability (and safe harbour) from platform governance from a regulatory design perspective and discusses the benefits of making this separation. **Section 7** concludes by acknowledging the lawmakers will continue to face difficult design questions in the creation of upcoming frameworks, but that intelligent regulatory design can ensure that stronger safe harbour and greater platform accountability are not mutually exclusive goals.

II. SECONDARY LIABILITY FOR INTERMEDIARIES

Intermediary liability as a tool for regulating the behaviour of intermediaries is fundamentally grounded in the notion of secondary liability. The common law doctrine of secondary liability recognises that where a party's harmful conduct is at least "partly conditioned" on the conduct of a second party, some legal responsibility for the harms attaches itself to this second party.[10] For example, if Party X knowingly assists Party Y in causing harms to Party Z, some responsibility for the harm attaches itself to Party X. In other words, Party X is *secondarily* liable for the harms caused to Party Z. Knowingly assisting in a wrongful act is one way to incur secondary liability, but it may also be incurred through other actions such as procurement, common design, and authorisation of a wrongful act.[11]

In the context of unlawful content, secondary liability could arise in a variety of situations. Some common situations include, a publisher's liability for its authors' defamatory writing, a bookstore's liability for a banned book, or in the digital age, an internet service provider or online intermediary's liability for transmitting or hosting unlawful texts, images, or videos. However, given the different role played by each of these secondary actors, the doctrine of secondary liability holds them differentially liable.[12] By determining when Party A is

[9] Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, G.S.R. 139(E) dated 25 February 2021.

[10] Jaani Riordan, 'Principles of Secondary Liability' in Jaani Riordan, *The Liability of Internet Intermediaries* (Oxford University Press 2016) 113.

[11] ibid 113–114.

[12] ibid 117.

secondarily liable for Party B's harmful conduct, the doctrine of secondary liability sets the outer limits of tortious responsibility that Party A bears for Party B's harmful actions.[13] Secondary liability thus determines *in what circumstances* 'a secondary wrongdoer may be made to answer for the wrongs of others where, by their own conduct, they have assumed some responsibility for the primary wrongdoing.'[14]

A. Secondary liability to curb unlawful content online

The advent of the internet, where intermediaries transmit and host both lawful and unlawful content, has given rise to a classical secondary liability question, but on an unprecedented scale.[15] Intermediaries may (intentionally or inadvertently) facilitate a wide range of unlawful conduct by internet users without ever themselves engaging in illegal acts.[16] For example, an intermediary may host defamatory content, facilitating the content's publication and dissemination without authoring, editing, or even knowing about the content. Intermediaries have thus been described as "necessary but insufficient causes" of unlawful activity online.[17] From the perspective of secondary liability, it was only a matter of time before litigants sought to hold intermediaries secondarily liable for their role in facilitating unlawful activity online.[18]

There is a parallel reason why online intermediaries are attractive targets for both litigants and governments. The decentralised, global, and often anonymous nature of activity on the internet means that primary wrongdoers are able to easily evade responsibility for unlawful acts.[19] Thus, the option of prosecuting or holding liable primary wrongdoers is often not open to litigants and governments, prompting them to turn to the next best actor that could prevent unlawful activity online.[20] As a result, despite an intermediary's liability being derivative of the primary

[13] ibid 113–114.

[14] ibid 114.

[15] Rebecca Tushnet, 'Power Without Responsibility: Intermediaries and the First Amendment' (2008) 76 The George Washington Law Review 986, 1007.

[16] Riordan, 'Principles of Secondary Liability' (n 10) 113.

[17] Jaani Riordan, 'A Taxonomy of Internet Intermediaries' in Jaani Riordan, *The Liability of Internet Intermediaries* (Oxford University Press 2016) 27.

[18] David S Ardia, 'Free Speech Savior or Shield for Scoundrels: An Empirical Study of Intermediary Immunity under Section 230 of the Communications Decency Act' (2010) 43 Loyala of Los Angeles Law Review 373, 378.

[19] Ronald J. Mann and Seth R. Belzley, 'The Promise of Internet Intermediary Liability' (2005) 47 William and Mary Law Review 239, 259, 268; Balkin (n 6) 2308.

[20] Mann and Belzley (n 19) 259.

wrongdoer's unlawful activity,[21] they are often key targets in lawsuits. The growth of social media communication platforms has only further enhanced this phenomenon.[22]

This is often referred to as a gatekeepers' liability approach to regulation. Where it is difficult or impractical to regulate unlawful conduct by prosecuting each primary wrongdoer, the law may deem it necessary to tackle the problem through holding 'gatekeepers' liable.[23] Gatekeepers are private parties who can disrupt unlawful activity by "withholding their cooperation from wrongdoers".[24] For example, a law that punishes bars for serving alcohol to underage customers uses the gatekeeping power of liquor license holders to tackle the problem of underage drinking. In the context of online content, where governments cannot hold numerous, anonymous, and foreign internet users responsible, they may seek to regulate online content by holding liable the services that constitute the digital architecture of the internet.[25] Intermediary liability thus serves the goals of governments and litigants seeking to regulate online content. However, unlike a bar serving alcohol, intermediaries are vital facilitators of free expression. Thus, imposing gatekeeping liability on them has consequences for free speech.

B. Free speech implications of secondary liability for intermediaries

There are three key reasons why imposing liability on intermediaries raises free speech concerns. First, intermediaries host and transmit both lawful and unlawful content. If a bar inadvertently refuses to serve an of-age customer, the consequence is a disgruntled customer. Where the threat of liability causes an intermediary to inadvertently restrict lawful content, an individual's free speech rights are interfered with. Second, intermediaries host and transmit content that is not their own, so they have few to no incentives to continue hosting and transmitting user content where such content represents a liability risk.[26] This is compounded by the fact that intermediaries typically do not receive sizeable revenue from any single piece

[21] Riordan, 'Principles of Secondary Liability' (n 10) 114.

[22] Ardia (n 18) 378.

[23] Reinier H Kraakman, 'Gatekeepers: The Anatomy of a Third-Party Enforcement Strategy' (1986) 2(1) Journal of Law, Economics, & Organization 53, 53; Chinmayi Arun, 'Gatekeeper Liability and Article 19(1)(A) of the Constitution of India' (2014) 7 NUJS Law Review 73, 76.

[24] Arun (n 23) 76.

[25] Balkin (n 6) 2308.

[26] ibid 2309; Eric Goldman, 'Why Section 230 Is Better Than the First Amendment' (2019) 95 (1) Notre Dame Law Review 33, 41.

of content.[27] Third, the flip side of intermediaries being well placed to restrict unlawful content is that users rely on intermediaries to exercise their free expression rights.[28] The cumulative effect of these factors is that if intermediaries are consistently held liable for all unlawful content on their networks, to avoid such liability they may 'err on the side of caution' and remove lawful content that raises even a remote risk of liability; this will ultimately lead to the suppression of lawful content of internet users.[29] Such an approach would also ultimately transform intermediaries into quasi-publishers who scrutinise every item they disseminate to avoid liability, and make impossible the rapid user-generated and content-driven internet as we know it.

For example, an intermediary may aggressively censor and filter user-generated content because it is concerned that a single user could post something unlawful, and if the intermediary is held secondarily liable for this user's unlawful content, such a verdict may bankrupt the intermediary or lead to the imprisonment of senior executives. While it may be economically sensible for the intermediary to adopt this aggressive approach to reviewing user content, this would have severe free speech consequences as several users who post legal content may have their content censored by the intermediary. Thus, holding intermediaries liable for unlawful content creates a situation where the law of secondary liability creates incentives for intermediaries to over-censor, restricting the lawful content of its users.[30]

To protect the free expression rights of internet users from private intermediaries who seek to avoid liability, and to incentivise intermediaries to host user-generated content, legislatures around the world have conferred on intermediaries a measure of immunity from secondary liability (i.e., safe harbours).[31] Cases such as *Stratton Oakmont v. Prodigy Services*[32] in the United States, and *Avnish Bajaj vs. State (NCT of Delhi)*[33] in India directly track this

[27] Goldman, 'Why Section 230 Is Better Than the First Amendment' (n 26) 41.

[28] Klonick (n 5) 1604.

[29] Balkin (n 6) 2309.

[30] ibid 2310.

[31] Graeme B Dinwoodie (ed), *Secondary Liability of Internet Service Providers* (1st ed. 2017, Springer International Publishing 2017) 19.

[32] 1995 WL 323710 (New York Supreme Court, 24 May 1995).

[33] 2008 SCC OnLine Del 688.

development, where intermediaries were held liable for content on their networks leading to legislative responses conferring a shield from secondary liability on intermediaries.[34]

It is true that safe harbours alone do not guarantee free expression online. In particular, the coercive actions of governments,[35] and the informal access and cooperation offered by platforms to States to restrict free speech online[36] represents an equally important area of concern. Similarly, as discussed above, provisions such as Section 69A of the IT Act, that allow for direct blocking of content at the government's behest, represent infringements on free speech and ought to be scrutinised closely. However, safe harbours drastically decrease the risk of an intermediary being held secondarily liable for unlawful content on its network and have been described as "among the most important protections of free expression" in the internet age.[37] This is because safe harbour allows intermediaries to host user-generated content without having to aggressively censor user content, thus protecting the free expression rights of users. Safe harbour has facilitated the creation of "user-generated content services" such as Wikipedia and YouTube while simultaneously protecting the voices of marginalised communities.[38] Finally, because safe harbour is also available where State authorities seek to restrict content by holding intermediaries liable (e.g., through criminal prosecutions), it also acts as a bulwark against government censorship.

C. Regulatory tool and balancing act

For the first fifteen years of the 21st century, debates regarding the curbing of online harms were primarily viewed through the lens of secondary liability. While the United States offers intermediaries largely unconditional immunity,[39] other jurisdictions have tied certain

[34] Balkin (n 6) 1310–2314; Klonick (n 5) 1608; Software Freedom Law Centre, 'Intermediary Liability 2.0: A Shifting Paradigm' (*Software Freedom Law Centre*, 1 March 2019) 10 <https://sflc.in/wp-content/uploads/2019/03/Intermediary_Liability_2_0_-_A_Shifting_Paradigm.pdf > accessed 26 May 2022.

[35] Derek E Bambauer, 'Against Jawboning' (2015) 100 Minnesota Law Review 51; Daphne Keller, 'When Platforms Do the State's Bidding, Who Is Accountable? Not the Government, Says Israel's Supreme Court' (*Lawfare*, 7 February 2022) <https://www.lawfareblog.com/when-platforms-do-states-bidding-who-accountable-not-government-says-israels-supreme-court> accessed 15 April 2023.

[36] Niva Elkin-Koren and Maayan Perel, 'Guarding the Guardians: Content Moderation by Online Intermediaries and the Rule of Law' in Giancarlo Frosio (ed), *Oxford Handbook of Online Intermediary Liability* (Oxford University Press 2020) 671–672.

[37] Balkin (n 6) 2313.

[38] Goldman, 'Why Section 230 Is Better Than the First Amendment' (n 26) 33–34, 42. "Not only are marginalized voices more likely to be targeted by people in positions of power, but Internet services are less likely to worry about marketplace or reputational consequences of removing content from marginalized communities."

[39] Communications Decency Act 1996, 47 U.S.C. s. 230. Jack L Goldsmith and Tim Wu, *Who Controls the Internet? Illusions of a Borderless World* (Oxford University Press 2006) 19–22. Noting that the Communication

conditions to the grant of safe harbour.[40] The contours and conditions for safe harbour have thus become a key regulatory tool to govern online content. For example, Section 79(3) of the IT Act stipulates that to retain safe harbour, when an intermediary receives "actual knowledge" of unlawful content on its network, it must "expeditiously" remove this content.[41] This is an example of a government utilising the carrot of safe harbour (and the stick of liability) to incentivise intermediaries to assist efforts to remove unlawful content.

However, pre-conditions to safe harbour can extend beyond removal of unlawful content and can even be used to facilitate censorship. For example, Balkin notes that states may use the promise of safe harbour (and the corresponding threat of liability if the safe harbour is revoked) to require intermediaries to filter content, shutdown accounts, or hand over personal information of users.[42] This demonstrates how the regulatory tool of safe harbour can also be misused to coerce platforms into doing the State's bidding. As discussed below in Section 5, there exists a case for simplifying safe harbour to avoid these risks. Adding numerous pre-conditions to safe harbour also has the effect of narrowing immunity and increasing the threat of liability, which may cause intermediaries to 'err on the side of caution' and over-censor their users' content to avoid liability.

As a regulatory tool, conditional safe harbour regimes such as that of India, aim to strike a balance between incentivising intermediaries towards minimising unlawful content through the threat of secondary liability, while offering them sufficient protection from liability that they do not over-censor the content of their users to avoid liability. Offer too little protection, and intermediaries may begin to aggressively curb the free expression of their users, offer too much protection, and intermediaries may wilfully be unresponsive to unlawful content. Much intricacy lies in exactly how this balance is best struck with different countries taking different

Decency Act's broad immunity is in part grounded in the American constitutional tradition of minimal government interference on speech courtesy First Amendment doctrine.

[40] Directive 2000/31/EC of 8 June 2000 on electronic commerce, Articles 12-14; Brazilian Civil Rights Framework for the Internet, (Federal) Law No 12.965 of 23 April 23 2014, Article 19. *See also* Park (n 7).

[41] The decision in *Shreya Singhal v Union of India* [2015] (5) SCC 1 has clarified that "actual knowledge" means a court or government order directing the removal of content. Rule 3(1)(d) of the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021 [**Intermediary Guidelines**] states that content must be removed within 36 hours.

[42] Balkin (n 6) 2311. Intermediary Guidelines, r. 3(1)(j).

approaches and much of intermediary liability scholarship being dedicated to how intermediary immunity ought to be structured.[43]

For the purposes of this Article however, it is sufficient to note that the historical practice of utilising secondary liability lawsuits to regulate online content has led to intermediary liability and safe harbours becoming a key site of contestation in regulating online harms. As this Article goes on to demonstrate, in India this has led to a conflation in the regulatory rhetoric between the systemic and structural harms caused by online platforms, and questions of liability for unlawful content. The next section of the article attempts to demarcate the two types of harms by demonstrating how the issue of liability for content is enforced through lawsuits in which individual pieces of content are in dispute. This is then contrasted with the systemic issues caused by the operations and business models of large intermediaries that occur on an ongoing basis.

III. "INTERMEDIARY" AND SAFE HARBOUR GROUNDED IN SECONDARY LIABILITY SUITS

As highlighted above, Section 79 of the IT Act provides intermediaries with a conditional safe harbour. Section 79(1) states that intermediaries "shall not be liable for any third-party information" made available or hosted by them, while Sections 79(2) and 79(3) stipulate the conditions intermediaries must satisfy to avail of this immunity.

It is important to note that only 'intermediaries' are eligible for safe harbour under Section 79. Thus, qualifying as an "intermediary" under Section 2(1)(w) of the IT Act is a threshold condition for availing safe harbour.[44] An examination of the definition of "intermediary" and Section 79 demonstrates how the regulatory tool of intermediary liability is enforced through suits against individual pieces of content.

A. Intermediary as an activity or function

The Indian Government has recently stated that an entity may lose its intermediary "status" or its safe harbour "status" for failing to satisfy the conditions set out in Section 79 of the IT Act

[43] Giancarlo F Frosio (ed), *The Oxford Handbook of Online Intermediary Liability* (First edition, Oxford University Press 2020); Vasudev Devadasan, 'Report on Intermediary Liability in India' (Centre for Communication Governance 2022) <https://papers.ssrn.com/abstract=4343781>; Jaani Riordan, 'Safe Harbours' in Jaani Riordan, *The Liability of Internet Intermediaries* (Oxford University Press 2016); Rishabh Dara, 'Intermediary Liability in India: Chilling Effects on Free Expression on the Internet' [2011] SSRN Electronic Journal <http://www.ssrn.com/abstract=2038214> accessed 18 February 2021; Eric Goldman, 'The Ten Most Important Section 230 Rulings' (2017) 20 Tulane Journal of Technology & Intellectual Property; Jonathan Zittrain, 'A History of Online Gatekeeping' (2006) 19 Harvard Journal of Law & Technology 253.

[44] Devadasan (n 43) 24.

and the Intermediary Guidelines.[45] For example, in a letter to Twitter, the Indian Government suggested that the Government could 'withdraw Twitter's immunity' under Section 79.[46] It is submitted that this is a misconception regarding how intermediary liability operates. The defence of Section 79's safe harbour is only triggered when an intermediary is sued for transmitting or hosting unlawful third-party content.[47] This has two key consequences: (i) the pre-conditions for safe harbour are only relevant once a suit for secondary liability is filed; and as a result (ii) intermediary liability and pre-conditions to safe harbour, as a regulatory schema, are enforced through individual secondary liability suits. Crucially, this means that for the purposes of liability, an entity can only be classified as an "intermediary" entitled to Section 79 "safe harbour" *in the context of a lawsuit* for unlawful content. Saying that an entity has lost its intermediary or safe harbour "status" (and is liable), or that such status could be "withdrawn" in the abstract, outside of a specific suit for content, is meaningless.

A perusal of the IT Act demonstrates how the questions of whether an entity is an "intermediary", and whether it is entitled to safe harbour, must be assessed vis-à-vis the specific content it is sued for. Section 2(1)(w) of the IT Act defines an "intermediary" as an entity which, receives, stores, transmits, or provides any service with respect to a particular electronic record, "on behalf of another person." A reading of this definition reveals that an "intermediary" only exists with respect to a particular electronic record(s); i.e., an intermediary is not a type of *entity per se* but rather a *function*. What is that function? The receiving, storing, transmitting, or providing of a service with respect to content on behalf of another person (i.e., third-party content). Thus, where an entity receives, stores, transmits, or provides a service in relation to third-party content, the entity is an intermediary *with respect to* that third-party content. If the entity is sued for *that* content, it may claim to be an intermediary, if it is sued

[45] 'Twitter Loses Intermediary Status over Non-Compliance with New Rules: Rpt' *Business Standard* (16 June 2021) <https://www.business-standard.com/article/news-ani/twitter-loses-its-status-as-intermediary-platform-in-india-due-to-non-compliance-with-new-it-rules-121061600199_1.html> accessed 4 April 2023; Sourabh Jain, 'Twitter Will Lose Its Intermediary Status If It Does Not Comply with the New IT Rules by July 4' *Business Insider* (29 June 2022) <https://www.businessinsider.in/tech/news/twitter-will-lose-its-intermediary-status-if-it-does-not-comply-with-the-new-it-rules-by-july-4-2022/articleshow/92534121.cms> accessed 4 April 2023.

[46] *X Corp. v Union of India* Writ Petition 13710 of 2022 (30 June 2023, High Court of Karnataka). Reference to letter dated 27 June 2022.

[47] *Amazon Seller Services Pvt. Ltd. v Amway India Enterprises Pvt. Ltd.* [2020] SCC OnLine Del 454 [143]. "The Plaintiffs have to first show that there had been a violation of any of their rights due to the Defendants' activities before the "affirmative defence" of Section 79 could be sought to be invoked." Further noting that Section 79 cannot be enforced by the Plaintiffs as a positive obligation on intermediaries. *Myspace Inc. v Super Cassettes Industries Ltd.* [2016] SCC OnLine Del 6382 [51]. "Section 79 is neither an enforcement provision nor does it list out any penal consequences for non-compliance. It sets up a scheme where intermediaries have to follow certain minimum standards to avoid liability; it provides for an affirmative defence."

for some other content with respect to which it does not perform this function, it cannot claim to be an intermediary.

For example, if YouTube is sued with respect to a video a user uploaded, it may claim to be an intermediary as it is hosting content on behalf of another person. But if YouTube is sued for a video, it created itself and subsequently uploaded, with respect to this second video, it would not be an intermediary as it was not hosting content "on behalf of another person". This understanding is further buttressed by the Explanation appended to Section 79, which states that, where an intermediary is sought to be held liable for "third-party information", the phrase "third-party information" means the information "dealt with by an intermediary *in his capacity as an intermediary* (emphasis supplied)." Thus, in the context of Section 79 and intermediary liability, it is the relationship or function that the entity has or performs vis-à-vis the content it is sued for that determines whether it is an intermediary or not.

This has two important consequences. First, the legal question of whether an entity is an "intermediary", and whether it is entitled to safe harbour protections, is to be adjudicated in relation to the third-party content it is sued for. Second, the potential liability of the intermediary is limited to the specific third-party content it is sued for. This understanding flows directly from the fact that in a suit for secondary liability against an intermediary, the wrongful act is the creation and sharing of the unlawful content, and the intermediary's liability is based on its role in hosting or transmitting *that* content. As the suit is based on the particular content, liability is limited to the subject matter of the suit, the unlawful content. Returning to the YouTube example, if YouTube is sued for the video uploaded by a user, its secondary liability in such a suit is limited to the liability associated with that one video, not any or all unlawful content on YouTube's platform.

B. An interpretation that may seem obvious

This interpretation is not novel, but rather flows from the secondary liability underpinnings of intermediary liability. An examination of two paradigmatic examples from the United States and Europe reveals that the above discussed interpretation is the appropriate lens of analysis. Both cases demonstrate that the focus of any intermediary liability query should be on the function or activity of an entity vis-à-vis specific content it is sued for.

Roomates.com was a roommate matching website where users could create profiles and indicate the kind of roommates they were looking for. User profiles had two principal components. First, the website provided a series of drop-down menus where users described

themselves (e.g., gender, sexual orientation, presence of children), and their preferences for roommates.[48] The drop-down menus for the "My Roommate Preferences" required users to select between "Straight or gay" males, only "Straight" males etc.[49] The second component of the user profile required users to add "Additional Comments", where users were provided with a blank text-box to indicate any additional preferences. In the "Additional Comments" section, users often filled out preferences such as "Prefer white Male roommates" or "NOT looking for black Muslims."[50] The website was sued for allegedly violating the Fair Housing Act, and in *Fair Housing Council v. Roommates.com*,[51] the Ninth Circuit was tasked with determining whether Roommates.com was eligible for safe harbour under Section 230 immunity of the Communication Decency Act (CDA).

Under the CDA, immunity from secondary liability is granted to 'interactive computer services', but not "information content providers". An information content provider is defined as "*any person or entity that is responsible, in whole or in part for the creation or development of information provided through the Internet*."[52] Thus, like the IT Act, where an interactive computer service deals with content on behalf of another person (third-party content), it is eligible for immunity, but if it is responsible for the creation of the content (in whole or in part), it is a content provider ineligible for immunity.

In its evaluation of liability and immunity, the Ninth Circuit first identified the content that Roomates.com was being sued for, noting that it was sued for both components of the user profiles, the questionnaires in the form of drop-down menus and for the content in the "Additional Comments" section. The Court then noted that Roommates.com had different functions vis-à-vis the two sets of content. With respect to the questionnaires, the Ninth Circuit ruled that by creating the options in the drop-down menus that ultimately formed the content of the allegedly discriminatory profiles, Rommates.com was responsible at least in part for the 'creation or development' of this content.[53] Thus, vis-à-vis the content created because of the drop-down menus, Roomates.com was ineligible for the Section 230 immunity because the content was not entirely created by third parties. However, vis-à-vis the "Additional

[48] *Fair Housing Council v Roommate.com* 521 F.3d 1157 (9th Circuit 2008) 5717.

[49] *Fair Housing Council v Roommate.com* 521 F.3d 1157 (9th Circuit 2008) 5717.

[50] *Fair Housing Council v Roommate.com* 521 F.3d 1157 (9th Circuit 2008) 5722.

[51] 521 F.3d 1157 (9th Circuit 2008).

[52] *Fair Housing Council v Roommate.com* 521 F.3d 1157 (9th Circuit 2008) 5716.

[53] *Fair Housing Council v Roommate.com* 521 F.3d 1157 (9th Circuit 2008) 5716.

Comments" section, the Court noted that the blank text-box provided to users "suggests no particular information that is to be provided by members; Roommate certainly does not prompt, encourage or solicit any of the inflammatory information provided by some of its members".[54] Thus, the Court concluded that for the content in dispute in the "Additional Comments" section, Roomates.com was not a content provider and thus eligible for the Section 230 immunity.

The question of whether drop-down menus constitute sufficient involvement to disqualify the answers from being third-party content may be subject to reasonable disagreement (indeed, even in *Roommates.com*, there was a dissent on this point). However, the *methodological* approach of the Court is clear and mirrors the analysis set out in the previous section. The court examined Roommates.com's eligibility for safe harbour based on its *relationship with the content it was sued for*. Where the website was sued for more than one piece or type of content, the court independently examined the website's relationship with each type of content it was sued for and determined the eligibility for immunity based on the functionality or activity undertaken vis-à-vis each type of content. Thus, the same entity may both be an intermediary, and not be an intermediary, based on its functionality with respect to the content it is sued for. This demonstrates how, because intermediary liability is enforced through lawsuits against allegedly unlawful pieces of content, both the eligibility for immunity and the securing of such immunity is fundamentally tied to content the intermediary is sued for.

Perhaps an even more emphatic statement of this interpretation comes from practice under the European E-Commerce Directive. The Directive served as the primary inspiration for India's own safe harbour provision,[55] and thus an understanding of how the immunities to liability are interpreted is relevant. In one of the earliest cases under the Directive, e-Bay was sued for allegedly making available trademark-infringing content and a question arose as to whether it was entitled to safe harbour under the e-Commerce Directive. Broadly, the Directive provides exemptions to three types of activities, 'transmission' (or mere conduit), 'caching', and 'hosting'. When a question arose as to which of these three categories e-Bay fell under, the opinion of the Advocate General noted:

> *Moreover, as a general remark on the three exceptions laid down in Articles 12, 13 and 14 of Directive 200/31, I should say something which may seem obvious. The three articles intend to create exceptions*

[54] *Fair Housing Council v Roommate.com* 521 F.3d 1157 (9th Circuit 2008) 5723.

[55] Software Freedom Law Centre (n 34) 11.

> *to certain types of activity exercised by a service provider. To my understanding, it is inconceivable to think that they would purport to exempt a service provider type as such.*
>
> *Indeed, it is difficult to see that Directive 2000/31 would impose three distinct types of activity which would only be exempted if each of them is exercised in a watertight compartment. If one company is caching and another one is hosting, they surely are both exempted. Yet such separation may be extremely rare. In my view, if one company does both – which does not appear at all exceptional in the real world, the exemptions should apply to that one entity too. The same should apply if one or more of the exempted activities are combined with an internet content provider's activity. It would be unworkable to reserve the exemptions to certain business types, especially in an area characterised by constant and almost unpredictable change.* […]
>
> *I do not think that it is possible to sketch out parameters of a business model that would fit perfectly to the hosting exemption. And even if it were, a definition made today would probably not last for long. Instead, we should focus on a type of activity and clearly state that while certain activities by a service provider are exempt from liability, as deemed necessary to attain the objectives of the directive, all others are not and remain in the 'normal' liability regimes of the Member States, such as damages liability and criminal law liability* (emphasis supplied).[56]

The Advocate General's Opinion demonstrates why determinations of safe harbour are based on the activity performed vis-à-vis the allegedly unlawful content, as opposed to conferring or denying safe harbour to a type of entity. A single entity may perform multiple functions or activities. The different activities performed by a single entity may be inside the scope of immunity, outside the scope of immunity, or treated slightly differently by a specific immunity regime. Within the three classifications of activities under the Directive (transmission, caching, and hosting), different elements of the intermediary's operations may come under different

[56] Opinion of Mr. Advocate General Jaaskinen in *L'Oréal SA v eBay International AG* (Case C-324/09, Opinion delivered on 9 December 2010). The Court adopted the reasoning adopted by the Advocate General but ultimately held that eBay was not entitled to immunity under Article 14 of the eCommerce Directive because it processed customer data to knowingly optimise sales offerings in a non-neutral manner.

classifications, while some activities may be entirely beyond the scope of immunity. The question of which activity of an intermediary should be examined, as demonstrated by *Roommates.com*, is to be determined by the content it is sued for.

The above-discussed statutory Explanation appended to Section 79 would indicate that the position is the same in India. Because ultimately safe harbour is an exception to *liability for third-party content*, it is the function or activity that the intermediary performs with respect to the allegedly *liability-generating content* that forms the heart of the adjudicatory question of liability and immunity. The Opinion also recognises that determining eligibility for immunity based on a "business model" or static classification would soon be rendered redundant by the rapid innovation in online services.

As the present author has argued elsewhere:

> *One may classify an entity as a 'borrower' only in relation to a loan it has taken but classify the entity as a 'company' or 'partnership' across all its functions. Similarly, a website may be an "intermediary" when it transmits 'Picture A' that is third-party content shared by a user, but not an "intermediary" when it transmits 'Picture B' that is its own content.*[57]

Therefore, eligibility for safe harbour must be determined by the activities performed by the entity with respect to the allegedly unlawful content, as opposed to the nature of the entity.

C. Confusions and conflations in India

Indian courts have often struggled to determine when an entity is acting as an "intermediary". For example, in several cases, judges have determined that the question of whether an entity is an "intermediary" eligible for safe harbour should be determined at trial.[58] This approach prevents defendants from getting cases against them dismissed at a preliminary stage, raising litigation expenses, and undermining a key procedural benefit of safe harbour. Immunity should ideally absolve entities entitled to safe harbour from engaging in costly trials.[59] It is submitted that while complex cases may require detailed evidence to be led on the exact

[57] Devadasan (n 43) 24.

[58] ibid 25. *Amazon Seller Services Pvt. Ltd. v Amway India Enterprises Pvt. Ltd.* 2020 SCC OnLine Del 454 [18]-[19], [141]; *Google India Pvt. Ltd. v Visaka Industries* 2020 (4) SCC 162 [153]; *Sorting Hat Technologies Pvt. Ltd. v Fermat Education* 2019 SCC OnLine Mad 33436 [20].

[59] Goldman, 'Why Section 230 Is Better Than the First Amendment' (n 26) 39. Even if some evidence is required to be led, safe harbour may narrow the scope of the evidentiary inquiry.

functionality of a particular website or service vis-à-vis the content it is sued for, in most cases the question of whether an entity is an intermediary eligible for safe harbour can be determined at the interim or preliminary stage through a cursory examination of its function vis-à-vis the disputed content. Avoiding lengthy trials will also lead to the saving of judicial time, as claims against intermediaries can be dismissed at a preliminary stage where the defence of safe harbour is made out.[60]

In other cases, Indian courts have strayed even further from the above-outlined understanding of when an entity is an "intermediary." For example, in (since overruled) *Christian Louboutin v. Nakul Bajaj*,[61] the High Court of Delhi was tasked with determining whether the online shopping platform 'Darveys.com' was eligible for Section 79 protection. Christian Louboutin alleged that the website was making available trademark-infringing listings on its website. In its judgement, the High Court of Delhi opined that if an online commerce platform performed certain 'active' services such as: uploading the entry of the product on the website, packaging and transporting products, and promoting or advertising the products on its platform, the platform would cease to be an intermediary eligible for safe harbour.

By setting out a prescriptive and static approach of the type of business that is eligible for safe harbour, the High Court runs into precisely the issue highlighted by the Advocate General in *L'Oréal SA v. eBay International AG*.[62] The Court effectively attempts to determine the type of business that is an intermediary eligible for protection. But this is a fraught endeavour, as the Advocate General observed, because the nature of online services is constantly evolving. The judgement was criticised precisely because it was unclear what the court's criteria for 'active' services was grounded in, as it finds no backing in either the definition of "intermediary" in Section 2(1)(w) of the IT Act, Section 79 of the IT Act or in trademark or consumer-protection legislation.[63] By stipulating the kinds of business activities that it believed intermediaries should perform, as opposed to examining whether the content in question was third party content, and what role the website played with respect to this content, the High Court risked stifling new and innovative platforms by raising the spectre of liability for e-commerce platforms that offered novel functionality. By failing to consistently interpret

[60] ibid 41.

[61] 2018 SCC OnLine Del 12215.

[62] Court of Justice of the European Union, Case 324/09 (12 July 2011).

[63] Vasundhara Majithia, 'The Changing Landscape of Intermediary Liability for E-Commerce Platforms: Emergence of a New Regime' 15 The Indian Journal of Law and Technology 470.

Sections 2(1)(w) and Section 79 in line with the secondary liability principles they represent, Indian courts have contributed to the confused regulatory discourse surrounding safe harbour in India.

Thus, rather than determine the type of entity or business model that constitutes an intermediary, it is submitted that courts should instead look at the activities they perform in relation to the content that forms the subject matter of the litigation. Examining the question of "intermediary" as an issue of function rather than status, is both faithful to the text of the IT Act and provides courts with a simple and elegant way of answering the preliminary question of whether an entity is eligible for safe harbour. Courts should thus refrain from asking, 'is this entity an intermediary?', and instead ask 'was this entity acting as an intermediary in relation to the content it is being sued for?'

Just as the question of whether an entity is acting as an "intermediary" must be determined in relation to specific content, the issue of safe harbour is also to be adjudged based upon the allegedly unlawful content. The statutory immunity provided by Section 79(1) is conditioned on the satisfaction of Sections 79(2) and 79(3). Thus, an intermediary can only be disentitled from safe harbour for a failure to satisfy Sections 79(2) and 79(3). Crucially, nowhere in these sub-sections does it say that an intermediary must have never been *previously* found liable or dis-entitled to Section 79(1) immunity. This makes it clear that safe harbour cannot be lost in the abstract or "withdrawn" by the Government. It can only be adjudicated upon in a suit against content. Even if an intermediary were to lose safe harbour in a specific case, this would not automatically result in the intermediary being disentitled to safe harbour in other (or future) lawsuits against content on its network. Rather, in those separate lawsuits, the question of whether the entity is an "intermediary" and has "safe harbour" would have to be independently adjudicated based on the entity's relationship and actions vis-à-vis the allegedly unlawful content in each dispute.

This discussion demonstrates how intermediary liability and safe harbour as a regulatory tool (properly interpreted and applied) is both grounded in and enforced through individual lawsuits alleging secondary liability for content. As lawmakers attempt to re-design safe harbours in the era of platform governance it is important to understand this inherent nature of intermediary liability. Where lawmakers seek to incentivise specific platform behaviour through the promise of safe harbour and threat of liability, they must be cognisant of how such liability rules are enforced. Crucially, as the next section argues, enforcement through lawsuits against content

has key implications for the types of obligations that can be imposed as pre-conditions to safe harbour. As an analysis of India's Intermediary Guidelines demonstrates, there are severe issues in attempting to impose complex obligations on intermediaries through liability rules that are enforced through lawsuits against content.

IV. LIABILITY RULES AS BLUNT INSTRUMENTS

Sections 79(2) and 79(3) broadly stipulate three key conditions for eligible intermediaries to avail of safe harbour: (i) intermediaries must not aid or abet unlawful acts;[64] (ii) when an intermediary receives "actual knowledge" of unlawful content of unlawful content on its network, it must remove such content;[65] and (iii) it must act with "due diligence".[66] The content of what acting with "due diligence" means is set out in the Intermediary Guidelines, which constitutes delegated legislation under the IT Act. The Guidelines classify intermediaries into two groups: intermediaries simpliciter, and 'significant social media intermediaries' ('**SSMIs**'), the latter being intermediaries which have more than 5 million registered users in India and primarily enable online interaction between users, allowing them to upload and share content.[67] Subsequent amendments to the Guidelines also defined an "online gaming intermediary."[68] The Intermediary Guidelines impose wide array of obligations, particularly on SSMIs, ranging from removal of content upon receiving actual knowledge, to data retention, appointing local officers, publishing transparency reports, proactive monitoring, and grievance redressal.[69] Crucially, the obligations in the Intermediary Guidelines effectively constitute pre-conditions to safe harbour. This is because the obligations specified in the Intermediary Guidelines constitute the "due diligence" activities intermediaries must undertake to qualify for safe harbour under Section 79.

Section 3 of the present article noted how safe harbour was evaluated on a case-by-case basis in suits against content. However, the nature of some of the obligations imposed in the Guidelines conflict with the enforcement mechanism of individual lawsuits. Some of the

[64] Information Technology Act, 2000, s. 79(3)(a).

[65] Information Technology Act, 2000, s. 79(3)(b).

[66] Information Technology Act, 2000, s. 79(2)(c).

[67] Intermediary Guidelines, r. 2(1)(v); Ministry of Electronics and Information Technology, Notification S.O. 942(E) dated 25 February 2021 setting the threshold of users to qualify as an SSMI at 5 million.

[68] Intermediary Guidelines, r. 2(1) (q b) as amended by the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2023 G.S.R. 275(E) dated 6 April 2023.

[69] Intermediary Guidelines, r. 3, 4.

obligations imposed through the Guidelines are better described as continuing obligations on entities as opposed to obligations that have any nexus with a suit against allegedly unlawful content. For example, Rule 4(1)(a) of the Intermediary Guidelines requires SSMIs to appoint a local compliance officer in India. Despite this requirement effectively constituting a pre-condition to safe harbour, the nature of such an obligation is logically one that: (i) would appear to apply even when an intermediary is not sued for content; and (ii) is unrelated to any specific content that an intermediary is being sued for.

In fact, this precise obligation caused much confusion when Twitter briefly failed to appoint a local officer after the Intermediary Guidelines were adopted, leading to the aforementioned claims that Twitter had lost its "intermediary" and "safe harbour" status.[70] The source of this confusion is clearly that the nature of the obligation is one that logically requires continual compliance; the intention is to ensure all SSMIs have local officers irrespective of whether they are sued for hosting allegedly unlawful content. But the manner of the obligation's imposition, as a pre-condition to safe harbour, leads to the legal conclusion that it can only be evaluated in the context of a lawsuit for secondary liability for hosting allegedly unlawful content.

The Intermediary Guidelines contain other such provisions that do not align with the enforcement mechanism of intermediary liability. For example, Rule 4(1)(d) of the Guidelines requires SSMIs to publish "compliance reports" documenting the number of complaints received and content removed consequently. While undoubtedly normatively desirable, as an obligation that can only be verified and enforced when an intermediary is sued for content, compliance may be inconsistent. If a SSMI is never sued for unlawful content, the SSMI may choose not to publish any transparency reports and theoretically face no legal consequences (though they may face government backlash in other areas). Further, even if it is sued, courts are ill placed to evaluate such an obligation in a suit for unlawful content. For example, where a SSMI is sued for hosting unlawful content, it would be a tangential inquiry at best for a court to have to examine whether the SSMI has been publishing monthly compliance reports since the advent of the Guidelines. Arguably, the courts efforts should be directed at determining key issues of secondary liability such as what function the intermediary carried out vis-à-vis the content and whether the content is illegal.

[70] Jain (n 45); 'Twitter Loses Intermediary Status over Non-Compliance with New Rules: Rpt' (n 45).

Going even further, it is questionable how some of the obligations imposed under the Intermediary Guidelines can even be engaged through a suit against unlawful content. For example, Rule 4(8) of the Guidelines requires SSMIs to grant users notice prior to removing content and offer users "an adequate and reasonable opportunity to dispute" the removal action taken by SSMIs and seek reinstatement. This is effectively an obligation to prevent platforms from arbitrarily removing user content without granting their users a measure of due process. However, because the obligation, strictly speaking, constitutes a pre-condition to safe harbour, it can only be evaluated when an intermediary is sued for unlawful content.

Suppose a platform removes content without providing a user notice or hearing, in breach of Rule 4(8). The user cannot initiate an action alleging a violation of Rule 4(8) directly. This is because Rule 4(8), and the entirety of the Intermediary Guidelines, constitute conditions that the intermediary must satisfy to avail of statutory immunity when sued for hosting or transmitting unlawful content.[71] This immunity is specifically for *liability* for third party content "made available" by the intermediary.[72] Thus, Rule 4(8) is only legally triggered when an intermediary is sued for its role in disseminating allegedly unlawful content, and a court subsequently evaluates the intermediary's safe harbour defence under Sections 79(2) and 79(3) and the Intermediary Guidelines. It follows then that to engage the operation of Rule 4(8), the user must sue the SSMI for hosting or transmitting unlawful content.

However, where a user is seeking the reinstatement of content, as they would be if they were aggrieved by the intermediary's non-compliance with Rule 4(8), the user cannot initiate a suit alleging the content is unlawful or generates liability. In fact, their claim is the contrary, that the content is lawful and must be carried by the intermediary.[73] Thus, it is unclear what kind of an action a user must bring to engage Rule 4(8), severely undermining the efficacy of such a provision. This is not merely an abstract discussion, as close to two years since the Intermediary Guidelines have been adopted, there is no evidence that SSMIs are complying with Rule 4(8).[74]

Similar issues plague the 'Grievance Appellate Committees' ("**GACs**"). Under Rule 3A of the Intermediary Guidelines, users may appeal to a GAC against the decision of an intermediary

[71] *Myspace Inc. v Super Cassettes Industries Ltd.* 2016 SCC OnLine Del 6382 [51].

[72] Information Technology Act, 2000, s. 79(1). "An intermediary shall not be liable for any third-party information, data, or communication link made available or hosted by him."

[73] Devadasan (n 43) 75.

[74] ibid 73–76.

to remove or keep-up content.[75] However, as with Rule 4(8), compliance with the orders of the GAC are a pre-condition to safe harbour under Section 79 of the IT Act.[76] Thus, if a GAC were to direct an intermediary to keep-up or reinstate content, and the platform refused, it is difficult to envision what the legal and monetary consequence of a loss of safe harbour in such a situation would be.[77] As with Rule 4(8), the obligation to comply with the GAC's order could not be engaged outside of a suit for unlawful content (for this is what triggers the applicability of Section 79 and the Intermediary Guidelines), and an intermediary cannot be sued for content it is not hosting or transmitting. The GAC is not a court with traditional contempt powers either.

Lastly, just as a court would be ill placed to verify the ongoing publication of compliance reports; in a lawsuit against a specific piece of unlawful content, a court is poorly placed to verify if the SSMI has been continually providing hearings to all its users whose content it has removed under Rule 4(8). This criticism plagues a host of obligations imposed on intermediaries through the Intermediary Guidelines including: the obligation to clearly identify promoted content,[78] proactive monitoring of certain types of content,[79] providing users with tokens to track complaints,[80] and providing users with a 'verified' mark.[81] Violations of these obligations are not tied to specific unlawful content, and thus: (i) it is difficult to see how these obligations can be engaged if there is no underlying claim to secondary liability against content that triggers Section 79 and the Intermediary Guidelines; and (ii) where actions for liability for content do exist, courts are poorly placed to evaluate compliance with these ongoing obligations in individual lawsuits concerning specific pieces of content.

Imposing such obligations as pre-conditions to safe harbour fails to recognise that the primary enforcement mechanism for intermediary liability is lawsuits for hosting unlawful content. This incongruence between the content of the obligation and its enforcement mechanism may lead to both chronic underenforcement, as seen in the case of Rule 4(8) and selective over-enforcement, as demonstrated by the pressure applied to Twitter to comply with Intermediary

[75] Intermediary Guidelines, r. 3A (2).

[76] Intermediary Guidelines, r. 3A (7).

[77] Devadasan (n 43) 98. Section 45 of the IT Act provides for a penalty of ₹ 25,000 for the contravention of a regulation for which no penalty has been specifically prescribed and could be utilised to fine non-compliant platforms.

[78] Intermediary Guidelines, r. 4(3).

[79] Intermediary Guidelines, r. 4(4).

[80] Intermediary Guidelines, r. 4(6).

[81] Intermediary Guidelines, r. 4(7).

Guidelines.[82] On the one hand, imposing numerous complex obligations to avail of safe harbour allows the government to pressurise intermediaries, because at any given point of time, an intermediary may be in breach of some obligation and lawsuits (including criminal proceedings) could be initiated. On the other hand, absent government pressure, imposing such complex obligations through liability rules allows platforms to pick and choose their compliance. They may satisfy simple, convenient, or government-facing obligations,[83] while waiting to be sued for content (and using their considerable resources to engage in lengthy litigation battles) for other obligations. For example, SSMIs have been submitting compliance reports to the Indian Government,[84] but as noted above, there is no evidence of them notifying users and granting them a hearing prior to the removal of content under Rule 4(8).[85] The Intermediary Guidelines are a paradigmatic case of how seeking greater accountability of online platforms can lead to a narrowing of safe harbour despite the incongruity between the desired outcomes and the regulatory approach.

These criticisms of intermediary liability as a regulatory tool are not novel. Scholars have long argued that liability rules are blunt instruments.[86] Jaani Riordan notes that secondary liability "lacks granularity" and its "binary nature" compels courts to choose between complete immunity and joint liability, offering no system to proportionately allocate responsibility or impact the behaviour of intermediaries in a nuanced manner.[87] This is strikingly evident in the Intermediary Guidelines which impose numerous complicated obligations on intermediaries

[82] Kushagra Sinha, 'India's IT Rules, Twitter MD, & UP Police: Actions & Options Ahead' (*The Quint*, 24 June 2021) <https://www.thequint.com/opinion/india-it-rules-twitter-chief-manish-maheshwari-up-police-loni-case-actions-options-ahead> accessed 15 April 2023; 'As Indian Police Raid Twitter's Office, Alarming New Internet Rules Take Effect' (*Access Now*, 25 May 2021) <https://www.accessnow.org/press-release/india-twitter-new-internet-rules/> accessed 5 April 2023; Manish Singh, 'Police in India Visited Twitter Offices over "manipulated Media" Label' (*TechCrunch*, 24 May 2021) <https://techcrunch.com/2021/05/24/delhi-police-run-by-indias-central-government-raids-twitter-offices-over-manipulated-label/> accessed 5 April 2023.

[83] For example, while large social media platforms have appointed local officers in India and publish compliance reports (and may choose to comply with GAC orders), such compliance may in part be attributed to the risk governmental backlash for public non-compliance. Such backlash may include new laws on content regulation, taxation, competition, or privacy, arrests of employees, service disruptions, or seizure of assets. *See* Keller, 'When Platforms Do the State's Bidding, Who Is Accountable?' (n 35).

[84] Anandita Mishra, 'Analysis of Social Media Compliance Reports of August, 2021' (*Internet Freedom Foundation*, 25 October 2021) <https://internetfreedom.in/social-media-giants-have-released-their-compliance-reports-for-the-month-of-august-weve-analysed-them/> accessed 5 April 2023; Vasudev Devadasan, 'Compliance Reports by Social Media Platforms Are Unhelpful' (*MediaNama*, 18 April 2022) <https://www.medianama.com/2022/04/223-transparency-reports-social-media-platforms-unhelpful/> accessed 27 April 2022.

[85] Devadasan (n 43) 73–76.

[86] Kraakman (n 23) 75.

[87] Riordan, 'Principles of Secondary Liability' (n 10) 126.

but the ultimate outcome of an intermediary liability-safe harbour lawsuit can only be liability or immunity.

As lawmakers seek to create comprehensive statutory frameworks for online platforms, there is likely to be a shift from such a liability or fault-based approach to the direct imposition of statutory duties. Yet as Section 2 of this article noted that safe harbour is a key free expression protection on the internet and intermediary liability allows lawmakers to leverage the gatekeeping powers of intermediaries to remove harmful content. Thus, if intermediary liability is required to continue performing these goals within the paradigm of platform governance, complex liability rules such as the Intermediary Guidelines must be refashioned. The next section argues for a simplification of liability rules to retain the benefits of a safe harbour approach, while Section 6 demonstrates how such a simplified approach does not undermine the broader goal of platform accountability.

V. THE NORMATIVE CASE FOR SIMPLIFYING SAFE HARBOUR

Section 2 of the present article discussed how lawmakers treat intermediaries as "gatekeepers" who can disrupt unlawful activity by "withholding their cooperation from wrongdoers".[88] Where the wrongdoing is difficult to detect or prosecute, or the wrongdoers lack the capacity or incentives to comply with the law, regulating gatekeepers may be an alternative.[89] For example, returning to the analogy of underage drinking, it may be an inefficient use of public money to identify and restrict every potential underage customer, thus the law places the onus of enforcement on the holder of the liquor license rather than supervise every alcohol-seeking customer. Further, direct enforcement may have adverse societal impacts, for example high penalties on every underage individual who attempts to consume alcohol may disrupt education and families.[90] This also highlights how, where gatekeepers are concerned, the allocation of secondary liability is not necessarily a neutral allocation of responsibility, but rather maps out the key policy objectives within a specific context.[91] This is particularly important given the free speech considerations involved with regulating online intermediaries.

[88] Arun (n 23) 76.

[89] Kraakman (n 23) 56.

[90] ibid.

[91] Dinwoodie (n 31) 4.

Gatekeeping itself also involves certain costs (e.g., a restaurant may be discouraged from seeking a liquor license due to the risk of sanctions if underage customers are, even unknowingly, served in their establishment). Thus, lawmakers must structure gatekeeping regimes to balance the costs of gatekeeping against the total harm prevented by gatekeeping within the specific context they seek to regulate.[92] Reinier Kraakman identifies four criteria to determine if gatekeeping is a beneficial approach; according to him gatekeeping is justifiable where: (i) there exists serious misconduct that direct penalties cannot deter; (ii) there are missing or insufficient incentives for entities to gatekeep on their own; (iii) there exist gatekeepers who can reliably prevent misconduct; and (iv) gatekeepers can be induced through legal rules to detect misconduct at a reasonable cost.[93]

While Kraakman's analysis predates the internet, these criteria help explain why online intermediaries are required to act as gatekeepers. The prevalence of unlawful online content is high, direct penalties are ineffective against primary wrongdoers, and intermediaries may not voluntarily (or cannot) always remove unlawful content. Crucially, because content can be widely disseminated on an intermediary's network and cause real world harm in a manner unrelated to the primary wrongdoing (posting or sharing), rapid gatekeeping action by intermediaries may at times be as important, if not more so, than direct enforcement. For example, content may be posted in a country where it is protected speech but be re-shared in a country or a context where it may be unlawful or even cause violence.

The third and fourth of Kraakman's criteria are of relevance to the structuring of online gatekeeping regimes. Gatekeeping is significantly shaped by the features of the market or ecosystem it seeks to regulate, including the framing of monitoring duties.[94] Based on our above discussion concerning the Intermediary Guidelines, it is evident that if gatekeeping duties are going to be imposed through liability rules for individual pieces of content, they must be framed in congruence with the nature of the online ecosystem and the enforcement mechanism for such duties. Similarly, the structuring of gatekeeping duties will also dictate whether the net cost of such duties is justified. As observed in Section 2, poorly designed gatekeeping regimes for online intermediaries can result in the suppression of lawful speech, a heavy (arguably unjustifiable) cost to pay.

---

[92] Kraakman (n 23) 93; Mann and Belzley (n 19) 266.

[93] Kraakman (n 23) 55.

[94] ibid 81.

A. Brightline gatekeeping duties

The goal of a gatekeeping regime is to induce gatekeepers to respond to unlawful activity at a justifiable cost. For example, a key reason gatekeeping is justifiable for controlling underage drinking is because the nature of obligation imposed on establishments is targeted and effective. Requiring customers to show proof of age is a highly focused monitoring duty that exhausts almost the entirety of what must be done to prevent wrongdoing.[95] Kraakman classifies this as a '*bouncer*' gatekeeping regime, where the gatekeeper's primary role is to deny wrongdoers access to a market.[96] Another example may be a pharmacist who will not dispense medication without a prescription, with the existence of a prescription constituting the entirety of the monitoring duty but effectively denying wrongdoers access to medicines.[97] In such situations, legal rules can co-opt gatekeepers into preventing unlawful activity at a minimal cost by setting a clear and targeted legal duty.

Kraakman contrasts '*bouncer*' regimes to '*chaperone*' regimes, which he described as situations where 'gatekeepers detect and disrupt wrongdoing in an unfolding relationship with enforcement targets' (e.g., accountants who have to regularly certify a company's financial information).[98] Kraakman is sceptical of '*chaperone*' regimes because he notes that they often target wrongdoing that is far too complicated to support clear monitoring rules for gatekeepers.[99] For example, lawmakers may impose "due care" (or "due diligence") standards on gatekeepers to have them prevent wrongdoing without clearly articulating exactly what a gatekeeper must do to fulfil their monitoring duty and avoid sanction.

Imagine a bar, instead of being required to have customers show proof of age, was required to take "due care" not to serve alcohol to underage customers. Would an ID check be sufficient, or should the bar have access to government or law enforcement databases to verify these IDs and catch individuals with counterfeit IDs? This latter mechanism may have an added cost to individual privacy. Imagine also the bar can lose its liquor license or suffer penal sanctions if it unknowingly serves an underage customer. The cost of the gatekeeping regime would now include not only the bar's high "performance cost" but also the bar's "residual legal risk and

---

[95] ibid 79.

[96] ibid 63–66. Kraakman notes that these distinctions (between '*bouncer*' and '*chaperone*') are shaped by the characteristics of the ecosystem within which gatekeepers operate.

[97] ibid 63.

[98] ibid.

[99] ibid 79.

the cost of strategies to limit such risk, such as disrupting the activities of risky but innocent clients or customers."[100] What if an of-age customer forgot their ID at home, or the system did not return a match on a valid ID due to a technical error, or the individual was a foreigner and the ID could not be verified? In such situations, otherwise innocent customers would be denied service due to the bar's stringent strategies to avoid sanction necessitated by the vague gatekeeping duties imposed on it. The bar is merely seeking to avoid sanction, but the costs are imposed on innocent customers.

As Kraakman notes, "the tertiary costs of gatekeeper liability are losses that fall upon parties other than particular gatekeepers and enforcement targets," making such expansive or ambiguous gatekeeping duties undesirable.[101] The greater the risk of liability (e.g., strict liability as opposed to a negligence standard), the greater the risk of tertiary costs on innocent parties as gatekeepers will be ever more cautious.[102] Thus, a vague gatekeeping duty can even be harmful overall, imposing significant tertiary costs on innocent users as gatekeepers adopt strategies to avoid liability. For example, Rule 3(1) of the Intermediary Guidelines requires intermediaries to "make reasonable efforts to cause" their users not to upload unlawful content. This may cause intermediaries to aggressively remove content to avoid liability as part of their effort to "cause" users not to share harmful content,[103] which may result in the removal of lawful content. Such 'due care' or 'best efforts' gatekeeping regimes thus create both high performance costs on intermediaries and high tertiary costs on innocent third parties.

Where gatekeepers are placed under broadly defined duties such as "due care" or required to act "reasonably", they will likely select the monitoring activities that reduce their expected penalties or liability.[104] This may involve adopting strategies that minimise their risks but ultimately impose tertiary costs on the users. These costs should also be evaluated when examining the desirability of a particular gatekeeping regime. Kraakman also notes that "due care" or "reasonable investigation" standards merely shift the problem elsewhere, most often to be resolved by courts over several years.[105] He notes that judges, who may have their own

[100] ibid 75.

[101] ibid.

[102] ibid 76.

[103] Intermediary Guidelines, r. 3(1)(b). The intermediary "shall make reasonable efforts to cause the user of its computer resource not to host, display, upload, modify, publish, transmit, store, update or share" any content that violates Indian law.

[104] Kraakman (n 23) 76.

[105] ibid 79.

beliefs about the appropriate level of gatekeeping, can "use almost any doctrinal vehicle to make plausible guesses about the monitoring abilities and enforcement opportunities of third parties."[106] Both courts and gatekeepers are likely to make errors, and these errors are likely to be proportional to the complexity of the wrongdoing and the open-ended nature of the obligations.[107]

B. India's chaperone regime

Kraakman's observations regarding "*chaperone*" regimes have largely held true in the context of intermediary liability in India. Even before examining the nature of the obligations imposed on intermediaries, the requirement in Section 79 of the IT Act that intermediaries "observe due diligence" to retain safe harbour and the widely worded obligations in the Intermediary Guidelines demonstrate how open-ended 'due care' obligations ultimately shift debates regarding the scope of gatekeepers' monitoring duties to courts.

For example, despite the Indian Government adopting the Intermediary Guidelines, there existed judicial and scholarly disagreement over whether the Intermediary Guidelines exhaustively encapsulated the contents of intermediaries' "due diligence" obligations vis-à-vis unlawful content.[108] Courts have also independently undertaken exercises to explore the technical capabilities of intermediaries to monitor and remove unlawful content, with the aim of expanding the gatekeeping role of online intermediaries.[109] This doctrinal uncertainty and open-ended exploration regarding the scope of online intermediaries' gatekeeping duties is in part facilitated by the lack of clearly defined and narrowly targeted gatekeeping duties (though it is also aided by motivated judicial actors).

Even if we (correctly) assume that the Intermediary Guidelines encompasses the entirety of online intermediaries' gatekeeping duties, the wide-ranging list of duties imposed on intermediaries prove too complex for an intermediary to clearly determine what it must do to avoid liability. For example, to retain safe harbour, Rule 4(4) of the Intermediary Guidelines

---

[106] ibid 87.

[107] ibid 76; Ardia (n 18) 381.

[108] *Google India Pvt Ltd v Visaka Industries Ltd* 2016 SCC OnLine Hyd 393 [76] overruled by *Google India Pvt Ltd v Visaka Industries Ltd* (2020) 4 SCC 162; T Prashant Reddy, 'Back to the Drawing Board: What Should Be the New Direction of Intermediary Liability Law?' (2019) 1 NLUD Journal of Legal Studies 38, 48; Devadasan (n 43) 40.

[109] *In re: Prajwala Letter* dated 18.2.2015 SMW (Cri) 3 of 2015 (Supreme Court of India); *Antony Clement Rubin v Union of India* WP 20774 of 2018 (High Court of Madras); *Janani Krishnamurthy v Union of India* WP 20214 of 2018 (High Court of Madras); Sabu Mathew George v Union of India [2017] (2) SCC 514.

requires SSMIs to "endeavour to deploy technology-based measures, including automated tools or other mechanisms to proactively identify information" that depicts rape, child sex abuse material, or content previously-removed pursuant to a court or government order. Such an obligation specifies no bright-line threshold for when it has been satisfactorily complied with, and in practice offers little more guidance to SSMIs than a simple 'due-care' obligation. Further, by suggesting SSMIs use proactive monitoring tools, it increases the risk of tertiary costs of gatekeeping on innocent internet users. Proactive monitoring entails a greater risk of removing lawful content,[110] and if SSMIs are incentivised to use proactive monitoring tools to avoid the risk of liability, the probability of lawful content being removed increases.[111] Even though Rule 4(4) includes several provisos that aim to reduce such tertiary costs, they suffer from similar lack of clear bright-line thresholds for conduct or result, and it is unclear how they are enforced.[112] Thus, there is a real risk that SSMIs adopt proactive monitoring tools to comply with Rule 4(4) but in the absence of any bright-line rules for satisfying the requirement, the quality of such proactive tools is poor leading to the removal of lawful content.

Similarly, as discussed above, Rule 3(1)(b) of the Intermediary Guidelines stipulates that every intermediary "shall make reasonable efforts to cause the user of its computer resource not to host, display, upload, modify, publish, transmit, store, update or share any information" that violates any Indian law. It remains unclear what types of actions an intermediary must take to "cause" its users not to upload or share unlawful content, nor what threshold of efficacy it must satisfy to retain safe harbour. Placing such obligations on intermediaries to act against unlawful content is a classic example of Kraakman's "*chaperone*" regime, where "gatekeepers detect and disrupt wrongdoing in an unfolding relationship with enforcement targets."[113] Such an approach suffers from the criticism discussed above, increasing tertiary costs on innocent third parties, and ultimately shifting the determination of exactly what the gatekeeper's monitoring

[110] Daphne Keller, 'Facebook Filters, Fundamental Rights, and the CJEU's Glawischnig-Piesczek Ruling' (2020) 69 GRUR International 616, 618.

[111] *See* Dinwoodie (n 31) 46. Dinwoodie states that the acceptability of proactive monitoring measures in the European Union at present is in part due to the fact these measures are imposed as "obligations without liability", i.e., they do not suggest exposure to monetary liability. He notes that if these measures were "imposed as conditions of immunity or as essential to negate an element of secondary liability", as they are in India, they would be more problematic.

[112] Devadasan (n 43) 71–73. The provisos to Rule 4(4) require that: (1) the proactive measures taken by the intermediary be "proportionate" and have regard to "the interests of free speech" and "privacy of users"; (2) the intermediary "implement mechanisms for appropriate human oversight"; and (3) the automated tools are reviewed to evaluate the "accuracy and fairness" and "propensity of bias and discrimination."

[113] Kraakman (n 23) 63.

duties are, to the courts. The problem is compounded by the incongruence between the nature of the obligations and the enforcement mechanisms of individual lawsuits highlighted in Section 4. This situation should prompt a re-imagining of the types of obligations placed on online intermediaries under a gatekeeping liability regime (i.e., pre-conditions to safe harbour).

C. Reconceptualising safe harbour

Two guiding principles emerge regarding the *nature* of the gatekeeping duties that ought to be imposed on online intermediaries. First, as we are discussing duties enforced through individual suits against content, the duties must have a direct nexus to the content in dispute. Duties that have a direct nexus to the allegedly unlawful content are the easiest to enforce in a suit against content and narrow the scope of judicial inquiry. As highlighted above, obligations that have no nexus to the role of an intermediary in facilitating the spread of unlawful content are difficult if not impossible to impose through secondary liability suits against content (e.g., reinstatement obligations under Rule 4(8)). Likewise, courts are poorly placed to verify the continual compliance with duties that have no nexus with the disputed content (e.g., local officer requirements or transparency reporting).

This approach aligns with the inherent goal of gatekeeping liability, to leverage the gatekeeping position of online intermediaries in the online ecosystem to prevent the dissemination of unlawful content without imposing disproportionate costs on intermediaries or internet users. Intermediaries are uniquely placed to remove or disable access to content, and their gatekeeping obligations should be focussed on these capabilities. Thus, the gatekeeping duties that ought to be imposed on online intermediaries (using the promise of safe harbour) should be obligations that have a direct bearing on the removal of unlawful content.[114] This allows for the easy engagement and enforcement of these obligations in suits for unlawful content while also appropriately harnessing the ability of online intermediaries to curb unlawful content.

The second principle that emerges is that the rules for online intermediaries should clearly specify: (i) the trigger for an intermediary to initiate action; and (ii) the actions to be taken by the intermediary to avoid liability once the trigger is engaged. Kraakman notes that well framed duties encourage gatekeepers to respond to clearly identifiable misconduct without creating liability for failing to police wrongdoing that is difficult to spot.[115] This lends itself to clear, 'IF

[114] Dinwoodie (n 31) 38. Dinwoodie notes that "to some extent" the conduct required of an intermediary to avail safe harbour will *also assist it* in avoiding secondary liability under tort law.

[115] Kraakman (n 23) 79.

x DO *y*' style regimes, where both the intermediary's trigger for action and response are clearly defined and unambiguous. This is likely to be the most effective at securing both the substantive and procedural benefits of safe harbour.

Broad 'due care' or 'best efforts' mandates such as those set out in Rules 3(1)(b) or 4(4) of the Intermediary Guidelines can cause intermediaries to choose compliance strategies that will limit their risk of liability but harm innocent internet users. The more discretionary obligations imposed, the greater the residual risk for gatekeepers, and the more cautious they will be, leading to a cycle of even more aggressive monitoring and greater harm to users. Such obligations will also lead to substantial litigation as various courts coalesce on the appropriate level of gatekeeping to be employed by intermediaries on a case-by-case basis over time.

This is not to state that detailed, or complex obligations (particularly of the procedural type) cannot be imposed to avoid liability. But that they should be unambiguous and constitute bright-line legal rules so intermediaries have certainty over the steps they must undertake to avoid liability. For example, Section 512 of the Digital Millennium Copyright Act (DMCA) requires intermediaries to remove content upon receiving a notification alleging infringing content on its network.[116] The DMCA goes on to specify in a detailed manner what the contents of valid notification must include.[117] The statute also requires service providers (i.e., intermediaries) to notify the user that their content has been removed, and deliver to the rightsholder the user's counternotification (if any), at which point the content may be reinstated within fourteen days unless the rightsholder seeks a court order.[118] Rule 75 of the Indian Copyright Rules, 2013 also stipulates similar, albeit even simpler procedures. Under Rule 75, a complainant may give a written notice to the intermediary facilitating infringement. Rule 75(2) sets out the requirements for a valid notice, which must include the location of the allegedly infringing content and details establishing the complainant's ownership or license of the concerned copyright. This ensures intermediaries can easily identify the disputed content and the veracity of the claim. Upon receiving a notice, the intermediary must: (i) remove the disputed content within thirty-six hours; and (ii) display a notice explaining why the content

[116] Digital Millennium Copyright Act 17 U.S.C. [**DMCA**] s. 512(c)(1)(C).

[117] DMCA s. 512(c)(3).

[118] DMCA s. 512(g)(2). This facilitation of notice and counternotice exempts the service provider from any liability arising from the service provider's removal of the user's content.

has been restricted.[119] However, unless the complainant can secure a court order within twenty-one days, the intermediary may reinstate the content.[120]

Such systems continue to cast online intermediaries in a role more akin to "*bouncers*" than "*chaperones*".[121] This is primarily because the specific trigger for removal of content (with respect to the elements of a valid notification) have been clearly demarcated. An intermediary is not required to carry out discretionary tasks such as proactively seeking infringing activity or adjudicating the legality of content.[122] This reduces the risk of residual legal risk and tertiary costs ultimately borne by internet users. This may be contrasted to the open-ended obligations such as the one specified in Rule 3(1)(b) of the Intermediary Guidelines to "cause" users not to share unlawful content.

In fact, Jonathan Zittarian notes that where the DMCA attempted to impose more discretionary obligations, such as terminating the accounts of "repeat infringers", it was largely ineffective.[123] He notes that there was ambiguity over who was a "repeat infringer"; did the determination have to be made by a court or by the intermediary, and what was the bright-line threshold for when somebody could be adjudged to be infringing 'repeatedly'?[124] Due to the open-ended nature of these obligations, this aspect of the DMCA has largely gone unenforced,[125] vindicating Kraakman's criticism of such "*chaperone*" regimes. This is not to suggest that the DMCA is perfect or has not resulted in significant litigation or been subject to criticism.[126] But when the bright-line *procedures* in the DMCA are compared with legislative efforts that lack this clarity and instead merely task intermediaries to remove suspect classes of content, the benefits of a clear procedural approach come to the fore. Writing on the Allow States and Victims to Fight Online Sex Trafficking Act and Stop Enabling Sex Traffickers Act (FOSTA and SESTA), which firmly fall within the second category legislation, Daphne Keller

[119] The Copyright Rules, 2023 G.S.R. 172(E) dated 14 March 2013, r. 75(3)-(4).

[120] The Copyright Act, 1956 s. 52(1)(c); The Copyright Rules, 2013 G.S.R. 172(E) dated 14 March 2013, r. 75(5).

[121] Zittrain (n 43) 260.

[122] DMCA s. 512(m).

[123] DMCA s. 512(i)(1)(A); Zittrain (n 43) 268–69.

[124] ibid.

[125] ibid; Mann and Belzley (n 19) 301.

[126] Jennifer M Urban, Joe Karaganis and Brianna L Schofield, 'Notice and Takedown in Everyday Practice' (Berkley School of Law 2016) <https://illusionofmore.com/wp-content/uploads/2016/04/Berkeley_Columbia-on-512-takedown.pdf> accessed 19 April 2023; Eric Goldman, 'How the DMCA's Online Copyright Safe Harbor Failed' (2014) 3 NTUT Journal of Intellectual Property Law & Management 195.

observes, "we should expect better outcomes if OSPs [intermediaries] can claim immunity by following clear operational steps under bright-line rules, with ample public transparency. We should expect worse ones if OSPs are charged with adjudicating difficult and nuanced questions about users' speech through unaccountable processes."[127]

Applying the two principles of 'nexus with disputed content' and 'bright line triggers and procedures' can help design a targeted gatekeeping liability regime for online intermediaries. The obligations imposed on intermediaries to retain safe harbour should be bright-line duties limited to the removal of unlawful content, and any allied duties that have a direct bearing on the curbing of specific unlawful content or a court's determination of liability for the content. For example, to retain safe harbour, an intermediary may be required to remove allegedly unlawful content once notified, and a limited data retention requirement to preserve the disputed content and metadata for an adjudication of eventual liability.

In the Indian context, this would require removing obligations in the Intermediary Guidelines that are unrelated to the issue of primary or secondary liability (such as local compliance officer, transparency reporting, and proactive monitoring). Such an approach would also significantly simplify the task of courts, who would not be required to conduct wide ranging inquiries into the conduct of intermediaries every time an intermediary is sued for content. Courts could limit their attention to whether the intermediary carried out the specific actions it was required to vis-à-vis the content it is sued for. For example, the European Union's e-Commerce Directive makes a sub-classification within intermediaries and prescribes different obligations for different types of intermediaries, but all these obligations have a direct nexus to the removal of unlawful content,[128] with national legislation specifying clear procedures.[129] Lawmakers can stipulate specific procedural steps to be taken by intermediaries to retain safe harbour provided they constitute bright-line legal rules that have a direct bearing with the removal of content and adjudicating liability for the content.

A discussion on the specific trigger and consequent actions an online intermediary must take to avoid liability is beyond the scope of the present article. However, a few observations may

[127] Daphne Keller, 'SESTA and the Teachings of Intermediary Liability' (The Centre for Internet & Society - Stanford University 2017) 5 <https://cyberlaw.stanford.edu/sites/default/files/SESTA-and-IL-Keller-11-2.pdf> accessed 13 November 2022; Goldman, 'Why Section 230 Is Better Than the First Amendment' (n 26) 45.

[128] Directive 2000/31/EC of 8 June 2000 on electronic commerce, Articles 12-14.

[129] Thibault Verbiest, Gerald Spindler and Giovanni Maria Riccio, 'Study on the Liability of Internet Intermediaries' (2007) SSRN Scholarly Paper 2575069 <https://papers.ssrn.com/abstract=2575069> accessed 5 April 2023.

be made. First, the above argument concerning the *character* and *scope* of gatekeeping obligations is not prescriptive with respect to the specific actions online intermediaries must take in relation to allegedly unlawful content. For example, questions of when an intermediary is required to remove content, what the rights of the user whose content is removed are, and what rights other internet users may have in such situations can continue to be debated.

A wealth of intermediary liability literature is devoted to the appropriate balance to be struck when structuring such content removal obligations, and the Manila Principles on Intermediary Liability represents a mature consensus on best practices.[130] In the Indian context, it is sufficient to note that *Shreya Singhal vs. Union of India* interpreted the term "actual knowledge" in Section 79(3) of the IT Act to mean a court or government order directing the removal of content.[131] While this is a highly speech protective standard,[132] models such as Section 512 of the DMCA, the Manila Principles, and the practice of various European nations in implementing the e-Commerce Directive suggest that detailed procedures in this area could further protect free expression.[133] Thus, rather than seeking to regulate platform behaviour *generally*, the focus of Indian lawmakers with respect to future intermediary liability provisions should be to prescribe clear procedures for intermediaries with respect to allegedly unlawful content. The final section of this paper argues that intermediary liability and safe harbour, conceived as such, can harmoniously operate within a comprehensive framework for platform governance.

VI. SAFE HARBOUR WITHIN A PLATFORM GOVERNANCE PARADIGM

Intermediaries are no longer the entities they were when intermediary liability laws were crafted and relying solely secondary liability for content has proven inadequate to regulate the myriad of harms that may occur within complex platform ecosystems. Dinwoodie notes that determining the legitimacy of modern-day intermediary conduct requires "greater flexibility than formal secondary liability doctrine might seem to allow."[134] Even more presciently, Mann writes, "*A focus on traditional tort law notions of fault necessarily diverts attention to*

---

[130] 'Manila Principles' <https://manilaprinciples.org/index.html> accessed 5 April 2023; Aleksandra Kuczerawy, 'From "Notice and Takedown" to "Notice and Stay Down": Risks and Safeguards for Freedom of Expression' in Giancarlo Frosio (ed), *Oxford Handbook of Online Intermediary Liability* (Oxford University Press 2020).

[131] *Shreya Singhal v Union of India* [2015] (5) SCC 1.

[132] Park (n 7).

[133] Dinwoodie (n 31) 43–44; Verbiest, Spindler and Riccio (n 129).

[134] Dinwoodie (n 31) 25.

subjective normative questions of blame and responsibility, and away from the more proper focus on question of effective regulatory design."[135]

Individual lawsuits against content are inherently *ex-post*, ad-hoc, and typically only apply to one intermediary and one piece of content. This places traditional secondary liability claims as a regulatory tool at significant tension with a network infrastructure that witnesses millions of speech infractions a day, and highlights the need to regulate the systems, processes, and administration of online intermediaries, particularly large platforms.[136] Finally, as India's experience with Rule 4(8)'s notice and hearing requirement demonstrates, not all harms arise from unlawful content, and thus not all harms can be remedied by suits against content.

Lawmakers across the world are debating a whole host of measures including regulating platforms' terms of service, risk assessments for content, audits for recommender systems, complaint mechanisms, regulating online and political advertising, and formalising the role of trusted flaggers. However, as discussed in the previous section, such obligations may be incongruous with the enforcement mechanism of intermediary liability (lawsuits against unlawful content). This begs the question; how should such obligations be imposed, and how does intermediary liability fit within this paradigm of platform governance?

An examination of the European Union's Digital Services Act (DSA) is evidence that the two can co-exist within the legislative framework by clearly separating the question of liability of intermediaries for unlawful content from the types of ongoing obligations necessary to ensure online safety, transparency, and accountability. Chapter 3 of the DSA provides for the "Liability of providers of intermediary services." Articles 4-6 within this Chapter provide intermediaries with safe harbour, with Article 6 providing hosting services safe harbour subject to the removal of unlawful content pursuant to receiving "actual knowledge. Distinctly, Chapter III of the DSA specifies "Obligations for a transparent and safe online environment." Under this Chapter, intermediaries have various statutory obligations such as ensuring points of contact and legal representatives,[137] transparency reporting,[138] providing users reasoned content moderation decisions,[139] providing dispute settlement processes for content

[135] Mann and Belzley (n 19) 249.

[136] Evelyn Douek, 'Content Moderation as Systems Thinking' (2022) 136 Harvard Law Review 526.

[137] Digital Services Act, Regulation (EU) 2022/2065, 19 October 2022 [**Digital Services Act**], Article 12-13.

[138] Digital Services Act, Article 15.

[139] Digital Services Act, Article 17.

moderation,[140] and regulating online advertisements and recommender systems.[141] Crucially, these obligations are imposed as distinct statutory obligations and not as pre-conditions to safe harbour. As Recital 41 of the DSA states, these obligations "*are independent from the question of liability of providers of intermediary services which need therefore to be assessed separately*."

Imposing such platform governance obligations as distinct statutory duties (or as "positive" obligations) on intermediaries has several distinct advantages. First, it preserves the statutory immunity granted to intermediaries for hosting and transmitting third-party content. The Intermediary Guidelines are demonstrative of how safe harbour can be narrowed through the imposition of complex regulatory obligations through the vehicle of intermediary liability, ultimately harming free expression by incentivising intermediaries to moderate aggressively. Thus, eliminating such obligations as pre-conditions to safe harbour protects free expression and preserving the current ecosystem built around the generation and transmission of third-party content. The shift from a mediated public sphere to one where users are central has had several positive democratic effects, and safe harbour is a key guarantor of these benefits.[142] Unlike a loss of safe harbour, which may impose uncertain (potentially crushing) monetary penalties, a violation of statutory duties entails suffering statutorily pre-defined fines. This mitigates the concern that intermediaries will 'err on the side of caution' and suppress lawful content to avoid legal risk.

Second, demarcating such regulatory-style obligations from discussions on intermediary liability and safe harbour allows for a focus on developing clear, bright-line gatekeeping duties that can be placed on intermediaries using the tool of liability. Intermediaries are uniquely placed to curb the spread of unlawful content, but the development of clear rules that have a direct nexus to the removal of content and the adjudication of liability is necessary for a gatekeeping liability regime for online intermediaries to be justifiable. The adoption of new legislation replacing the Information Technology Act is an opportunity to further nuance the "actual knowledge" standard and takedown procedures, and as Mann observes, it should proceed independent of broader discussions surrounding the accountability of online platforms.

[140] Digital Services Act, Article 21.

[141] Digital Services Act, Article 27, 28.

[142] Balkin (n 6) 2313; Goldman, 'Why Section 230 Is Better Than the First Amendment' (n 26) 33–34. *See supra,* Section 2(B).

This should also contribute to greater doctrinal certainty amongst courts, who will not have to devise standards of what constitutes appropriate gatekeeping duties.

Third, imposing such platform governance style obligations as distinct statutory duties (as opposed to pre-conditions to safe harbour) allows these obligations to be independently enforced absent a suit alleging the hosting of unlawful content. Section 3 demonstrated how, because several of these obligations have no nexus with actual unlawful content (e.g., dispute settlement obligations or transparency reporting), they are hard to engage and enforce through suits for secondary liability. By requiring intermediaries to adhere to such obligations by way of independent statutory obligations, they can directly be enforced even absent a suit for secondary liability. While such obligations will undoubtedly require some form of enforcement mechanism, potentially through an independent regulator or complaint mechanism, any regulatory body tasked with such enforcement will be better placed to ensure ongoing, uniform, and industry-wide compliance with such obligations as opposed to courts adjudicating individuated cases of unlawful content.

Fourth, separating question of liability and regulatory-style obligations allows for a simplification of the Gordian knot that is the classification of intermediaries. As discussed in Section 3, for questions of liability, the question of 'what is an intermediary' is best examined through the lens of *functionality* vis-à-vis the content the entity is sued for. However, when it comes to questions of platform governance obligations such as transparency reporting or appointing a local officer, a functional approach does not work; such obligations must be imposed on specific entities. Because such obligations are monitored and enforced on an ongoing basis, the definition of intermediaries for such regulatory-style obligations must clearly identify the subjects of such regulation. For example, from the perspective of liability, every "intermediary" (i.e., an entity that performs the function of an intermediary with respect to third party content) would be eligible for safe harbour. However, only certain types of intermediaries (e.g., "online platforms" or "marketplaces") would be required to comply with statutory duties such as transparency reporting or providing hearings.

By separating the question of liability from questions of these platform governance obligations, the definitional architecture can be separated. With respect to intermediary liability, the definition can remain functional and be interpreted by courts in relation to specific unlawful content. Separately, definitions for specific entities (e.g., "online platforms" or "marketplaces") can be utilised to impose independent statutory obligations such as transparency reporting or

local officer requirements. For example, the Intermediary Guidelines define a "significant social media intermediary" as an intermediary which primarily or solely enables online interaction between users and has more than five million users.[143] This is ultimately a definition that applies to specific entities and is not interpreted based on functions vis-à-vis content.

Ensuring that: (i) the term "intermediary" is interpreted functionally when issues of *liability* arise; and (ii) all "intermediaries" are eligible for safe harbour, then leaves lawmakers open to craft distinct, entity specific definitions to impose discrete platform governance obligations without undermining safe harbour or comprising their regulatory goals. This does not hamper lawmakers from imposing specific obligations on entities that broadly perform the role of intermediaries outside of the issue of liability. For example, the DSA defines "intermediary services" as conduits, caching, and hosting services[144] and imposes certain minimal obligations on them, such as designating points of contact for communication and informing users when their terms of service change significantly.[145] However, crucially, these conditions are not imposed as pre-requisites to safe harbour and are set out in a separate Chapter of the legislation, distinct from the issue of liability.[146] This is then further distinguished from the issue of "online platforms" which are defined separately as a sub-set of hosting providers and have their own set of obligations.[147]

This by no means exhausts the question of how to define intermediaries. There are still many thorny questions about how to categorise and define different types of intermediaries, with respect to the statutory duties, lawmakers may seek to impose. However, the above arguments demonstrate that the issue of defining intermediaries must be congruent with not only the type of obligation imposed on them, but also the enforcement mechanism. Where the enforcement mechanism is that of suits for secondary liability, a functional approach (as discussed in Section 3) is beneficial. However, where the enforcement mechanism is a regulator imposing

[143] Intermediary Guidelines, r. 2(1)(v); Ministry of Electronics and Information Technology, Notification S.O. 942(E) dated 25 February 2021 setting the threshold of users to qualify as an SSMI at 5 million.

[144] Digital Services Act, Article 3(g).

[145] Digital Services Act, Article 11(1), 14(2).

[146] Digital Services Act, Chapter II (Liability of Providers of Intermediary Services); Digital Services Act, Chapter III, Section 1 (Provisions applicable to all providers of intermediary services).

[147] Digital Services Act, Article 3(i), Chapter III, Section 2 (Additional provisions to providers of hosting services, including online platforms).

independent statutory duties, a definition that focuses on entities as opposed to functions may be suitable.

VII. CONCLUSION

When presenting a roadmap for India's replacement to the Information Technology Act, 2000, India's Minister of State for Information and Technology noted that it was open to debate whether certain intermediaries would be eligible for safe harbour at all.[148] This is cause for significant concern given that statutory immunities for intermediaries have been key protections for free speech on the internet and have given rise to the modern day internet as we know it.[149] They allow intermediaries, including online platforms to host and transmit user-generated content, giving voice to millions of internet users, without saddling them with the debilitating risk of liability that that would cause them to remove content that bears a remote risk of unlawfulness.[150] This is especially crucial for unpopular content that may shock, offend, and disturb, but is nonetheless lawful and often of vital public importance.

The present article sought to demonstrate that safe harbours (for all intermediaries) are not incompatible with a comprehensive statutory framework necessary to guarantee the accountability of platforms. However, such compatibility hinges on a proper appreciation of how liability rules (pre-conditions for safe harbour) operate and should be structured, and why they should be separated from other continual and complex duties that platform governance envisages. Therefore, this article has argued that it is important for legislators and courts to properly appreciate how the functional definition of an 'intermediary' and safe harbour operates vis-à-vis specific pieces of content. Once this understanding is set out, the incongruity between the types of obligations imposed by the Intermediary Guidelines and the enforcement mechanism of intermediary liability comes to the fore, suggesting the need to re-design India's safe harbour for the age of platform governance.

The principles that govern a revised safe harbour regime must be informed by the enforcement mechanism of intermediary liability. The vague "due diligence" standard and open-ended obligations currently part of the Intermediary Guidelines must be dispensed with. Such complex obligations must be replaced with clear, unambiguous bright-line rules that have a

[148] Ganesan (n 2); 'Govt Rethinking "Safe Harbour" in Digital India Bill: How This Could Change Internet Landscape' (n 3).

[149] Balkin (n 6) 2313; Klonick (n 5) 1604.

[150] Balkin (n 6); Goldman, 'Why Section 230 Is Better Than the First Amendment' (n 26).

direct nexus to the issue of liability for the unlawful content in dispute. Regulatory design that acknowledges differential enforcement mechanisms dictates that the proper place for continuing, regulatory-style obligations such as local officer requirements, transparency reporting, and dispute resolution, is through independent statutory duties distinct from questions of liability. Crucially, the imposition of these obligations does not need to result in a dilution of safe harbour. Indeed, by permitting all intermediaries to avail of safe harbour, and then separately imposing regulatory-style obligations in a targeted manner as distinct statutory obligations, the IT Act's replacement could both strengthen and expand safe harbour, while securing greater accountability from platforms.