The Dialogue™
INFORM ENGAGE IDEATE

WHITE PAPER

ENABLING DIGITAL RIGHTS AND SAFETY THROUGH SAFE HARBOUR

DIGITAL INDIA BILL SERIES | PART 1

WHITE PAPER

ENABLING DIGITAL RIGHTS AND SAFETY THROUGH SAFE HARBOUR

DIGITAL INDIA BILL SERIES | PART 1

Authors: Shruti Shreya & Garima Saxena
Copy editor: Akriti Jayant
Designer: Shivam Kulshreshtha

The Dialogue™ is a public policy think tank with a vision to drive a progressive narrative in India's policy discouse. Founded in 2017, we believe in facilitating well-researched policy debates at various levels to help develop a more informed citizenry, on areas around technology and development issues. The Dialogue™ has been ranked as the world's Top 10 think tanks to watch out for, by the Think Tank and Civil Societies Programme (TTCSP), University of Pennsylvania in their 2020 and 2021 rankings.

For more information
https://thedialogue.co

Suggested Citation
Shreya S. & Saxena G. (2023). *Enabling Digital Rights and Safety Through Safe Harbour: Digital India Bill Series Part 1*. New Delhi. The Dialogue™.

Catalogue No.
TD/PR/WP/1123/08

Publication Date
November 30, 2023

Acknowledgement

The authors express their sincere appreciation to Mr. Kazim Rizvi and Mr. Kamesh Shekar for their insightful comments and thorough review of the paper. Additionally, heartfelt thanks are extended to Ms. Akriti Jayant for her invaluable assistance in copy-editing and to Mr. Shivam Kulshrestha for his skillful design contributions to the paper.

Contents

ABBREVIATIONS

BOSEs	Basic Online Safety Expectations
CDA	Communications Decency Act
CSAM	Child Sexual Abuse Material
DSA	Digital Services Act
EARN IT Act	Eliminating Abusive and Rampant Neglect of Interactive Technologies Act
EU	Europe Union
IT Act, 2000	Information Technology Act, 2000
IT Rules, 2021	Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021
LAED Act	Lawful Access to Encrypted Data Act
LEAs	Law Enforcement Agencies
NCMEC	National Center for Missing & Exploited Children
NetzDG	Network Enforcement Act
OSA 2023	Online Safety Act 2023
SESTA-FOSTA	The Stop Enabling Sex Traffickers Act and the Fight Online Sex Trafficking Act
SMEs	Small and Medium Enterprises
U.K.	United Kingdom
UNICEF	United Nations Children's Fund
U.S.A.	United States of America
VLOPs	Very large online platforms
VLOSEs	Very large search engines

Executive Summary

Context and Scope

This white paper is developed amidst the ongoing consultations to introduce a new Digital India Act in India, re-evaluating the scope of safe harbour immunity. Safe harbour protects online intermediaries from liability for user-generated content, underpinning the democratic essence of the Internet.

However, the proliferation of online harms has triggered global regulatory shifts, causing intermediaries to assume greater due diligence to maintain their safe harbour immunity. While aimed at enhancing online safety, this trend risks diluting the very foundation of safe harbour, potentially curtailing user freedoms and stifling digital innovation. The paper scrutinises such regulatory approaches across various jurisdictions, including the EU, USA, UK, and Australia, noting the adverse implications for digital rights and economic growth.

In light of these international experiences, the paper posits that increasing due diligence requirements for safe harbour is counterproductive. It compromises user privacy and speech, disproportionately affects marginalised groups, and places undue burdens on intermediaries. Instead, the paper recommends a nuanced, evidence-based policy that respects constitutional rights while addressing the enforcement gaps and complexities of online content regulation.

Key Policy Recommendations for the Upcoming Digital India Bill

The paper urges India's policymakers to consider these recommendations while formulating the new IT Law, to foster an internet that is both safe and free, supporting democratic values and encouraging digital innovation.

- **Maintain Actual Knowledge Definition:** Align with the *Shreya Singhal* judgement to safeguard intermediaries from excessive liability.
- **Establish Clear Legal Framework:** Define clear and precise rules for intermediaries, allowing action against illegal content through lawful orders to protect due process.
- **Implement Good Samaritan Principles:** Protect intermediaries from punitive actions when they take voluntary measures in good faith to combat illegal content.
- **Regulate Platform Operations:** Focus on regulating platform procedures to promote transparency and ensure user rights are respected.
- **Promote Collaborative Engagement:** Encourage joint efforts among platforms, users, civil society and the government to build a robust defence against online harms while upholding digital rights.

1 Background

The Internet, a symbol of seamless global connectivity, undoubtedly enhances our daily lives. Today, users can access a vast repository of human knowledge, stay updated with real-time news, connect with others, and express themselves with a single click. The Internet's essence lies in its immediacy and democratisation, symbolising a transformative shift from the days of physical libraries and yesterday's newspaper editions.[1]

One of the foundations of this digital revolution is the emergence and evolution of intermediary platforms. These platforms function as the gears of the Internet machinery, facilitating diverse interactions and services. Internet Service Providers ensure our connectivity, search engines assist in navigating the web, social media platforms enable us to engage, and e-commerce websites facilitate trade. Each intermediary fulfils a distinct role. A particularly critical segment among these intermediaries is the platforms that host user-generated content. Platforms like YouTube, Facebook, and Twitter democratise content creation, ensuring they empower every voice.[2] They stand as proof of the transformative power of free speech on the Internet, allowing millions to share ideas, build communities, and advocate for change.

However, the democratisation of content creation raises the question: Who bears responsibility if the content is illegal or harmful? This is where the principle of 'Safe Harbour' comes into play. Designed as a protective shield, it insulates intermediaries from liability for user-generated content based on the premise that they act as mere conduits without actively controlling the content. The intention is clear: to foster an environment where freedom of speech thrives without overburdening platforms that are primarily facilitators.

In India, Section 79 of the Information Technology Act, 2000 (IT Act, 2000) embodies this safe harbour principle, providing intermediaries with immunity from third-party content liability in the absence of 'actual knowledge' regarding its illegality.[3] According to the pronouncement in the Shreya Singhal judgment, this actual knowledge must come in the form of a court order or notification by the appropriate government.[4] Furthermore, such requests must align with the limitations listed under Article 19(2) of the Constitution, which provides the conditions for reasonable restrictions on the fundamental right to freedom of speech and expression.[5]

Over the last two decades of India's platform regulation experience, safe harbour has been the key facilitator of all our internet freedoms and the digital economy, enabling more

1. Komaitis, K. (2023). The democratic Nature of the Internet's Infrastructure. https://doi.org/10.31752/idea.2023.35
2. Nadeem, R. (2022, September 15). Experts on the Future of Democracy at A Time of Digital Disruption. Pew Research Center. Retrieved 10 September 2023, from https://www.pewresearch.org/internet/2020/02/21/broader-thoughts-from-key-experts-on-the-future-of-democracy-at-a-time-of-digital-disruption/
3. Information Technology Act, 2000, Section 79.
4. Shreya Singhal v. Union Of India, AIR 2015 SC 1523.
5. The Constitution of India, 1950, Art. 19(2).

meaningful conversations,[6] cultural exchanges,[7] and improved economic empowerment opportunities.[8] Nevertheless, as in the physical world, the Internet also confronts its own set of challenges. The increased popularity and usability of intermediary platforms have transformed them into a force for social good, but they have also been more frequently misused by malicious hackers and other cybercriminals.[9] This has led to the proliferation of online harms, including child pornography, misinformation, radicalising content, and online gender-based abuse. These challenges pose threats to user safety and security and have compelled authorities across the globe to contemplate mechanisms for controlling online vices.[10]

One major proposal has been to increase the due diligence requirements that intermediaries must fulfil to enjoy safe harbour protection. This proposal arises from the rationale that intermediaries should work more proactively to tackle the harms on their platforms. This discussion is even more critical considering that the rationale behind imposing a minimum level of due diligence requirements on intermediaries is to ensure that private companies do not alter user-generated speech.[11] Increasing the due diligence requirements of intermediaries concerning user-generated content on their platform would lead them to exert over the content.[12]

What are the most sustainable mechanisms through which intermediaries can be motivated to put more effort into enhancing user safety on their platforms? Is increasing due diligence requirements and diluting the safe harbour the right approach to make intermediaries more responsive and tackle online harms? What lessons can we draw from the experiences of other countries in this regard? These questions must be addressed to determine the best way to regulate the Internet and digital platforms.

6. Shreya, S., & Tiwari, P. (2020, December). Analysing the American Safe Harbour Regime: Takeaways for India. The Dialogue. Retrieved 10 September 2023, from https://thedialogue.co/wp-content/uploads/2020/12/Analysing-the-American-Safe-Harbour-Regime_Takeaways-for-India_The-Dialogue.pdf

7. Walko, P. (2022, November 29). Internet access as a tool for boosting economic and social equality. LSE International Development. Retrieved 10 September 2023, from https://blogs.lse.ac.uk/internationaldevelopment/2022/11/29/internet-access-as-a-tool-for-boosting-economic-and-social-equality/

8. Hoboken, J. van, & Keller, D. (2019). Design Principles for Intermediary Liability Laws. Transatlantic High Level Working Group on Content Moderation Online and Freedom of Expression. Retrieved 10 September 2023, from https://www.ivir.nl/publicaties/download/Intermediary_liability_Oct_2019.pdf

9. Li, Y., & Liu, Q. (2021). A comprehensive review study of cyber-attacks and cyber security; Emerging trends and recent developments. Energy Reports, 7, 8176–8186. Retrieved 10 September 2023, from https://doi.org/10.1016/j.egyr.2021.08.126

10. Schmon , C., & Pedersen, H. (2022, July 19). Platform liability trends around the globe: Recent noteworthy developments. Electronic Frontier Foundation. Retrieved 10 September 2023, from https://www.eff.org/deeplinks/2022/05/platform-liability-trends-around-globe-recent-noteworthy-developments

11. Jain, K. (2022, July 16). Safe harbour protection for social media not so safe for users - The Sunday Guardian Live. Retrieved 10 September 2023, from https://sundayguardianlive.com/legally-speaking/safe-harbour-protection-social-media-not-safe-users

12. Shreya, S. & Tiwari, P. (2020, December). Analysing the American Safe Harbour Regime: Takeaways for India. The Dialogue. Retrieved 10 September 2023, from https://thedialogue.co/wp-content/uploads/2020/12/Analysing-the-American-Safe-Harbour-Regime_Takeaways-for-India_The-Dialogue.pdf

Figure 1 The Cornerstone of Digital Freedom in India

Section 79 of the IT Act builds a free speech enabling open internet by preventing blogs, websites and forums from being held liable for the speech of their users.

SECTION 79 OF IT ACT, 2000

THE CORNERSTONE OF DIGITAL FREEDOM IN INDIA

Intermediaries are protected from liability provided they meet the conditions laid down under Section 79 and the additional due diligence requirements prescribed under the Rules made therein.

What if there was no Section 79?

E-commerce websites could have been held liable for user reviews.

Intermediaries might have felt pressure to limit real-time posts in order to review them to avoid liability.

(Un) Ease of Doing Business

LITIGATION COSTS

Intermediaries, including startups, would have faced increased litigation costs as they would have to spend more time and resources proving that they are not liable for third-party content.

MONETARY DAMAGES

Intermediaries would have also faced monetary fines in case of their inability to detect and takedown illegal third-party content.

Overmoderation

Content Blocking

ISPs would have engaged in excessive content-blocking to avoid legal liability.

Suppressing Creative Expression

AI moderation tools may not have been able to appreciate artistic nuances, satire, and art representing societal taboos and commentaries.

Impact on Vulnerable Users

Overmoderation would have disproportionately affected vulnerable users like sexual assault survivors and anti trafficking advocate who relied on these platforms for support.

Who would be affected?

Free Speech of **692 million Internet** users in India.

Innovation and growth of over **25,000 tech companies.**

Foreign investments of **over $1.82 billion**.

2 Safe Harbour Trends Around the World: Global and Domestic Perspectives

Over the years, we have observed a rise in online harms, prompting jurisdictions worldwide to impose stricter due diligence requirements on intermediaries. This results in a narrowing of the safe harbour protection as intermediaries must now do more to enjoy this immunity. Although well-intentioned, these additional due diligence requirements have raised questions from noted policy and technical experts.[13] They have led to the consequent dilution of safe harbour protection, the alteration of the fundamental definition of an intermediary, and an unintended but disproportionate impact on user rights and internet freedom.[14]

This chapter examines the evolving online regulations introduced in other prominent jurisdictions, including the European Union (E.U.), the United Kingdom (U.K.), Australia, and the United States of America (U.S.A), to understand how other jurisdictions are regulating intermediaries, how new types of due diligence requirements are affecting the scope of safe harbour protection in these countries, and how effective legislative efforts have been in enhancing online safety.

The analysis of global jurisprudence is followed by a brief review of the evolution of the intermediary liability regime in India and recent developments, such as the promulgation of the IT Rules, 2021, and the ongoing deliberations surrounding the introduction of a new Digital India Act to replace the existing IT Act.

2.1 Experience from other jurisdictions

2.1.1 Europe Union (EU)

Existing Regulations: The EU's Platform Regulation regime derives its key principles from the E-commerce Directive adopted in 2000.[15] The Directive aimed to promote the development of electronic commerce in the EU and ensure the free movement of information society services across Europe.[16] It extended safe harbour protection to online platforms when their activity was merely technical, automatic, or passive in nature.[17]

[13] Johnson, A., & Castro, D. (2021, February 22). How Other Countries Have Dealt with Intermediary Liability. Information Technology and Innovation Foundation. Retrieved 10 September 2023, from https://itif.org/publications/2021/02/22/how-other-countries-have-dealt-intermediary-liability/

[14] Shreya, S. & Tiwari, P. (2020, December). Analysing the American Safe Harbour Regime: Takeaways for India. The Dialogue. Retrieved 10 September 2023, from https://thedialogue.co/wp-content/uploads/2020/12/Analysing-the-American-Safe-Harbour-Regime_Takeaways-for-India_The-Dialogue.pdf

[15] E-Commerce Directive. (2023, February 1). Retrieved 10 September 2023, from https://digital-strategy.ec.europa.eu/en/policies/e-commerce-directive

[16] E-Commerce Directive. (2023, February 1). Retrieved 10 September 2023, from https://digital-strategy.ec.europa.eu/en/policies/e-commerce-directive

[17] Madiega, T.(2020, May). Reform of the EU liability regime for online intermediaries. European Parliamentary Research Service. Retrieved 10 September 2023, from https://www.europarl.europa.eu/RegData/etudes/IDAN/2020/649404/EPRS_IDA(2020)649404_EN.pdf

However, in response to the rapid increase in illegal content and the global trend to enhance due diligence requirements for intermediaries, the EU Council approved the Digital Services Act (DSA) last year. The DSA is part of a broader initiative to revamp the platform regulation regime in Europe. Its goal is to establish a secure digital environment that protects the fundamental rights of users and ensures a level playing field for businesses.[18] The law categorises intermediaries into different types, including mere conduits, caching, hosting, and online platforms. Furthermore, it introduces the classification of very large online platforms (VLOPs) and very large search engines (VLOSEs) based on the number of users. VLOPs and VLOSEs refer to online platforms or search engines with more than 45 million average monthly active users in the EU.[19]

The core ideas and legal concepts, including the fundamental safe harbour principle established under the E-Commerce Directive,[20] remain unchanged. The DSA maintains that there is no general obligation to monitor illegal or infringing content or activity. Article 6 provides safe harbour protection to hosting services against illegal third-party-generated information if they lack awareness or actual knowledge of such content and promptly remove it upon becoming aware of its presence.[21] Article 7 introduces an exemption to the actual knowledge or awareness requirement.[22] The 'Good Samaritan' clause under Article 7 of the DSA shields intermediaries from voluntary own-initiative investigations against illegal content and does not disqualify them from safe harbour protection.[23] The Good Samaritan protections are available only if undertaken in good faith and with due diligence.[24] Recital 26 offers some guidance, emphasising that 'the mere fact that providers undertake such activities does not lead to the unavailability of the exemptions from liability' if these activities are carried out in good faith and diligently.[25] This pertains to activities conducted to comply with EU law, including those outlined in the DSA and the platform's terms and conditions.

Impact on Safe Harbour: Over all the DSA has been one of the most progressive legislations[26] which preserves the essence of safe harbour immunity.[27] However, certain

[18] The Digital Services Act Package. (2023, September 6). Retrieved 10 September 2023, from https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package
[19] European Commission. DSA: Very large online platforms and search engines. Retrieved 10 September 2023, from https://digital-strategy.ec.europa.eu/en/policies/dsa-vlops
[20] E-Commerce Directive. (2023, February 1). Retrieved 10 September 2023, from https://digital-strategy.ec.europa.eu/en/policies/e-commerce-directive
[21] Digital Decade by Hannes Snellman. (2023, March 2). Recitals - Digital Decade by Hannes Snellman. Retrieved 10 September 2023, from https://digitaldecade.hannessnellman.com/digital-service-package/digital-services-act-dsa/recitals
[22] Digital Decade by Hannes Snellman. (2023, March 2). Recitals - Digital Decade by Hannes Snellman.Retrieved 10 September 2023, from https://digitaldecade.hannessnellman.com/digital-service-package/digital-services-act-dsa/recitals
[23] La Rosa, A., & Mazzilli, M. G. (2022). DSA: the European 'good samaritan' rule and the 'safe harbour' regime. Lexology. Retrieved 10 September 2023, from https://www.lexology.com/library/detail.aspx?g=889e875c-9f43-477d-8b3b-7d17fa0e3499
[24] Digital Decade by Hannes Snellman. (2023, March 2). Recitals - Digital Decade by Hannes Snellman. Retrieved 10 September 2023, from https://digitaldecade.hannessnellman.com/digital-service-package/digital-services-act-dsa/recitals
[25] Digital Decade by Hannes Snellman. (2023, March 2). Recitals - Digital Decade by Hannes Snellman. Retrieved 10 September 2023, from https://digitaldecade.hannessnellman.com/digital-service-package/digital-services-act-dsa/recitals
[26] Madiega, T. (2020, May). Reform of the EU liability regime for online intermediaries. European Parliamentary Research Service. Retrieved 10 September 2023, from https://www.europarl.europa.eu/RegData/etudes/IDAN/2020/649404/EPRS_IDA(2020)649404_EN.pdf
[27] Kuczerawy, A. (2021). The Good Samaritan that wasn't: voluntary monitoring under the (draft) Digital Services Act. Verfassungsblog: On Matters Constitutional. Retrieved 10 September 2023, from https://doi.org/10.17176/20210112-181758-0

aspects have led to implementational[28] and compliance[29] concerns as highlighted by experts[30]

To illustrate, the DSA introduces certain other burdensome and potentially unworkable user redress requirements, including appeals for restrictions of visibility,[31] demotions,[32] an expansion of appeals to user-flagged content[33] where the online platform chooses not to remove content, and the availability of out-of-court redress. This shift from a systemic approach to a more individualised content moderation decision-making process can pose increased risks and challenges, especially considering the potential to overwhelm the review teams. For instance, if content moderation decisions are made on a case-by-case basis on a social media platform that receives thousands of complaints daily, it may affect the moderation team's resources and efficiency. The DSA also outlines expansive data access requirements, including law enforcement access,[34] without stringent safeguards to protect user data and privacy. The DSA also leaves important questions unanswered about protecting personal data accessed by researchers.[35]

In another significant move, the DSA requires intermediaries to publicly disclose information on how they deploy automated content moderation tools and the error rates of the tools used. Transparency in the way user-generated content is processed is one of the foundations of ensuring openness, trust, and accountability. However, it's important to assess the technical and legal feasibility of such additional due diligence requirements before introducing them as a necessary obligation for enjoying safe harbour. Algorithms are commercially valuable and are often held as trade secrets, making their disclosure contentious.[36] Moreover, public access to these algorithms may end up providing hackers and other malicious actors with the tools to bypass the algorithms and misuse the platform, leading to an increase in the online harms that we aim to curtail.[37]

[28] Digital Decade by Hannes Snellman. (2023, March 2). Recitals. Digital Decade by Hannes Snellman.Retrieved 10 September 2023, from https://digitaldecade.hannessnellman.com/digital-service-package/digital-services-act-dsa/recitals
[29] Digital Decade by Hannes Snellman. (2023, March 2). Recitals. Digital Decade by Hannes Snellman. Retrieved 10 September 2023, from https://digitaldecade.hannessnellman.com/digital-service-package/digital-services-act-dsa/recitals
[30] Digital Decade by Hannes Snellman. (2023, March 2). Recitals. Digital Decade by Hannes Snellman. Retrieved 10 September 2023, from https://digitaldecade.hannessnellman.com/digital-service-package/digital-services-act-dsa/recitals
[31] Barata, J., Budzinski, O., Cole, M., Ledger, M., McGonagle, T., Pentney, K., Rosati, E., de Streel, A. IRIS Special 2021-01 DSA. (2021, October). Council of Europe. Retrieved September 10, 2023, from https://rm.coe.int/iris-special-2021-01en-dsa-package/1680a43e45
[32] Barata, J., Budzinski, O., Cole, M., Ledger, M., McGonagle, T., Pentney, K., Rosati, E., de Streel, A. IRIS Special 2021-01 DSA. (2021, October). Council of Europe. Retrieved September 10, 2023, from https://rm.coe.int/iris-special-2021-01en-dsa-package/1680a43e45
[33] Barata, J., Budzinski, O., Cole, M., Ledger, M., McGonagle, T., Pentney, K., Rosati, E., de Streel, A. IRIS Special 2021-01 DSA. (2021, October). Council of Europe. Retrieved September 10, 2023, from https://rm.coe.int/iris-special-2021-01en-dsa-package/1680a43e45
[34] Digital Decade by Hannes Snellman. (2023, March 2). Recitals. Digital Decade by Hannes Snellman. Retrieved 10 September 2023, from https://digitaldecade.hannessnellman.com/digital-service-package/digital-services-act-dsa/recitals
[34] EU's Digital Services Act Just Became Applicable: Outlining Ten Key Areas of Interplay with the GDPR. Future of Privacy Forum. Retrieved 10 September 2023, from https://fpf.org/blog/eus-digital-services-act-just-became-applicable-outlining-ten-key-areas-of-interplay-with-the-gdpr/
[36] Lath, A. L. (2022, June). Algorithmic transparency and the smart state. Spicyip. Retrieved 10 September 2023, from https://spicyip.com/2022/06/algorithmic-transparency-and-the-smart-state.html
[37] Lath, A. L. (2022, June). Algorithmic transparency and the smart state. Spicyip. Retrieved 10 September 2023, from https://spicyip.com/2022/06/algorithmic-transparency-and-the-smart-state.html

2.1.2 United States (U.S.A.)

Existing Regulations: The Communications Decency Act (CDA), passed by the United States Congress in 1996, provides the origin of the safe harbour for online platforms. The provision stipulated that, with some exceptions, online service providers are not liable for content posted by their users: "No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider." In recent years, legislators have actively worked to more clearly limit the protection under Section 230 CDA through the passage of several legislations, including the SESTA-FOSTA (The Stop Enabling Sex Traffickers Act and the Fight Online Sex Trafficking Act) laws, the proposed EARN IT (Eliminating Abusive and Rampant Neglect of Interactive Technologies), and LAED Act (Lawful Access to Encrypted Data Act).

Due Diligence Requirements: The SESTA-FOSTA legislations establish an exception to Section 230 by rendering online platforms accountable for third-party advertisements related to prostitution, which includes consensual sex work, on their websites. Consequently, intermediaries must ensure that they do not host advertisements related to prostitution, including consensual sex work, on their websites. The Acts are framed so expansively that they could potentially be applied against platform owners, even in cases where they lack 'actual knowledge' of their platform being used for trafficking activities.[38]

In another significant development, the proposed EARN IT Act empowers government officials to compel access to private communications by weakening encryption, enabling them to take action against illegal activities.[39] Similarly, the LAED Act allows law enforcement agencies to request access to encrypted data from intermediaries. This, for all practical purposes, would entail breaking encryption and would constitute a violation of the First Amendment and Fourteenth Amendment rights.[40]

Impact on Safe Harbour: The SESTA-FOSTA legislations have signified a significant departure from the original safe harbour jurisprudence. Essentially, this means that intermediaries must proactively identify and remove prohibited advertisements from their platform, in contrast to earlier times when they were only expected to act upon receiving actual knowledge. During the debate over the Bill in the Senate, the proponents of FOSTA-SESTA provided little to no evidence that increased platform liability would help restrict trafficking. However, the opponents of the Act presented a plethora of arguments, suggesting that shutting down platforms that advertise sexual services exposes trafficking victims to greater danger. In fact, evidence suggests that this proactive approach has not improved online safety. Five years after their

[38] What is Sesta/Fosta?. Decriminalize Sex Work. (2023, June). Retrieved 10 September 2023, from https://decriminalizesex.work/advocacy/sesta-fosta/what-is-sesta-fosta/

[39] EARN IT Act, 2023. congress.gov. (2023, April). Retrieved 10 September 2023, from https://www.congress.gov/bill/118th-congress/house-bill/2732/text

[40] Shreya, S. & Tiwari, P. (2020, December). Analysing the American Safe Harbour Regime: Takeaways for India. The Dialogue. Retrieved 10 September 2023, from https://thedialogue.co/wp-content/uploads/2020/12/Analysing-the-American-Safe-Harbour-Regime_Takeaways-for-India_The-Dialogue.pdf

enactment, numerous reports[41] emerged, suggesting that these laws have further endangered the lives of sex workers, making it more challenging to gather evidence to investigate and prosecute traffickers. Moreover, ample evidence, both anecdotal[42] and researched, has emerged, suggesting that providing sex workers with a way to advertise, vet, and choose clients online makes them much safer[43] than they are without an online system. When they are forced onto the streets to find clients, they have fewer advanced safety precautions in place, no ability to effectively pre-screen clients, and no way to ensure that they work in safe and secure locations.

The EARN IT Act targets the safe harbour enjoyed by intermediaries by mandating that they would not automatically be exempt from liability against content related to CSAM but will have to 'earn it.'[44] Under the amended version of the law, the Attorney General has the power to notify a 'broad category' of best practices for unrestricted regulation of the platform's editorial activities, amounting to a violation of the First Amendment. The law also allows for the selective removal of Section 230 immunity for CSAM content and creates an unconstitutional condition that violates the First Amendment. Similarly, the LAED Act restricts good-faith providers from using encryption on their platforms without considering that savvy bad actors will shift to other or their own encrypted platforms. The creation of backdoors also renders the platforms vulnerable to foreign surveillance. While it will be harder to catch savvy criminals, law-abiding citizens will be left susceptible to cyber vulnerabilities in the digital age, in which the right to privacy is held to be a part of the right to life.[45]

2.1.3. United Kingdom (U.K.)

Proposed Regulation: First proposed in 2017, the UK government passed the Online Safety Act (OSA) 2023 on October 26, 2023, with the stated intent to make "the UK the safest place in the world to be online".[46] The new regulatory regime takes a different approach to regulating the Internet by imposing legal responsibility on service providers to keep the Internet safe for children and give adults more choice over what they see online.[47] The Act

41. Grant, M. G. (2021, June 23). The Real Story of the Bipartisan Anti–Sex Trafficking Bill That Failed Miserably on Its Own Terms. The Soapbox. Retrieved 10 September 2023, from https://newrepublic.com/article/162823/sex-trafficking-sex-work-sesta-fosta
42. Williams, T. (2017, March 11). Backpage's Sex Ads Are Gone. Child Trafficking? Hardly. New York Times. Retrieved 10 September 2023 https://www.nytimes.com/2017/03/11/us/backpage-ads-sex-trafficking.html
43. (1, January 1). Why do sex workers need online spaces? REFRAME HEALTH AND JUSTICE. Retrieved 10 September 2023, from https://survivorsagainstsesta.files.wordpress.com/2018/03/onlinespaces_impact-003.pdf
44. Shreya, S. & Tiwari, P. . (2020, December). Analysing the American Safe Harbour Regime: Takeaways for India. The Dialogue. Retrieved 10 September 2023, from https://thedialogue.co/wp-content/uploads/2020/12/Analysing-the-American-Safe-Harbour-Regime_Takeaways-for-India_The-Dialogue.pdf
45. Shreya, S. & Tiwari, P.(2020, December). Analysing the American Safe Harbour Regime: Takeaways for India. The Dialogue. Retrieved 10 September 2023, from https://thedialogue.co/wp-content/uploads/2020/12/Analysing-the-American-Safe-Harbour-Regime_Takeaways-for-India_The-Dialogue.pdf
46. UK children and adults to be safer online as world-leading Bill Becomes Law. GOV.UK. (2023, October 26). Retrieved 10 September 2023, from https://www.gov.uk/government/news/uk-children-and-adults-to-be-safer-online-as-world-leading-bill-becomes-law
47. UK children and adults to be safer online as world-leading Bill Becomes Law. GOV.UK. (2023, October 26). Retrieved 10 September 2023, from https://www.gov.uk/government/news/uk-children-and-adults-to-be-safer-online-as-world-leading-bill-becomes-law

requires platforms to prevent and remove illegal content, like terrorism, revenge pornography, child exploitation, hate crimes, or fraud.[48] The law takes a zero-tolernace to protecting children and requires platforms from preventing children from seeing harmful content, such as bullying, self-harm and eating disorders promoting content, and pornography.[49] This represents a significant departure from the EU's E-Commerce Directive, adopting a more interventionist approach to regulate harms such as child sexual abuse, hate crimes, fraud, and terrorism.[50]

Due Diligence Requirements: The OSA 2023 has been introduced to protect children and adults by imposing a broad "duty of care" on user-to-user services and search engines, which host user-generated content or facilitate public or private communication between users, to monitor and remove content. The law mandates online platforms to regulate primarily two types of content: illegal content and content that is harmful to children. The Act includes an extensive list of illegal content that platforms must take down, including CSAM, terrorism-related content, fraud, violence, suicide, and more. The Act also differentiates between two categories of content that are subject to age-appropriate protections: "primary priority" and "priority" content that is harmful to children. Platforms are required to prevent children from encountering primary priority content, including pornography, promotion of self-harm, eating disorders, and legal suicide. Platforms are also required to ensure that priority content, such as harassment, health and vaccine misinformation, or violence, is "age appropriate" for children, and have a complaint system for parents to report any violations of these provisions they may come across online. If companies fail to comply with the law, Ofcom has the power to fine them up to £18 million or 10% of their global annual revenue, whichever is bigger.[51]

Adopting such a broad duty-of-care approach to content moderation can create legal ambiguity and uncertainty for services and regulators. This approach may also pose potential risks to users' fundamental rights. As a result, it could lead to significant penalties for failing to meet their obligations, or companies may be motivated to proactively remove legitimate content rather than safeguarding the rights to free expression and privacy and facing enforcement actions. Additionally, the prohibition on general monitoring is likely to result in the excessive removal of content and increased reliance on unreliable automated tools, as services will need to monitor all content on their platforms.[52]

[48] Haves, E. (2023, January 25). Online Safety Bill: HL bill 87 of 2022–23 - House of Lords Library. Retrieved 10 September 2023 https://lordslibrary.parliament.uk/research-briefings/lln-2023-0005/

[49] UK children and adults to be safer online as world-leading Bill Becomes Law. GOV.UK. (2023, October 26). Retrieved 10 September 2023, from https://www.gov.uk/government/news/uk-children-and-adults-to-be-safer-online-as-world-leading-bill-becomes-law

[50] Porter, J. (2023, May 4). The UK's Online Safety Bill, explained. The Verge. Retrieved 10 September 2023, from https://www.theverge.com

[51] UK children and adults to be safer online as world-leading Bill Becomes Law. GOV.UK. (2023, October 26). Retrieved 10 September 2023, from https://www.gov.uk/government/news/uk-children-and-adults-to-be-safer-online-as-world-leading-bill-becomes-law

[52] Senftleben, M. (2020, October 22). The odyssey of the Prohibition on General Monitoring Obligations on the way to the Digital Services Act: between Article 15 of the E-Commerce Directive and Article 17 of the Directive on Copyright in the Digital Single Market. Retrieved 10 September 2023, from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3717022

Before services remove content, they must adequately substantiate its illegality to reduce the risk of platforms unintentionally deleting lawful and legitimate content. However, the Act establishes a low standard of proof for content removal. The OSA standard, which requires providers to have 'reasonable grounds to infer that content is illegal,' may result in platforms using a less stringent criterion than that employed by a court or regulator to assess content legality. Lowering the threshold for content determination and assigning platforms the responsibility to evaluate what constitutes legal speech in the online realm may result in two different standards for permissible speech. This could lead to online speech censorship and an over-removal of legitimate content without sufficient evidence that the content qualifies as an offence. The Act also grants the enforcement authority OfCom the power to instruct online platforms to employ 'accredited technology' for identifying child pornography and radicalising content, as well as preventing users from encountering such content, whether it is communicated publicly or privately.[53]

Impact on Safe Harbour: The Online Safety Act introduces several positive principles in its approach, including adopting a systemic approach to enhance user safety and protection, implementing a risk-based strategy for platform regulation, and fostering regulatory dialogue with the industry. However, the Act places a new additional due diligence mandate on intermediaries to employ "accredited technology" for identifying harmful content across all types of platforms, including private messaging platforms, potentially threatening end-to-end encryption. While the Act doesn't explicitly target end-to-end encryption, private E2EE services may need to either remove or weaken encryption to comply with the law.

Furthermore, this heightened due diligence requirement may not necessarily enhance safety. As evidenced by The Dialogue's study, the threat to revoke safe harbour immunity might only affect well-intentioned intermediaries who already cooperate with law enforcement agencies (LEAs) by providing metadata for investigations.[54] In contrast, those involved in distributing CSAM may simply migrate to dark web sites or potentially create their encrypted systems, making it more challenging for LEAs to apprehend them. End-to-end encryption plays a vital role in preserving confidentiality in technologically driven communications. This technology ensures that individuals can converse with one another without any admissible proof that could be presented in a court of law to demonstrate the exchanged messages. Consequently, it safeguards user privacy in everyday conversations with friends and partners and empowers women and other marginalised groups to express their opinions on public platforms without the fear of offline repercussions.

Similarly, the criminal liability against senior managers for not complying with a direction to provide Ofcom with information raises concerns about encouraging excessive content removal, thereby infringing upon users' fundamental rights. The apprehension

[53] Online safety bill - UK parliament. Online Safety Bill Volume 724: debated on Monday 5 December 2022. (2022, December). Retrieved 10 September 2023, from https://hansard.parliament.uk/Commons/2022-12-05/debates/E155684B-DEB0-43B4-BC76-BF53FEE8086A/OnlineSafetyBill

[54] Azad, Y., Venkatnarayanan, A. Tiwari, P. & Chatterjee, S. (2022, January). Analysing the national security implications of weakening encryption. The Dialogue. Retrieved 10 September 2023, from https://thedialogue.co/wp-content/uploads/2022/01/Report-_-National-Security-Encryption-_-The-Dialogue-DeepStrat-_-Jan-12-2022.pdf

of facing criminal sanctions is likely to pressure platforms into overcompensating by restricting any content falling into a gray area, potentially promoting the over-removal of user-generated content and creating a 'chilling effect' on free speech. While the demand for clear and responsive reporting lines between relevant authorities and intermediaries is valid, the ability to subject individual employees to personal criminal liability is unnecessary and likely to undermine intermediaries' trust in the government's intentions, as well as user trust in intermediaries' capacity to uphold their rights. These extensive penalties and enforcement authorities, when combined with the substantial expectations for intermediaries to prohibit and actively police vaguely defined forms of content, engineer data access, and fulfill onerous due diligence requirements, will place significant burdens on intermediaries of all sizes and across the technology stack.

2.1.4 Australia

Existing Regulations: The Australia Online Safety Act of 2021 is the country's new law introduced to expand and strengthen the existing online safety regulations aimed at keeping up with abusive behaviour and toxic content. Prior to the 2021 Act's enactment, online content was governed by a patchwork of legislations, including the Enhancing Online Safety Act 2015,[55] the Broadcasting Services Act 1992,[56] and the Criminal Code Amendment (Sharing of Abhorrent Violent Material) Act 2019.[57]

In 2019, Australia amended its Criminal Code to classify the sharing of 'abhorrent video material' as a criminal offence.[58] 'Abhorrent video material' is narrowly defined to encompass content depicting a terrorist act, murder, attempted murder, torture, rape, or kidnapping.[59] The Amendment Act imposes criminal liability on service providers if they fail to ensure the prompt removal or cessation of hosting abhorrent violent material. However, it doesn't specify a timeframe for such removal or blocking access to the content in question. Non-compliance with the law can result in penalties of up to 3 years of imprisonment and fines of £2.1 million for individuals and up to £10.5 million or 10% of a corporate body's annual turnover under the Act.

Due Diligence Requirements: Building upon the existing Enhancing Online Safety Act of 2015, the Online Safety Act of 2021 establishes clear responsibilities for online service providers, making them more

55. Office of Parliamentary Counsel, Canberra. (2017, July). Enhancing online safety act 2015 - International Labour Organization. ILO. Retrieved 10 September 2023, from https://www.ilo.org/dyn/natlex/docs/ELECTRONIC/105255/128681/F249839433/AUS105255%202017.pdf
56. Office of Parliamentary Counsel, Canberra. Broadcasting services act 1992. Broadcasting Services Act 1992. Retrieved 10 September 2023, from https://www.legislation.gov.au/Details/C2022C00079
57. Contact, C. P. A. H. C. a. 2. (2021, June 2). Criminal Code Amendment (Sharing of Abhorrent Violent Material) Bill 2019.Retrieved 10 September 2023, from https://www.aph.gov.au/Parliamentary_Business/Bills_Legislation/Bills_Search_Results/Result?bId=s1201#:~:text=Amends%20the%20Criminal%20Code%20Act,Federal%20Police%20within%20a%20reasonable
58. Australian Government (2019), Criminal Code Amendment (Sharing of Abhorrent Violent Material) Act 2019. Retrieved 10 September 2023 https://parlinfo.aph.gov.au/parlInfo/search/display/display.w3p;db=LEGISLATION;id=legislation%2Fbills%2Fs1201_aspassed%2F0001;query=Id%3A%22legislation%2Fbills%2Fs1201_aspassed%2F0000%22#c76e04f2304b4b55b1128637ff286979
59. Australian Government (2019), Criminal Code Amendment (Sharing of Abhorrent Violent Material) Act 2019. Retrieved 10 September 2023 https://parlinfo.aph.gov.au/parlInfo/search/display/display.w3p;db=LEGISLATION;id=legislation%2Fbills%2Fs1201_aspassed%2F0001;query=Id%3A%22legislation%2Fbills%2Fs1201_aspassed%2F0000%22#c76e04f2304b4b55b1128637ff286979

accountable for user safety.[60] The Act introduces five primary "schemes" aimed at addressing the proliferation of illegal and harmful online content: the cyberbullying scheme, the adult cyber-abuse scheme, the image-based scheme, the abhorrent violent material blocking scheme, and the online content scheme.[61]

The previous 2015 version solely addressed cyberbullying against minors, while the new law expands its scope to include cyber abuse and cyberbullying against adult users by creating a "world-first Adult Cyber Abuse Scheme for Australians aged 18 and older."[62] Various online service providers, including social media platforms, electronic messaging services, search engines, app distribution services, internet service providers, hosting service providers, and internet carriage providers, fall under the Act's jurisdiction.

The Online Safety Act formalises the role of the eSafety Commissioner,[63] empowering them to implement and enforce the Act by issuing orders to remove and block access to non-consensual intimate images and violent or abhorrent material.[64]

The law also mandates that online service providers meet Basic Online Safety Expectations (BOSE),[65] which represent the government's expectations for improving online safety. It acknowledges that service providers are best equipped to identify emerging forms of harmful end-user conduct or material and determine the most effective methods for addressing them on their platforms.

The Minister for Communications, Urban Infrastructure, Cities, and the Arts further articulated these expectations, and the BOSE Determination came into effect on 23rd January 2022.[66] Section 46 of the Act outlines core expectations for online platforms, such as taking reasonable measures to ensure user safety and proactively minimising illegal or harmful content or activity on their services. These measures may include the development and implementation of mechanisms for detecting and addressing such harmful content without compromising encryption.[67] However, these expectations could potentially incentivise platforms to increase the use of content moderation tools, resulting in over-moderation and the blanket removal of protected speech.[68]

60. Australian Government. (2021). Online safety act 2021 . eSafety. Retrieved 10 September 2023, from https://www.esafety.gov.au/sites/default/files/2021-07/Online%20Safety%20Act%20-%20Fact%20sheet.pdf
61. Australian Government. (2021). Online safety act 2021 . eSafety. Retrieved 10 September 2023, from https://www.esafety.gov.au/sites/default/files/2021-07/Online%20Safety%20Act%20-%20Fact%20sheet.pdf
62. Australian Government. (2021). Online safety act 2021. eSafety. Retrieved 10 September 2023, from https://www.esafety.gov.au/sites/default/files/2021-07/Online%20Safety%20Act%20-%20Fact%20sheet.pdf
63. Australian Government. (2021). Online safety act 2021 . eSafety. Retrieved 10 September 2023, from https://www.esafety.gov.au/sites/default/files/2021-07/Online%20Safety%20Act%20-%20Fact%20sheet.pdf
64. Australian Government. (2021). Online safety act 2021 . eSafety. Retrieved 10 September 2023, from https://www.esafety.gov.au/sites/default/files/2021-07/Online%20Safety%20Act%20-%20Fact%20sheet.pdf
65. Basic Online Safety Expectations | eSafety Commissioner.Retrieved 10 September 2023, from https://www.esafety.gov.au/industry/basic-online-safety-expectations
66. Basic Online Safety Expectations | eSafety Commissioner. Retrieved 10 September 2023, from https://www.esafety.gov.au/industry/basic-online-safety-expectations
67. Basic Online Safety Expectations | eSafety Commissioner. Retrieved 10 September 2023, from https://www.esafety.gov.au/industry/basic-online-safety-expectations
68. Australian Mobile Telecommunications Association (AMTA), BSA | The Software Alliance (BSA), Communications Alliance Ltd (CA), Consumer Electronics Suppliers Association (CESA), Digital Industry Group Inc. (DIGI) and Interactive Games and Entertainment Ass. (2022). Submissions log and industry associations' responses to public consultation feedback. Retrieved 10 September 2023, from https://onlinesafety.org.au/wp-content/uploads/2022/11/221118_Submissions-log-responses_FINAL.pdf

Expectation 8 of the Act also requires platforms to take reasonable steps to prevent anonymous accounts from engaging in unlawful or harmful material or activities.[69] Such steps may involve preventing the same individual from repeatedly using anonymous accounts to disseminate and engage in unlawful or harmful content and activities,[70] as well as verifying the identity or ownership of accounts.

Impact on Safe Harbour: The additional due diligence requirement under the law to proactively remove harmful content will incentivise and encourage platforms to increase the deployment of content moderation tools, leading to over-moderation and blanket removal of protected speech due to the fear of losing their safe harbour protection.[71] As discussed before, the compulsory use of automated tools for content moderation will also be inconsistent with the role and nature of an intermediary, which is supposed to be a mere conduit.[72] Moreover, the strict one-size-fits-all approach to content moderation may also disproportionately harm smaller platforms.[73] Further, the mandate on user verification will likely impinge upon the active participation of all communities on the internet.[74] Anonymity can be inconvenient from a regulatory standpoint. However, it keeps the free flow of information and opinions in a digital space - an aspect that may be lost permanently if intermediaries are compelled to verify accounts as a precondition to avail safe harbour mandatorily.

2.2 India

Existing Regulations: As previously elucidated, the IT Act of 2000, Section 79, establishes the regulatory framework for platforms in India and provides safe harbour protection to intermediaries. In its original form and following the *Shreya Singhal* judgement,[75] safe harbour immunity required that intermediaries only needed to exercise due diligence by blocking or taking down content when they received actual knowledge of its illegality through a government notice or a court order. However, this legal framework underwent significant changes with the introduction of the IT Rules in 2021, which imposed various additional due diligence requirements that intermediaries must comply with to preserve their safe harbour immunity.

69. Basic Online Safety Expectations | eSafety Commissioner. Retrieved 10 September 2023, from https://www.esafety.gov.au/industry/basic-online-safety-expectations
70. Consultation on the Online Safety Bill 2021 . Global Partners Digital. (2021). Retrieved 10 September 2023, from https://www.rssfeeds.aph.gov.au/DocumentStore.ashx?id=f9650d70-6522-47c9-a6e2-c1fe2f612031&subId=703318
71. Aarathi Ganesan. (2023, May). Can taking away safe harbour fix free speech issues in an internet dominated by a few Tech Cos?. MediaNama. Retrieved 10 September 2023, from https://www.medianama.com/2023/05/223-safe-harbour-free-speech-internet-competition-nama/
72. João Pedro Quintais, Naomi Appelman, & Ronan Ó Fathaigh. (2023). Using Terms and Conditions to Apply Fundamental Rights to Content Moderation. OSF. Retrieved 10 September 2023, from https://osf.io/f2n7m/download
73. Content Moderation: Best practices towards effective legal and procedural frameworks for self-regulatory and co-regulatory mechanisms of content moderation. Council of Europe. (2021, June). Retrieved 10 September 2023, from https://rm.coe.int/content-moderation-en/1680a2cc18
74. Content Moderation: Best practices towards effective legal and procedural frameworks for self-regulatory and co-regulatory mechanisms of content moderation. Council of Europe. (2021, June). Retrieved 10 September 2023, from https://rm.coe.int/content-moderation-en/1680a2cc18
75. primary Singhal v. Union Of India, AIR 2015 SC 1523.

Due Diligence Requirements: In the pursuit of enhancing user safety, the IT Rules 2021 impose various due diligence requirements on intermediaries that they must adhere to in order to secure safe harbour protection. Some of the noteworthy requirements include providing information assistance to law enforcement agencies within 72 hours,[76] executing content blocking orders within 36 hours,[77] and expeditiously removing content in response to user complaints about sexually offensive material within 24 hours.[78] Furthermore, significant social media intermediaries (SSMIs) must also ascertain the first originator of content[79] on private messaging platforms and deploy automated AI tools to proactively identify harmful content to the best of their abilities.[80]

Moreover, the promulgation of a new Digital India Act is expected to bring about significant changes to these due diligence requirements.

Impact on Safe Harbour: Experts have emphasised the inefficiency of the current due diligence requirements, especially the fixed timelines for takedowns and information assistance, as they can lead to increased censorship due to the limited time available for careful consideration of each order.[81] The Dialogue's primary research indicates that due to time constraints, platforms may end up censoring content excessively[82], especially in cases like nudity, which might be taken out of context.[83] For instance, educational content related to human anatomy or sexual health could be flagged for nudity even when intended for educational and awareness purposes.[84] The time constraints also have negative implications for posts and content related to human rights abuses, public emergency information, and whistleblower disclosures.[85] Research demonstrates that approximately 50% of takedown requests

76. FAQ on IT Rules 2021: Ministry of Electronics and Information Technology, Government of India. MeitY. (2021).Retrieved 10 September 2023, from https://www.meity.gov.in/writereaddata/files/FAQ_Intermediary_Rules_2021.pdf
77. FAQ on IT Rules 2021: Ministry of Electronics and Information Technology, Government of India. MeitY. (2021). Retrieved 10 September 2023, from https://www.meity.gov.in/writereaddata/files/FAQ_Intermediary_Rules_2021.pdf
78. FAQ on IT Rules 2021: Ministry of Electronics and Information Technology, Government of India. MeitY. (2021). Retrieved 10 September 2023, from https://www.meity.gov.in/writereaddata/files/FAQ_Intermediary_Rules_2021.pdf
79. The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021. Retrieved 10 September 2023, from https://prsindia.org/billtrack/the-information-technology-intermediary-guidelines-and-digital-media-ethics-code-rules-2021#:~:text=Identification%20of%20the%20first%20originator%20of%20information%3A%20The%20Rules%20require,provision%20under%20the%20parent%20Act
80. The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021. Retrieved 10 September 2023, from https://prsindia.org/billtrack/the-information-technology-intermediary-guidelines-and-digital-media-ethics-code-rules-2021#:~:text=Identification%20of%20the%20first%20originator%20of%20information%3A%20The%20Rules%20require,provision%20under%20the%20parent%20Act.
81. Shreya S. & Tiwari P. (2022, July 4). IT Rules, 2021: A Regulatory Impact Assessment Study (Vol. 1). New Delhi. The Dialogue and Internet And Mobile Association of India. https://thedialogue.co/wp-content/uploads/2022/07/IT-RULES-2021-interactive.pdfn/
82. Shreya, S., Tiwari, P., Rizvi, K., & Saxena, G. (2023, July 18). IT Rules, 2021: A Regulatory Impact Assessment Study (Vol. 2). New Delhi. The Dialogue and Internet And Mobile Association of India. https://thedialogue.co/wp-content/uploads/2023/07/IT-Rules-2021-Analysis-Volume-2.pdf
83. Content Moderation: Best practices towards effective legal and procedural frameworks for self-regulatory and co-regulatory mechanisms of content moderation. Council of Europe. (2021, June). Retrieved 10 September 2023, from https://rm.coe.int/content-moderation-en/1680a2cc18
84. Oosterhoff, P., Müller, C. & Shepherd, K. (2017). Sex education in the Digital Era. IDS Bulletin, 48(1), 61-62. Retrieved 10 September 2023, from https://doi.org/10.19088/1968-2017.102.
85. United Nations Office of the High Commissioner for Human Rights. (2021, July 23). Moderating online content: Fighting harm or silencing dissent? Retrieved 10 September 2023, from https://www.ohchr.org/en/stories/2021/07/moderating-online-content-fighting-harm-or-silencing-dissent

target potentially legitimate or protected speech.[86]

Considering the significant impact of such takedowns on free speech and user privacy, it is crucial to adopt a graded approach. This approach would involve extending the timeline for taking down content that poses a lesser degree of harm, such as defamation or contempt, while shorter timelines would be provided for content with grave security implications.[87]

Furthermore, the norm of content takedown upon receiving a user complaint raises jurisprudential questions regarding its consistency with the *Shreya Singhal* judgement.[88] In that judgement, the Apex Court interpreted actual knowledge to only be established through a government notice or a court order. However, this due diligence requirement compels intermediaries to take action based on individual complaints. Additionally, the traceability and criminal liability norms raise concerns about compromising end-to-end encryption and potentially leading to increased mass censorship, as explained earlier.

Looking ahead, a new law is set to replace the existing platform regulation regime in the country, and the scope and extent of safe harbour protection will be a critical consideration. Safe Harbour jurisprudence has already evolved significantly over the years, with intermediaries facing a growing list of due diligence requirements to enjoy this immunity. While the intent behind intensifying intermediary responsibilities is legitimate, making them a prerequisite for safe harbour protection may not be a feasible solution, given its impact on the openness of the Internet. Safe harbour protection is particularly crucial for intermediaries that enable user-generated content, such as communication platforms, social media platforms, and online forums, as they play a vital role in facilitating user speech and promoting free expression in the digital realm.

Furthermore, the upcoming Digital India Act is expected to introduce a classification system for intermediaries to distinguish between various online intermediaries, including social media platforms, e-commerce sites, AI-based intermediaries, and more.[89] While a one-size-fits-all approach may not be the most effective path forward, it's essential to recognise the limitations of rigid classification. With the evolving business models of intermediaries, a principle-based regulatory model should be envisioned, avoiding stringent categorisation of companies into specific groups. Responsibilities should be defined in consideration of various factors, including the nature and function of the intermediary, the potential harm, and self-regulatory measures, among others.

85. United Nations Office of the High Commissioner for Human Rights. (2021, July 23). Moderating online content: Fighting harm or silencing dissent? Retrieved 10 September 2023, from https://www.ohchr.org/en/stories/2021/07/moderating-online-content-fighting-harm-or-silencing-dissent

86. Bar-Ziv & Elkin-Koren (2018) Behind the Scenes of Online Copyright Enforcement: Empirical Evidence on Notice & Takedown, Connecticut Law Review, Retrieved 10 September 2023, from https://core.ac.uk/download/pdf/327018404.pdf

87. The Dialogue's Comment to MeiTY on Draft Intermediary Guidelines, 2018. Retrieved 10 September 2023, https://www.meity.gov.in/writereaddata/files/Addendum1_Public_comments_on_draft_intermediary_guidelines.pdf

88. Shreya Singhal v. Union Of India, AIR 2015 SC 1523.

89. Proposed Digital India Act, 2023. (2023). Ministry of Electronics and Information Technology, Government of India. Presented at Digital India Dialogues. Retrieved 10 September 2023, from https://www.meity.gov.in/writereaddata/files/DIA_Presentation%2009.03.2023%20Final.pdf.

3 Diluting Safe Harbour to Achieve Safety: A Counterproductive Approach

Safe harbour aims to promote an open Internet and foster innovation[90] while allowing reasonable checks to safeguard user safety. It is based on the principles of providing legal protections or liability exemptions to intermediaries that host or facilitate user-generated content.[91] To qualify for these exemptions, intermediaries must take action upon gaining actual knowledge or awareness of illegal content. However, regulatory frameworks governing intermediaries have introduced numerous due diligence requirements over time, encompassing legal, operational, and user safety obligations.

Initially, these requirements focused on ensuring that content posted by third parties would not lead to legal repercussions for intermediaries. In recent years, though, they have expanded to regulate aspects unrelated to safe harbour protection from illegal or harmful third-party content. Unintentionally, these requirements have broadened the scope of actual knowledge for intermediaries, potentially diluting safe harbour protection.

In essence, safe harbour already acknowledges the necessity of shared responsibility and accountability among intermediaries as prerequisites for enjoying this protection. Nevertheless, the global trend of diluting this protection by significantly expanding due diligence requirements is a cause for concern.[92] The importance of safe harbour extends beyond intermediaries, as it underpins users' fundamental rights and the essential properties and ethos of the core Internet architecture, rooted in openness and inclusivity.[93]

From an economic perspective, studies consistently demonstrate that safe harbour protections stimulate innovation and economic growth. The contributions of online intermediaries to the economy would not be at their current level without these protections. For instance, a 2017 study by NERA Economic Consulting found that weakening safe harbour protections would result in the loss of 4.25 million jobs and £330 billion in GDP every decade, with SMEs (Small and Medium Enterprises) being the hardest hit.[94]
While it is commendable that governments

90. Bambauer, D. E. (2020, July 16). How Section 230 reform endangers internet free speech. Brookings Institution. Retrieved 10 September 2023, https://www.brookings.edu/articles/how-section-230-reform-endangers-internet-free-speech/
91. Schmon , C., & Pedersen, H. (2022, July 19). Platform Liability Trends Around the Globe: From Safe Harbors to Increased Responsibility. Electronic Frontier Foundation. Retrieved 10 September 2023, https://www.eff.org/deeplinks/2022/05/platform-liability-trends-around-globe-safe-harbors-increased-responsibility
92. Schmon , C., & Pedersen, H. (2022, July 19). Platform Liability Trends Around the Globe: From Safe Harbors to Increased Responsibility. Electronic Frontier Foundation. Retrieved 10 September 2023, https://www.eff.org/deeplinks/2022/05/platform-liability-trends-around-globe-safe-harbors-increased-responsibility
93. Schmon , C., & Pedersen, H. (2022b, July 19). Platform liability trends around the globe: Taxonomy and tools of intermediary liability. Electronic Frontier Foundation. Retrieved 10 September 2023, https://www.eff.org/deeplinks/2022/05/platform-liability-trends-around-globe-taxonomy-and-tools-intermediary-liability
94. Asia Internet Coalition. (2020, January). Industry letter on India (IT-Intermediary-Guidelines). Retrieved 10 September 2023, https://aicasia.org/download/307

aim to enhance user protection, it is equally important to ensure that these efforts are well-directed. Evidence-based research should focus on areas of policy implementation, addressing enforcement gaps, and acting on pending complaints without over-regulating the Internet, which could create new challenges threatening user rights.

- **Censorship:** Facing the risk of ruinous prosecution, the dilution of safe harbour leaves little room for intermediaries to adopt a more restrictive approach towards the type of discussions and users allowed on their platforms. For some platforms, this may entail the impositio n of more stringent terms of service, while others might increasingly rely on automated filters. Both of these mechanisms result in unwarranted restrictions on free speech, violations of the right to privacy, and the stifling of innovation. Regardless of the approach chosen by the platforms to mitigate risk, intermediaries leaning towards censorship are likely to disproportionately restrict marginalised voices and reduce the inclusivity of the internet. For example, Germany's Network Enforcement Act (NetzDG) encouraged companies to excessively block legitimate content to avoid the risk of substantial NetzDG fines,[95] based on violations of their terms and conditions.

- **Causing New Challenges:** Furthermore, the threat of revoking safe harbour immunity for failing to adhere to the notified 'best practices' has been found only to deter good-faith service providers who already assist law enforcement agencies (LEAs) by providing metadata to aid investigations. Conversely, those involved in the pornography trade simply migrate to dark websites or potentially develop their private systems, further complicating LEAs' efforts to apprehend them. Additionally, platforms exclusively dedicated to pornography are already ineligible for safe harbour immunity, so the removal of this protection has no impact on them. For example, the FOSTA legislation in the US has been observed to redirect companies' resources away from supporting law enforcement initiatives, focusing primarily on administrative compliance.[96]

- **Good Faith Requirements:** Importantly, recent regulations in this sphere that diluted safe harbours have demonstrated that they do not enhance the effectiveness of combating harmful content. Instead, they disproportionately increase compliance costs for Internet intermediaries and have a chilling effect on investment and innovation in the technology sector. Indeed, unclear and untested standards within the law will create uncertainty for years to come. This, in turn, will discourage investment in startups, raise litigation-related costs for companies of all sizes, harm the entire online ecosystem, and stifle innovation at a time when new innovation is most crucial, especially as Moore's Law[97] is already slowing.

[95.] Masnick, M.(2023, April). The unintended consequences of internet regulation . CCIA Research Center. Retrieved 10 September 2023, from https://research.ccianet.org/reports/unintended-consequences-of-internet-regulation/
[96.] Masnick, M. (2023, April). The unintended consequences of internet regulation . CCIA Research Center. Retrieved 10 September 2023, from https://research.ccianet.org/reports/unintended-consequences-of-internet-regulation/
[97.] Roser, M. (2023, March). What is Moore's law? Our World in Data. Retrieved 10 September 2023, from https://ourworldindata.org/moores-law

While often seen as opposing principles, digital rights and safety represent two complementary aspects of the same concept. Empowering citizens with the Right to Free Speech and privacy is crucial for ensuring online safety and national security. The regulatory experiences of various jurisdictions around the world have already shown that creating broad exceptions to safe harbour protections in the name of promoting online safety is, in reality, counterproductive. These exceptions result in making the internet less safe, especially for women, children, and people from marginalised communities. Prominent international voices, such as UNICEF[98] and the U.N. Special Rapporteur[99] on Freedom of Speech and Expression, share a similar perspective on the importance of human rights in ensuring online safety.

[98.] Kardefelt-Winther, D., Day, E., Berman, G., Witting, S. & Bose, A. (2020, October). Encryption, privacy and children's right to protection from harm. UNICEF. Retrieved 10 September 2023, from https://www.unicef-irc.org/publications/pdf/Encryption_privacy_and_children%E2%80%99s_right_to_protection_from_harm.pdf

[99.] Kaye, D. (2015, May). Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression,. UN Official Documents. Retrieved 10 September 2023, from https://documents-dds-ny.un.org/doc/UNDOC/GEN/G15/095/85/PDF/G1509585.pdf?OpenElement

4 Recommendations for content regulation in India

The cross-jurisdictional experience explains that sacrificing free speech and privacy to enhance online safety does not yield the desired results and, in turn, undermines the freedom and inclusivity of the internet. Consequently, there is a requirement to tackle online content challenges with a more pragmatic approach and envision solutions that can simultaneously achieve safe harbour and safety.

4.1 Envisaging a Balanced Liability Approach for Third-Party Content

India's forthcoming Digital India Bill should ensure that online intermediaries continue to benefit from safe harbour protection and are not held liable for user-provided content in the absence of actual knowledge of its illegality. Furthermore, the definition of 'actual knowledge' should remain in line with the principles established in *Shreya Singhal v. Union of India*.[100] This actual knowledge should be interpreted as either notification through a court order or upon notification by the relevant government authority. Additionally, such notifications must be in accordance with the limitations outlined in Article 19(2) of the Constitution,[101] which specifies the conditions for reasonable restrictions on the fundamental right to freedom of speech and expression.

4.2 No Mandated Content restrictions without Actual Knowledge

The new law must crucially establish clear, precise, and accessible rules for governing intermediaries. Most importantly, intermediaries should not be held liable for choosing not to remove content in the absence of 'actual knowledge' simply because they received a private notification from a user.

4.3 No Mandatory Monitoring or Filtering

Platforms monitoring what users share online may have a disproportionate effect on free speech and privacy. The new law should, therefore, maintain the mandate for proactive monitoring on a best-effort basis and refrain from imposing obligations on digital service providers to actively monitor their platforms or networks for illegal content posted, transmitted, or stored by users.

4.4 Good Samaritan Principles

Good Samaritan laws ensure that an online platform that voluntarily takes steps to address illegal content will not face penalties: actions taken in good faith to identify and act

[100] Shreya Singhal v. Union Of India, AIR 2015 SC 1523.
[101] The Constitution of India, 1950, Art. 19(2).

against illegal content will not strip intermediaries of these liability protections. An intermediary should be able to manually review content voluntarily regarding one type of unlawfulness (e.g., illegal terrorist content) without being deemed to possess knowledge of all the other potential ways in which that same content might be unlawful (e.g., defamation).

4.5 Regulate Processes

Digital India Bill should focus on establishing standards for platform processes. These standards could encompass changes to terms of service and the prompt publication of their compliance reports. Ensuring responsible governance, including informing users and providing explanations whenever platforms modify their terms of service, can help rectify the information imbalance between users and intermediaries. Users should be empowered to gain a better understanding of how to report problematic content to platforms and express their concerns about questionable takedown decisions. They should also receive insights into the workings of content moderation on major platforms.

Privacy should be the default setting, and enhanced transparency and procedural safeguards, such as due process and effective mechanisms for redress in case of removal or blocking decisions, should be implemented to safeguard online safety and digital rights. This must include the introduction of a transparent and robust 'notice-and-action' approach that obliges intermediaries to remove specific illegal content upon notification. However, this removal should occur only when the content is manifestly illegal and doesn't require adjudication by judicial bodies. This approach offers clarity and flexibility, crucial for providing businesses with predictability and the ability to operate and expand. It also allows companies to address illegal content without unduly compromising fundamental rights.

A noteworthy aspect of the notice framework is that, to mitigate the risk to fundamental rights, companies should not be compelled to prioritise speed of removal over careful decision-making.

Moreover, given the breadth and intricacy of contemporary platforms, a data-driven approach that assesses overall processes and outcomes, as opposed to isolated incidents, is imperative. This is especially relevant when it comes to imposing sanctions. Dealing with illegal content is an ongoing challenge without a one-size-fits-all solution. Therefore, it is vital to penalise systemic, recurring lapses rather than individual ones and leverage data-driven methods to determine whether specific errors are anomalies or indicative of more significant issues.

4.6 Flexibility and Suitability

Adapting and tailoring regulations to cater to different types of services constitute a vital component, as exemplified by the European Union's approach under the DSA. It is essential that content regulation takes into consideration and recognises the relevant distinctions between services. What may be suitable for a content-sharing platform may

not be applicable or even technically feasible for a search index or a platform that hosts various applications. Services such as electronic communications service providers and B2B cloud infrastructure providers inherently face limitations when addressing illegal content uploaded or shared by their users. These limitations stem from the technical architecture of their services and the contractual relationships they uphold with users. Anticipating these services to apply the same level of content management as publicly-facing content-sharing services disregards their technical and operational characteristics, potentially resulting in unwarranted intrusions into privacy, security, and commercial aspects.

Furthermore, regulations should demonstrate flexibility to accommodate and embrace new technologies and businesses of varying sizes. Given the ever-evolving nature of technology and the continuous emergence of new forms of communication, governments should embrace an adaptable, cooperative approach that endorses best practices and fosters research and innovation.

4.7 Transparency

Collaborating with governments and other stakeholders, content-sharing platforms should furnish clear information regarding legal removal processes, how to submit complaints and appeals, and promptly assess and act on those submissions. They should also report on the outcomes of these actions. This approach not only offers insights into how the platform enforces its content policies but also provides transparency on the removal requests received by Internet intermediaries from both users and government bodies.

In addition, platforms should ensure that their content policies are publicly accessible in a clear, comprehensible, and user-friendly language and format. This will provide users with clarity on acceptable online behaviour and how to report inappropriate conduct. Transparency reporting should be meaningful and tailored, with a focus on systems and processes that enhance understanding for users and other stakeholders, including policymakers, about the regulations governing the information they share and consume online.

4.8 Enhance Capacity and Effectiveness of the LEAs

Enhancing economic support for the underfunded criminal justice machinery is also crucial to ensure time-bound investigation and prosecution of cybercrimes. The American Invest in Child Safety Act[102] sets a commendable example in this regard. The Act mandates funding of 5 billion dollars and the addition of 100 new FBI agents and 65 positions at the National Center for Missing & Exploited Children (NCMEC) to address online sexual abuse.[103] Substantial funding under the legislation is also allocated to support community-based efforts in preventing child abuse in the digital realm. Additionally, a new office at the White House is established to coordinate these endeavours aimed at combating Child Sexual Abuse Material (CSAM) and promoting child rights in the country.[104] The new Indian Law can draw inspiration from these provisions and develop

[102] Invest in Child Safety Act, H.R.807, 117 Cong. (2021-2022). Retrieved 10 September 2023, from https://www.congress.gov/bill/117th-congress/house-bill/807

measures to enhance the capabilities of law enforcement personnel.

It is important to frame law enforcement assistance carefully and maintain a balance. Law enforcement plays a vital role in ensuring that wrongdoers, both in the physical world and online, are brought to justice for the harm they cause to others. However, obligations imposed on intermediaries, including intrusive user data requests and retention processes, especially those with potential extraterritorial implications, can have serious consequences that often affect a much broader range of users than just the wrongdoers. Moreover, as mentioned earlier, certain intermediaries (e.g. cloud service providers or encrypted messaging service providers) have contractual obligations and technical constraints on the information they can share.

4.9 Sensitisation and Collaboration

In addition to the continuous efforts of intermediaries and the government, it is vital to emphasise the importance of including users as equal participants in the fight against online harms. Many countries have already initiated efforts in this direction. For example, the Australian eSafety Commissioners' office has established a Youth Advisory Council,[105] which will provide feedback to the government regarding online safety issues and measures to counter cyber harms against children. Such feedback, educational initiatives, and consultation mechanisms can be valuable in integrating vulnerable stakeholders into the national decision-making process.

Furthermore, raising awareness about online safety is equally significant. In the long term, ensuring a safe and enjoyable Internet experience can only be achieved through a holistic societal approach that gradually equips citizens to become informed users capable of applying critical thinking and safety tips to the information and content they access online. For instance, India's largest news and information literacy program, FactShala,[106] has designed several nationwide programs aimed at empowering citizens, from students to communities in remote villages, by fostering critical thinking skills and institutionalising media literacy initiatives in different regions/zones of India. Ultimately, these initiatives empower individuals and communities to build resilience against misinformation.

[103] Congresswoman Anna G. Eshoo. (2020, May 8). Invest in Child Safety Act creates mandatory funding to quadruple DOJ child exploitation prosecutors, add 100 new FBI agents and 65 positions at NCMEC to respond to online sex abuse [Press Release]. Retrieved 10 September 2023, https://eshoo.house.gov/media/press-releases/eshoo-wyden-and-colleagues-introduce-legislation-fight-online-child.

[104] Congresswoman Anna G. Eshoo. (2020, May 8). Invest in Child Safety Act creates mandatory funding to quadruple DOJ child exploitation prosecutors, add 100 new FBI agents and 65 positions at NCMEC to respond to online sex abuse. Retrieved 10 September 2023, from https://eshoo.house.gov/media/press-releases/eshoo-wyden-and-colleagues-introduce-legislation-fight-online-child.

[105] The Online Safety Youth Advisory Council. eSafety Commissioner. Retrieved 10 September 2023, from https://www.esafety.gov.au/about-us/consultation-cooperation/online-safety-youth-advisory-council

[106] Factshala. Retrieved 10 September 2023, from https://factshala.com/

AUTHORS

Shruti Shreya

Senior Programme Manager, Platform Regulation and Gender and Tech, The Dialogue™

Shruti Shreya is a Senior Programme Manager at The Dialogue, overseeing two key verticals: 'Gender and Tech' and 'Platform Regulation'. Trained as a lawyer and honored with a Gold Medal from Symbiosis International University, she is deeply engaged in conducting interdisciplinary research on various aspects of social media governance and online safety.

Garima Saxena

Research Associate, The Dialogue™

Garima Saxena is a Research Associate at The Dialogue. She pursued her undergraduate degree from Rajiv Gandhi National University of Law, Punjab. Her prime interest lies in how our society interacts with technology and its impact on individuals. She actively advocates for privacy and digital freedom through her work.

More From Our Research

RESEARCH REPORT

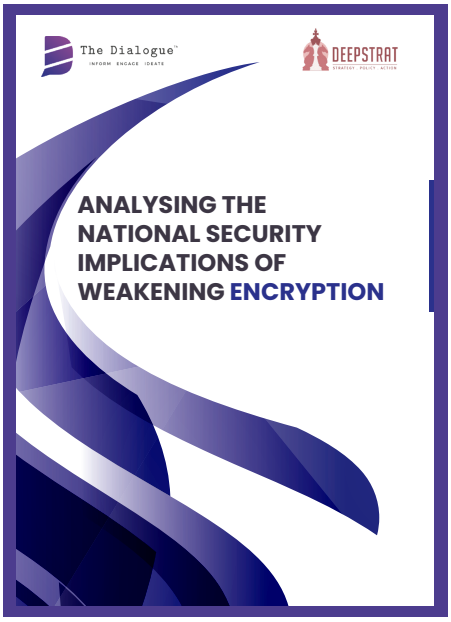
Analysing the American Safe Harbour Regime

RESEARCH REPORT

IT Rules, 2021: A Regulatory Impact Assessment Study | Volume 1

RESEARCH REPORT

IT Rules 2021: IT Rules 2021: A Regulatory Impact Assessment Study [Volume 2]



STUDY

Analysing the National Security Implication of Weakening Encryption

LinkedIn | The Dialogue

Twitter | The Dialogue

Facebook | The Dialogue

Instagram | The Dialogue