Module 5

Trends in Censorship by Private Actors

Advanced Modules on Digital Rights and Freedom of Expression Online

MEDIA DEFENCE

ISBN 978-0-9935214-1-6
Published by Media Defence: www.mediadefence.org
This module was prepared with the assistance of ALT Advisory: https://altadvisory.africa/

Originally published in November 2022
Revised in March 2024

TABLE OF CONTENTS

MODULE 5

TRENDS IN CENSORSHIP BY PRIVATE ACTORS

This module aims to:

- **Give an overview of ways in which non-state actors facilitate online censorship;**
- **Set out the international and regional legal principles that are implicated by online censorship;**
- **Unpack the concept of net neutrality;**
- **Examine the misuse of intermediary liability to curb expression and access;**
- **Explore the right to be forgotten; and**
- **Explain the monitoring obligations of search engines and platforms.**

1. INTRODUCTION

States' obligations to uphold and respect rights, including digital rights, are a cornerstone of international law.[1] However, there is growing appreciation in international law and human rights that much of the digital space, and the technology used to access it, is owned, or controlled by multinational companies, giving the private sector unprecedented power to either uphold or infringe on an array of expressive rights. Litigators and activists must now contend not only with state abuses of digital rights but also violations by private actors.

In 2011, the United Nations Special Rapporteur on Freedom of Expression (UNSR on FreeEx) noted that: "Generally, companies have played an extremely positive role in facilitating the exercise of the right to freedom of opinion and expression," but that "pressure exerted upon them by States, coupled with the fact that their primary motive is to generate profit rather than to respect human rights" creates risks for the private sector to engage in or enable censorship.[2]

In 2021, a subsequent UNSR on FreeEx highlighted the gendered dimensions of these risks, noting that social media companies' failure to address the proliferation of online gender-based

[1] UNHRC, 'The promotion, protection and enjoyment of human rights on the Internet' (2012) (accessible here). See further UNHRC, 'Report of the Special Rapporteur on the rights to freedom of peaceful assembly and of association' (2019) (accessible here).
[2] UNHRC, 'Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression' A/HRC/17/27 (2011) at para 44 (accessible here).

violence, and gender biases in content moderation and other AI-driven processes, have led to the silencing of women's voices online.[3]

This module grapples with some of the long-term threats to freedom of expression from non-state actors, as well as emergent threats. Alongside a brief overview of relevant topics, it provides practical guidance on how to ensure that fundamental rights and freedoms are respected, protected, and promoted online.

2. NET NEUTRALITY

2.1. An overview of net neutrality

The principle of net neutrality is that internet service providers (ISPs) should treat all internet traffic equally, without imposing restrictions or preferential treatment based on factors like the source, destination, or type of data being transferred, or any profit motive. For example, an ISP cannot block, slow down or alter access to service A or make it faster and easier to access service B.[4] This aims to ensure that users have equal access to all online content and services. It means that ISPs must remain neutral and impartial when providing internet access.[5]

Net neutrality is now a well-established principle of contemporary human rights and international law.[6]

- A 2017 report of the UNSR, for example, found that: "In the digital age, the freedom to choose among information sources is meaningful only when Internet content and applications of all kinds are transmitted without undue discrimination or interference by non-State actors, including provider."[7]

- In 2021, a resolution of the Human Rights Council on the promotion and protection of human rights on the internet included a clarion call for states to ensure net neutrality, and to prohibit ISPs from giving preferential access to particular types of content or services for commercial gain.[8]

In principle, net neutrality protections are designed to safeguard freedom of expression and access to information online by ensuring that such freedoms are not determined by market forces or curtailed by network providers. Net neutrality aims to promotes diversity, pluralism, and innovation, and to ensure that people can freely access information and impart ideas across the information society. The Steering Committee on Media and Information Society of

[3] UNHRC 'Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression' (accessible here).
[4] Center for Democracy and Technology 'The importance of internet neutrality to protecting human rights online' (2013) (accessible here).
[5] Carrillo, 'Having Your Cake and Eating It Too? Zero-Rating, Net Neutrality, and International Law' 19 *Stanford Technology Law Review* (2016) at 367 (accessible here).
[6] See further Media Defence 'Training Manual on Digital Rights and Freedom of Expression Online Litigating digital rights and online freedom of expression in East, West and Southern Africa' at 24, (accessible here).
[7] UNHRC, 'Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression' (2017) (accessible here).
[8] UNHRC, 'Resolution on the promotion, protection and enjoyment of human rights on the Internet' (2021) (accessible here).

the Council of Europe, in its report on Protecting Human Rights through Network Neutrality, explained that net neutrality encourages internet users to freely elect how they use their internet connection. The Center for Technology and Democracy explains that:

> "Preserving internet neutrality means preserving the power of individuals to make choices about how they use the Internet – what information to seek, receive, and impart, from which sources, and through which services."[9]

2.2. Net neutrality, development, and human rights

Given net neutrality's role in the advancement of freedom of expression, it should be viewed through a human rights lens. Some have gone as far as suggesting that it is an emerging international human rights norm.[10] Ensuring network neutrality is seen as central to the protection of fundamental human rights and an enabler of fair competition and innovation, as it promotes freedom and enhances network access.[11]

Yet despite the demonstrable link between human rights and net neutrality and the clearly defined position of the UNSR, the past decade has seen growing threats to net neutrality. It has been the subject of regulatory debates and radical shifts in regulations across the world. Additionally, norms and standards have started to develop, and, equally, attempts by state and non-state actors to influence net neutrality and individuals' freedom of expression online are pervasive. This will be outlined below.

2.3. Current challenges and debates

There are two common approaches that interfere with net neutrality:

- **Blocking or throttling of content**, either by state or non-state actors, may include entirely blocking or significantly slowing down access to specific websites, content, or platforms, or restricting access to content in specific geographic regions. This form of restriction contravenes international human rights norms. The Net Neutrality Compendium explains that "blocking certain information resources or restricting what information Internet users can impart over their connection would have serious implications for the right to free expression. For example, blocking access to a particular lawful blog because its content is disfavoured by the access provider would raise obvious concerns." The 2017 Report of the UNSR notes that "States' use of blocking or filtering technologies is frequently in violation of their obligation to guarantee the right to freedom of expression."

- **Zero-rating** involves the differential treatment of content by making certain content available with a zero-download cost.[12] This method is less drastic than blocking and throttling of content and is often framed in terms of public benefit. The 2017 Report of the UNSR describes zero-rating as "the practice of not charging for the use of internet

[9] See Center for Democracy and Technology, above n 4.
[10] See Carillo, above n 5.
[11] Audibert and Murray, 'A Principled Approach to Network Neutrality' *LSE Research Online (*2016) at 120 (accessible here).
[12] Marsden 'Zero Rating and Mobile Net Neutrality' Belli and De Filippi (ed) *Net Neutrality Compendium: Human Rights, Free Competition, and the Future of the Internet* (2016) at 241.

> data associated with a particular application or service; other services or applications, meanwhile, are subject to metered costs." The impact of zero-rating can depend on who implements it, the purpose of the zero-rating, how decisions are made about what content is zero-rated, and the nature of the content itself. In low-income contexts, it can be an effective way to provide widespread access to information in the public interest.

States have responded differently to debates about net neutrality and zero-rating, with some legislating strong protections for the former and others developing policies to promote zero-rating of certain content as a public service.

Certain developed states have shown a trend towards complete bans of zero-rating, perhaps as a reflection of better and more affordable connectivity. Canada, Norway, Slovenia, and the Netherlands are some of the states that have prohibited service providers from differentiating between tariffs for internet access services.[13]

Among developing countries, zero-rating is more likely to be viewed as a policy approach to address challenges such as limited internet access, high data prices and widespread digital divides. Notably, the global COVID-19 pandemic prompted a range of temporary zero-rating initiatives in both developed[14] and developing nations,[15] in which online education, health, and other resources were zero-rated. In many instances, ISPs voluntarily provided zero-rated access to certain resources, such as in **Tanzania** and **Kenya**,[16] while in **South Africa** the government issued regulations requiring ISPs to zero-rate certain resources.[17]

While these measures were enacted as once-off exceptions in the unprecedented challenges of a global pandemic, in the long run, zero-rating could be seen to cause complications in relation to net neutrality. Access Now explains:

> "Activists in advanced economies are struggling to communicate the importance of Net Neutrality for free expression, innovation, and competition, in some cases to audiences that are increasingly anti-regulation. Many in developing countries are facing down critics who argue that non-neutral internet access somehow functions as an "on-ramp" for the free and open internet."

The following examples illustrate the complexity of this debate.

The fight over net neutrality in India

The net neutrality debate came to the fore in India in 2015 with two zero-rated options being offered to Indian users – Facebook's 'Internet.org' and Bharti Airtel's 'Airtel Zero'. Facebook (now Meta) launched Internet.org with the stated intention of providing free basic internet services to people in India, but only to selected online content.[18] At around the same time,

[13] Marsden in Net Neutrality Compendium above n 12 at 248.
[14] Body of European Regulators for Electronic Communication, *BEREC Report on COVID-19 crisis – lessons learned regarding communications networks and services for a resilient society (*2021) (accessible here).
[15] Bhandari, *Improving internet connectivity during Covid-19,* Digital Pathways at Oxford Paper Series no. 4 (2020) (accessible here).
[16] GSMA, 'Education for all during COVID-19: Scaling access and impact of EdTech' (2020) (accessible here).
[17] Bhandari, above n 15 at 19.
[18] Carrillo above n 5 at 367. See further Chaudhry, 'Spotlight on India's Internet: Facebook's Free Basics or Basic Failure' University of Washington Henry M. Jackson School of International Studies (2016) (accessible here).

Airtel launched Airtel Zero, a platform for zero-rated services, offering access to a range of content. Content providers paid Airtel to be included in this service. By April 2015, Airtel was the largest mobile ISP in India with 226 million customers.[19]

That year, the Telecom Regulatory Authority of India (TRAI) called for public comment on its consultation paper on net neutrality. This sparked a national debate on the topic, with many individuals and civil society actors providing comments on the importance of net neutrality. While Meta argued that some access is better than no access, digital rights activists lobbied to introduce regulations to safeguard net neutrality. The process led to significant changes to safeguard net neutrality in India's digital policy:

- In 2016, TRAI released regulations titled "Prohibition of discriminatory tariffs for data services" which, among other things, prohibited any service provider from offering or charging discriminatory tariffs for data services on the basis of content.[20]
- In 2017, TRAI tabled further recommendations for net neutrality with the Department of Technology.[21]
- In 2018, the Indian Government pledged its commitment to the fundamental principles and concepts of net neutrality and was heralded for adopting the world's strongest net neutrality norms.

However, in 2023 TRAI published a policy discussion paper[22] which invited public comment on the possibility of policy changes which would mark a shift away from net neutrality, including a framework for authorisation and network usage feeds for internet services, and a mechanism for 'selective banning' of such services. This drew widespread criticism from Indian civil society organisations and technical experts, who framed the policy discussion as a rollback of the government's previous support for net neutrality.[23] The outcome of this policy process was still pending at the time of this publication.

[19] Marsden in Net Neutrality Compendium at 251.
[20] Telecom Regulatory Authority of India, Regulation no 2 of 2016 (2016) (accessible here).
[21] Telecom Regulatory Authority of India, Recommendations on Net Neutrality (2017) (accessible here).
[22] Telecom Regulatory Authority of India, Consultation Paper on Regulatory Mechanism for Over-The-Top (OTT) Communication Services (2023) (accessible here).
[23] Access Now" Open letter: no discriminatory fees or licencing; TRAI must uphold net neutrality" 2023 (accessible here).

The fight over net neutrality in the United States

The legislative and policy conflicts over net neutrality in the United States reflect larger ideological contests on the role of government and business between successive political administrations.

In 2015, following a Federal Court of Appeals ruling, the Federal Communications Commission (FCC) in the US enacted the historic Open Internet Rules, which prohibited internet providers from engaging with differential pricing for certain content or from giving preferential treatment to certain websites.[24]

However, during the Trump presidency, the US government's view on net neutrality changed. In 2017, under new leadership, the FCC voted to repeal the Open Internet Rules.[25] This decision was viewed as a negative step for many digital rights and free expression activists.[26] Net neutrality advocates challenged this decision, but in 2019 the DC Circuit Court ruled in favour of the FCC and upheld its repeal of the 2015 Rules.[27] In 2020, the DC Court of Appeals dismissed an appeal seeking to reverse the decision.[28]

However, the position was reversed again shortly after President Joe Biden assumed office in 2021 when Biden signed an Executive Order which urged the FCC to reinstate net neutrality rules.[29] In October 2023, the Federal Communications Commission (FCC) voted to proceed with a proposal to restore the net neutrality rules that were repealed during the Trump administration. At the time of publication, the FCC was set to begin public consultations on the proposal with a final decision expected in early 2024.[30]

However, given broader partisan divides in the US political system, it seems likely that the net neutrality debate will continue in the US.

2.4. Practically engaging with net neutrality

As illustrated above, state and non-state actors often seek to depart from the principles of net neutrality and materially change the conditions of people's access to the internet, which impacts the right of freedom of expression and access to information. Overcoming the threats to net neutrality involves two key considerations: the need to ensure adequate safeguards that preserve net neutrality; and the need to understand what limitations are permissible in relation to net neutrality. According to the Net Neutrality Compendium:

[24] See Pouzin, 'Net Neutrality and Quality of Service' in Net Neutrality Compendium above n 12 at 78. See further Access Now 'Net Neutrality matters for human rights across the globe' (2017) (accessible here).
[25] See Washington Post, 'The FCC just voted to repeal its net neutrality rules, in a sweeping act of deregulation' (2017) (accessible here). See further Electronic Frontier Foundation 'Team Internet Is Far From Done: What's Next For Net Neutrality and How You Can Help' (2017) (accessible here).
[26] AccessNow 'The world responds to the U.S. FCC vote against Net Neutrality' (2017) (accessible here).
[27] Washington Post, 'Appeals Court Ruling Upholds FCC's Cancelling of Net Neutrality Rules' (2019) (accessible here).
[28] Endgated, 'US Appeals Court Will Not Rule on Repealing Net Neutrality Laws' (2020) (accessible here)
[29] Office of the US Presidency, 'Fact Sheet: Executive Order on Promoting Competition in the American Economy' (2021) (accessible here).
[30] Vox "Net neutrality is back, but it's not what you think" 2023 (accessible here).

> "To an unprecedented degree, the Internet transcends national borders and reduces barriers to the free flow of information, enabling free expression, democratic participation, and the enjoyment of other rights … Establishing rules to preserve net neutrality – or more precisely, Internet neutrality – is one way to prevent the imposition, by those in a position to control access, of structural inequalities that would distort this environment."31

As discussed above, states should preserve net neutrality to promote the widest possible non-discriminatory access to information. Calling on states to enact laws or regulations to protect net neutrality is an important step in holding states accountable and pushing them to fulfil their responsibilities of protecting freedom of expression.[32]

Tips for good net neutrality protections

The Net Neutrality Compendium provides five principles to guide the substantive development of net neutrality protections that will ensure that states fulfil their obligations in relation to free expression and other human rights online:[33]

- There should be a clear expectation that internet access services must be provided in a neutral manner, without discrimination based on the content, applications or services subscribers choose to access.
- The scope of the neutrality obligation should be clearly defined and should account for the crucial distinction between internet access services and specialised services.
- The neutrality obligation should apply equally to fixed and mobile internet access services.
- There should be clear guidelines for evaluating exceptions for reasonable network management practices.
- The neutrality obligation should not apply to over-the-top services available on the internet.

While adequate legislative and regulatory provisions are the goal, it is, as with all rights, imperative to know what limitations are permissible. The 2011 Joint Declaration on Freedom of Expression and the Internet by a group of Special Rapporteurs on Freedom of Expression from around the world stated:

> "Freedom of expression applies to the Internet, as it does to all means of communication. Restrictions on freedom of expression on the Internet are only acceptable if they comply with established international standards."

[31] McDiarmid and Shears, 'The Importance of Internet Neutrality to Protecting Human Rights Online' in Net Neutrality Compendium at 31-32.
[32] Id at 38.
[33] Id 38-41.

Minimum standards and safeguards for network neutrality regulation

The Net Neutrality Compendium in its Policy Statement on Network Neutrality provides further recommendations for net neutrality regulation:

- **Principle of network neutrality**: Network neutrality is the principle according to which internet traffic is treated without unreasonable discrimination, restriction, or interference regardless of its sender, recipient, type or content.

- **Reasonable traffic management**: While ISPs should follow the principle of network neutrality, there should be provision for reasonable traffic management, meaning any acceptable deviation from these principles is necessary to:
  - Preserve network security and integrity.
  - Mitigate the effects of temporary and exceptional congestion, primarily through protocol-agnostic measures.
  - Prioritise emergency services in the case of public emergencies.

- **Law enforcement**: None of the above should prevent ISPs from giving force to a court order or a legal provision by human rights norms and international law.

- **Transparent traffic management**: ISPs should be transparent about the internet access services they offer, the connection speeds that are to be provided, and their traffic management practices, including how internet access may be affected by the simultaneous use of other services provided by the ISP.

- **Privacy**: All players in the internet value chain, including governments, shall provide robust and meaningful privacy protections for individuals' data in accordance with human rights norms and international law. In particular, any techniques to inspect or analyse internet traffic shall be in line with privacy and data protection obligations and subject to clear legal protections.

- **Implementation:** The relevant authorities should promote independent testing of internet traffic management practices to ensure they meet the principle of network neutrality and other human rights norms and international law. There should be enforceable complaint procedures to address network neutrality violations. All individuals and stakeholders should be able to contribute to the detection, reporting and correction of violations of the principle of network neutrality.

Simply put limitations to net neutrality should only be permitted when provided by law and where necessary and proportionate to the achievement of a legitimate aim.[34] This three-part test is rooted in article 19(3) of the International Covenant on Civil and Political Rights (ICCPR) and must be passed for the legitimate and legal restriction of the right to freedom of expression.

[34] For a detailed outline of the limitation of freedom of expression see Module 2 on Restricting Access and Content at 4 – 5. See also Belli, 'End-to-End, Net Neutrality and Human Rights' in Net Neutrality Compendium at 12.

In a 2018 Report, the UNSR made the following notable statements regarding state and company liability that should be kept in mind when litigating issues around net neutrality:

- **In relation to state responsibility**: Human rights law imposes duties on states to ensure enabling environments for freedom of expression and to protect its exercise. The duty to ensure freedom of expression obligates states to promote, among other things, media diversity, independence, and access to information. Additionally, international and regional bodies have urged states to promote universal internet access. States also have a duty to ensure that private entities do not interfere with the freedoms of opinion and expression. The UN Guiding Principles on Business and Human Rights (Guiding Principles), adopted by the Human Rights Council in 2011, emphasises state duties to ensure environments that enable businesses to respect human rights.

- **About state responsibility**: The Guiding Principles establish a framework according to which companies should, at a minimum, avoid causing or contributing to adverse human rights impacts, and seek to prevent or mitigate such impacts directly linked to their operations, products, or services by their business relationships, even if they have not contributed to those impacts.

2.5. Conclusion

Developing countries continue to face challenges in relation to net neutrality and the suggestion that some access is better than no access. While there is a need for a nuanced approach to zero-rating to enable access to public interest information, the international human rights framework is clear on the need to protect equal access, and states should not enable infringements on net neutrality to serve as justification for failing to take steps toward full and meaningful internet access for all. It is necessary for civil society actors and human rights litigators to ensure that net neutrality is protected through lobbying states, sending complaints to regulators, strategic litigation, and public advocacy, in order to achieve the goal of equal opportunity in access.

3. INTERMEDIARY LIABILITY

3.1. Internet intermediaries – an overview

'Internet intermediary' is a broad, constantly developing term referring to the many services and stakeholders involved in providing access to internet services. The Council of Europe suggests the term encompasses "a wide, diverse and rapidly evolving range of service providers that facilitate interactions on the internet between natural and legal persons." Their functions include connecting users to the internet; hosting web-based services; facilitating the processing of data; gathering information and storing data; assisting searching, and; enabling the sale of goods and services.[35]
Examples of internet intermediaries include:

[35] Media Defence above n 6 at 6.

- Internet service providers who offer connectivity;
- Web hosting companies that provide the infrastructure;
- Search engines and social media platforms, that provide content and facilitate communication.[36]

Simply put, "internet intermediaries are the pipes through which internet content is transmitted and the storage spaces in which it is stored and is therefore essential to the functioning of the internet."[37] Internet intermediaries dominate a pivotal role in the current digital climate impacting social, economic and political exchanges. They can influence the dissemination of ideas and have been described as the "custodians of our data and gatekeepers of the world's knowledge".[38]

It is not difficult to see the link between internet intermediaries and the advancement of an array of human rights. As gatekeepers to the internet, they occupy a unique position in which they can enable the exercise of freedom of expression, access to information and privacy rights. The 2016 Report of the UNSR noted that:

> "The contemporary exercise of freedom of opinion and expression owes much of its strength to private industry, which wields enormous power over digital space, acting as a gateway for information and an intermediary for expression."

3.2. Intermediary liability

Given the important roles that intermediaries play in society, with influence on either upholding or infringing on a myriad of implicated rights, it is imperative to understand their legal liability. The Association for Progressive Communications (APC) explains that intermediary liability refers to the extent that internet intermediaries should be held responsible for what users do through their services. Where intermediary liability exists, ISPs have an obligation to prevent unlawful or harmful activity by users of their services, and failure to do so may lead to legal consequences such as orders to compel or criminal sanctions.

For example, in 2023 the Malaysian Communications and Multimedia Commission (MCMC) announced that it would take legal action against Meta for what it saw as a failure to promptly remove content deemed harmful.[39] This reportedly included matters related to race, royalty, religion, and instances of defamation, impersonation, online gambling, and fraudulent advertisements. Digital rights advocates argued that the MCMC's threat of legal action against

[36] ARTICLE 19, 'Internet intermediaries: Dilemma of Liability', 2013, at 3 (accessible here). See further Li, 'Beyond Intermediary Liability: The Future of Information Platforms' Yale Law School *Information Society Project (*2018) at 9 (accessible here).
[37] Id at 6.
[38] Riordan, 'The Liability of Internet Intermediaries' DPhil thesis, Oxford University (2013,) at 1 (accessible here).
[39] MCMC press statement (2023) (accessible here).

a social media platform for its content moderation decisions poses a potential risk to intermediary liability principles and online freedom of expression.[40]

In a report on the liability of internet intermediaries in Nigeria, Kenya, South Africa, and Uganda, APC captured the following ways in which intermediary liability can arise:

- Copyright infringement.
- Digital privacy.
- Defamation.
- National and public security.
- Hate speech.
- Child protection.
- Intellectual property disputes.

While intermediary liability can be associated with a legitimate interest, there are growing concerns, as noted by the UNSR in the 2016 Report, about the "appropriate balance between freedom of expression and other human rights" and the misuse of intermediary liability to curb expression and access.[41] The legal liability of intermediaries has a direct impact on users' rights, as intermediaries are more likely to be pre-emptively restrictive, and even prevent lawful activity, to avoid possible legal consequences. In this regard, there is a direct correlation between restrictive liability laws – the over-regulation of content – and the increased censorship, monitoring and restrictions of legitimate and lawful online expression.

There are three general approaches to intermediary liability, each with differing considerations and implications: strict liability, the broad immunity model, and the safe-harbour model.

3.2.1. Strict liability

In terms of this approach, intermediaries are liable for third-party content. The abovementioned UNESCO report states that the only way to avoid liability is to proactively monitor, filter, and remove content in order to comply with the state's law. Failing to do so places an intermediary at risk of fines, criminal liability, and revocation of business or media licenses. The UNESCO report notes that China and Thailand are governed by strict liability. This approach is largely considered inconsistent with international norms and standards.

Strict Liability in China

The Stanford CIS World Intermediary Liability Map, which documents intermediary laws around the world, has captured the following in relation to China:

[40] ARTICLE 19 'Malaysia: Halt legal action against Meta over content moderation' 2023 (accessible here).
[41] A 2014 UNESCO report on fostering freedom online and the role of internet intermediaries provides a comprehensive overview of the above regulatory objectives pursued by the states, which in turn have a direct impact on how, and to what extent, intermediaries are compelled to restrict freedom of expression online.

- In 2000, China's State Council imposed obligations on "producing, assisting in the production of, issuing, or broadcasting" information that contravened an ambiguous list of principles (for example opposing the basic principles as they are confirmed in the Constitution; disrupting national policies on religion, propagating evil cults and feudal superstitions; and spreading rumours, disturbing social order, or disrupting social stability).

- China has followed through with its strict liability approach and continues to hold internet companies liable if they fail to comply. This has led to wide-scale filtering and monitoring by intermediaries. This level of oversight has resulted in social media companies being the principal censors of their users' content.

3.2.2. Broad immunity model

On the other end of the spectrum is the broad immunity model, which provides exemptions from liability without distinguishing between intermediary function and content. The UNESCO report cites the Communications Decency Act in the United States as an example of this model, which protects intermediaries from liability for illegal behaviour by users when they do remove content in compliance with private company policy. ARTICLE19 explains that under this model, intermediaries are not responsible for the content they carry, but are responsible for the content they disseminate. The Organisation for Economic Co-operation and Development (OECD), in its Council Recommendation on principles for internet policy, makes reference to this as the preferred model, as it conforms with the best practices, discussed below, and gives due regard to the promotion and protection of the global free flow of information online.

3.2.3. Safe harbour model

The safe harbour model, otherwise known as *conditional liability,* seemingly adopts a middle-ground approach. This approach gives intermediaries immunity provided they comply with certain requirements. Through this approach, intermediaries do not have to actively monitor and filter content but rather are expected to remove or disable content upon receipt of notice that the content includes infringing material. Central to this approach is the idea of 'notice and takedown procedures', which can be content- or issue-specific. There are mixed views on this approach; for some, it is a fair middle-ground; for others, it is a necessary evil to guard against increased filtering or a complete change in the intermediary landscape.[42] As noted in the UNESCO report, there are others who express concern about this approach because of its susceptibility to abuse, as it may lend itself to self-censorship, giving the intermediaries quasi-judicial power to evaluate and determine the legality of content.

[42] Koren, Nahmia and Perel, 'Is It Time to Abolish Safe Harbor? When Rhetoric Clouds Policy Goals' *Stanford Law & Policy Review, Forthcoming* (2019) at 47 (accessible here).

Conditional liability in South Africa

The Freedom of Expression Institute explains the position in South Africa as follows:

- Chapter 11 of the South African Electronic Communications Act 25 of 2002 provides for limited liability of internet intermediaries subject to a takedown notice condition. These provisions apply to members of the Internet Service Providers Association. If an ISP receives a takedown notice to remove harmful content, they must respond immediately; failing which their immunity from liability is forfeited.

- Criticism of South Africa's framework matches broader concerns about the safe harbour approach: that ISPs err on the side of caution and are quick to remove content without providing the content provider with an opportunity to defend the content, and there are no existing appeal mechanisms for content creators or providers. This is concerning given the fact that any individual can submit a take-down notice.[43]

- The potential for these mechanisms to be abused became clear in 2019 when an ISP briefly took down the South African news portal Mail & Guardian Online in response to a fraudulent takedown request which appears to have been submitted in retaliation for an investigative report about a convicted fraudster at the centre of a controversial South African oil deal.[44]

At the core of the debate between the various models is the need to understand the difference between lawful and unlawful content. There is a chilling effect on expression when internet intermediaries are left to their own devices to determine what is good or legal, as it is likely they will tend towards more censorship than less, out of fear of liability.

Keeping in line with a human rights perspective, this guide advocates that "[t]he right to freedom of expression online can only be sufficiently protected if intermediaries are adequately insulated from liability for content generated by others."[45] The following section provides some guidance on applicable international human rights frameworks that can be relied on when advocating for rights in relation to intermediary liability.

Intermediary liability in the courts

Intermediary liability has been dealt with at some length in the European Court of Human Rights (ECtHR). The seminal case of *Delfi AS v Estonia* found that an online news portal was liable for offensive comments they allowed to be posted below one of their news articles.

[43] See further Comninos, 'Intermediary liability in South Africa' (2012) (accessible here). See also Rens, 'Failure of Due Process in ISP Liability and Takedown Procedures' in Global Censorship, Shifting Modes, Persisting Paradigms (2015) (accessible here)..
[44] Mail & Guardian, 'The digital breadcrumbs behind the M&G's censorship attack' (2019) (accessible here).
[45] Media Defence above n 6 at 28.

In *Magyar Tartalomszolgáltatók Egyesülete and Index.hu Zrt v. Hungary*, however, found that imposing objective liability for unlawful comments made by readers on a website placed "excessive and impracticable forethought capable of undermining freedom of the right to impart information on the Internet."

More recently, Media Defence and the Electronic Frontier Foundation (EFF) have intervened in a case at the Grand Chamber of the ECtHR, which concerns online users being held liable for third-party comments. In *Sanchez v France* a French politician was charged with incitement to hatred on religious grounds following comments posted on the 'wall' of his Facebook account by other parties. Because he failed to delete those comments promptly, he was convicted of that offence. The individuals who posted the comments were convicted of the same offence. The Fifth Section of the ECtHR held that his conviction for failing to promptly delete unlawful comments published by third parties on the public wall of his Facebook account did not breach his Article 10 rights despite his apparent lack of knowledge of the comments. The judgment was referred to the Grand Chamber of the ECtHR. In 2023, the Grand Chamber dismissed the application.

3.3. Human rights best practices for intermediary liability

In 2021, the UN Special Rapporteur warned against the trend of states passing regulations and issuing orders to pressure online platforms to police speech, rather than creating rights-preserving processes that can be adjudicated through the courts, noting:

> "The risk with such laws is that intermediaries are likely to err on the side of caution and "over-remove" content for fear of being sanctioned."[46]

Different interest groups continue to push different agendas in relation to internet intermediaries and their liability. Many countries either have non-existent laws or vague and inconsistent laws that make it difficult to enforce rights. There are, however, applicable international human rights frameworks that guide how laws should be enacted or how restrictions may be imposed. With any rights-based advocacy or litigation, it is necessary to establish the rights invoked. As discussed above, it is clear that internet intermediaries play a vital role in the advancement of an array of rights. Thereafter, the next step is to determine responsibility.

In relation to internet intermediaries, the triad of information rights is clearly invoked. The 2010 UN Framework for Business and Human Rights finds that states are primarily responsible for ensuring that internet intermediaries act in a manner that ensures the respect, protection and promotion of fundamental rights and freedoms of internet users. But at the same time, the intermediaries themselves have a responsibility to respect the recognised rights of their users.

Although there might be complexities regarding the cross-jurisdictional scope of intermediaries' powers and responsibilities, international human rights norms should always be at the fore.

[46] UNHRC, 'Disinformation and freedom of opinion and expression' The promotion, protection and enjoyment of human rights on the Internet' (2021) (accessible here

Given the link between internet intermediaries and the fundamental right to freedom of expression, it is best to engage with this topic and test laws, regulations and policies against prescribed human rights standards and understand the restrictions and limitations that may be applicable. As discussed in previous sections, restrictions on the right to freedom of expression have been formulated as a strict, narrow, three-part test – namely, that the restriction must:

- Be provided by law;
- Pursue a legitimate aim; and
- Conform to the strict tests of necessity and proportionality.[47]

Laws content restriction orders and practices must comply with this test. Practically, the need to assess the compliance of legislative frameworks is most likely to be needed in jurisdictions that adopt the strict liability model and the safe-harbour model. The strict liability model can be easily tested and found to be compliant. The safe-harbour model requires slightly deeper engagement to determine compliance, as the following example – namely Kenya's Copyright (Amendment) Act of 2022 – shows.

Copyright reform in Kenya

In 2022, Kenya passed into law the Copyright (Amendment) Act. While the final Act did not deal substantively with intermediary liability, this was due to drafting changes during the public participation process: in its earlier forms, the Copyright (Amendment) Bill, provided some interesting proposals regarding intermediary liability in the African context. A key feature of earlier versions of the Bill was the introduction of the safe-harbour approach, providing for "conduit" safe harbours and "caching" safe harbours. The former would have protected intermediaries from liability for copyright infringements if their involvement was limited to "providing access to or transmitting content, routing or storage of content in the ordinary course of business".

Under these circumstances, the intermediary is not under an obligation to take down or disable content if a takedown notice is received. As per (former) section 35A(1)(b), intermediaries would have been protected if their role was related to content storage that is "automatic, intermediate and temporary". This protection would be conditional upon the removal of content following a take-down notice.[48]

Civil society criticised the lack of clarity and vague notice-and-takedown procedures in the Bill, noting that it fell short of international standards on freedom of expression. ARTICLE 19 listed five problems with the Bill in terms of notice-and-takedown procedures:

[47] For a detailed outline of the limitation of freedom of expression see Module 2 on Restricting Access and Content at 4 – 5. See further OSCE, 'Media Freedom on the Internet: An OSCE Guidebook' (2016) (accessible here).
[48] For a more detailed discussion on the Bill see Walubengo and Mutemi, 'Treatment of Kenya's Internet Service Providers (ISPs) under the Kenya Copyright (Amendment) Bill, 2017', The African Journal of Information and Communication (2019) (accessible here).

- **Lack of proportionality**: criminal sanctions would have been imposed on intermediaries who failed to remove content. As discussed above, this would cause intermediaries to lean toward censorship and blocking, which infringes on freedom of expression.
- **Lack of clarity**: the procedures were vague and did not provide clarity on the issue of counter-notices.
- **Lack of due process**: there was no mention of judicial review or appeal mechanisms. There was also no requirement to notify the content publisher of the alleged infringement. The 48-hour timeframe for content removal would not have allowed for a counter-notice.
- **Lack of transparency**: there was no obligation to maintain records of takedown requests or provide access to such records.
- **Severe sanctions**: the harsh sanctions for false takedown notices would have been disproportionate to the purpose of deterring such.

It is apparent that the necessity and proportionality legs of the test proved to be the sticking points in relation to this Bill. While the safe harbour method might serve a legitimate aim, if the guiding regulations are not clear, necessary, and proportionate, then there is an unjustifiable limitation on freedom of expression.

These sections of the Bill were removed, and the Act was passed in 2022 without addressing intermediary liability.

In 2015, a group of civil society organisations drafted a framework of baseline safeguards and best practices to protect human rights when intermediaries are asked to restrict online content. Known as the Manila Principles, these were drafted with the intention of being "considered by policymakers and intermediaries when developing, adopting, and reviewing legislation, policies and practices that govern the liability of intermediaries for third-party content." Advocates and litigators should similarly rely on these best practice principles, which are based on international human rights instruments and other international legal frameworks when advancing online rights.

Manila Principles

The key tenets of the Manila Principles on Intermediary Liability:

- Intermediaries should be shielded from liability for third-party content.
- Content restrictions should not be required without an order from a judicial authority.
- Requests for restrictions of content must be clear, unambiguous, and follow due process.
- Laws and content restriction orders and practices must comply with the tests of necessity and proportionality.
- Laws and content restriction policies and practices must respect due process.
- Transparency and accountability must be built into laws and content restriction policies and practices.

Digital rights advocates have used these principles to test whether states' legal frameworks and regulations for intermediary liability are adequate. For example, in 2018 India's ICT

ministry published draft regulations that would add new restrictions to that country's existing intermediary liability, including for example that internet intermediaries should automatically, proactively filter out content that promotes cigarettes and alcohol.[49] The Centre for Internet and Society (CIS) made submissions showing that the draft 2018 Rules were unaligned to the Manila Principles and had the potential to infringe on the right to freedom of expression. At the time of this publication, the provisions in the 2018 Draft Rules had not been put into regulation, and the CIS approach is a useful illustration of how the Manila Principles can be used to test domestic legislation against international best practices.

However, from 2021 to 2023 the Ministry subsequently proposed new, more extensive changes to the intermediary liability framework.[50] While these do not include the controversial provisions of the 2018 Draft Rules, the changes extend new liabilities to the online game industry and include new restrictions on publishing information which is "patently false and untrue or misleading in nature". In follow-up submissions, the CIS argued that this effectively requires intermediaries to factcheck any content published through their services, which they argue is unconstitutional.

The apparent successes in having the draft 2018 Rules withdrawn illustrate the importance of digital rights advocates bringing international law to bear in their policy engagements. Yet the subsequent developments in India's intermediary liability framework illustrate the ongoing debates and need for further engagement to ensure emerging policies uphold the principles of freedom of expression online.

3.4. Conclusion

Internet intermediaries play a crucial role in the advancement of human rights. Intermediary liability needs to be understood holistically in relation to the prevention of harm, the protection of free speech and access to information, and encouraging innovation and creativity.[51] While there is a growing trend of online harms and unlawful content:

> "The law must find a way to flexibly address these changes, with an awareness of the ways in which existing and proposed laws may affect the development of information intermediaries, online speech norms, and global democratic values."[52]

4. RIGHT TO BE FORGOTTEN

4.1. Overview of the right to be forgotten

The right to be forgotten, which is also described as the right to be delisted, or the right to erasure, involves an entitlement or right to request that search engines remove links to private

[49] Ministry of Electronics and Information Technology 'Draft Information Technology [Intermediaries Guidelines (Amendment)] Rules, 2018' (2018) (accessible here).
[50] Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules (2021) (accessible here).
[51] Keller, 'Build Your Own Intermediary Liability Law: A Kit for Policy Wonks of All Ages' in Li, 'New Controversies in Intermediary Liability Law Essay Collection Yale Law School' *Information Society Project* (2019) at 20 (accessible here)
[52] Li, 'Beyond Intermediary Liability: The Future of Information Platforms' Yale Law School *Information Society Project* (2018).

information taking into account the right to privacy weighed against public interest considerations.[53]

In India, the Digital Personal Data Protection Bill of 2022, specifically in section 14, introduces the concept of the "Right to be Forgotten."[54] This right grants individuals, known as data principals, the authority to correct or erase their personal data. If a data fiduciary receives such a request, they are obligated to either correct, complete, or update the data principal's information, or erase it if it is no longer necessary for the original processing purpose, unless retention is legally required. It's important to note that the Digital Personal Data Protection Bill is yet to be passed, and currently, the Information Technology Act of 2000 provides relevant protections.

Case Note : *Google Spain SL v Agencia Española de Protección de Datos*

The right to be forgotten was given prominence following the 2014 Court of Justice of the European Union (CJEU) judgement in what has come to be known as the Google Spain case.[55] This judgement has altered the online privacy landscape and has far-reaching legal implications.

In brief, Mr Gonzalez, a Spanish national, took issue with the fact that when internet users searched his name on Google, the search results revealed a news story from 1998 regarding his debt. He requested that the personal information be removed as the matter had been resolved and was no longer relevant. The findings of the CJEU can briefly be summarised as follows:

- The CJEU held that it has jurisdiction to adjudicate the matter, search engines are data controllers, and the right to be forgotten means that personal information that is "inadequate, irrelevant or no longer relevant, or excessive in relation to the purposes of the processing" must be erased by the search engine.
- The CJEU, however, ruled that the right to be forgotten should not apply to information that is relevant to the public interest.

This wide discretion for search engines to balance the competing elements of relevance and the public interest left some digital rights activists concerned. The decision also triggered a debate regarding the tension between the right to privacy and the right to freedom of expression and access to information. Some privacy proponents welcomed the legal development for creating space for people to have some level of control over their personal information, arguing that it "restores the balance between free speech and privacy in the digital world."[56] Others were more circumspect, noting that when information is delisted it affects

[53] See Media Defence above n 6 at 35.
[54] The Indian Express 'Plea in Delhi High Court: What is the 'Right to be Forgotten?' 2023 (accessible here).
[55] For a fuller case note see Media Defence above n 6 at 35.
[56] Cook, 'The Right to be Forgotten: A Step in the Right Direction for Cyberspace Law and Policy', 6 Journal of Law, Technology & the Internet (2015) at 121-123 (accessible here).

other fundamental rights, including freedom of expression and the right to receive and impart information and ideas.[57]

4.2. Evolution of the right to be forgotten

Following the abovementioned judgment, the right to be forgotten has been recognised in domestic contexts,[58] regional legislation and again by the CJEU. For example, the High Court of Orissa, India held in *Rout v State of Odisha* (2020) that the right to be forgotten is an integral part of the right to privacy. Nevertheless, some countries' courts continue to push back against such a right. In *Curi et al v Globo Comunicação e Participações S/A* (2021), the Brazilian Federal Supreme Court held that a general right to be forgotten is incompatible with the Federal Constitution.

As of 2022, Google's Transparency Report revealed that it had delisted nearly 50% of the URLs requested for removal under these terms, having received over 1.3 million requests from users to be "forgotten" since 2014. The relevance of this new right cannot be disputed; however, its scope, applicability and effects are still being debated.

In May 2018, the European Union (EU) elevated the status of the right through article 17 of the General Data Protection Regulation. Article 17 provides data subjects with the right to the erasure of their personal data from search engines. It further obliges search engines to erase personal data without undue delay subject to listed grounds. When erasure is required, article 17(2) stipulates that all reasonable steps must be followed – taking into account the available technology and the cost of implementation – to inform all controllers processing the personal information that any links, copies or replication of the personal data should also be erased. Article 17(3) includes instances when the right to be forgotten does not apply, namely for exercising the right of freedom of expression and information; for compliance with a legal obligation; for reasons of public interest in the area of public health; for archiving purposes in the public interest, scientific or historical research or statistical purposes; or for the establishment, exercise or defence of legal claims.

Further jurisprudence on the right to be forgotten

In September 2019, the CJEU handed down a further ruling in *Google LLC v Commission Nationale de l'Information et des Liberties (CNIL)*. The case dealt with whether a de-listing order made in a member state of the EU meant that the search results had to be removed from all the search engine's domain name extensions globally.

In 2015, the French Data Protection Agency (**CNIL**) requested Google to globally remove information concerning a data subject. Google refused and limited its removal only to EU member states, resulting in CNIL fining Google. Google appealed this decision. Many interested parties, including Wikimedia, Microsoft, governments of EU member states, and civil society actors made submissions to the CJEU. The CJEU acknowledged that the right to

[57] Kulk and Borgesius, 'Freedom of expression and 'right to be forgotten' cases in the Netherlands after Google Spain' 2 European Data Protection Law Review (2015) at 116 (accessible here) See also ARTICLE 19, 'The "Right to be Forgotten": Remembering Freedom of Expression' (2016) (accessible here).
[58] See Media Defence above n 6.

be forgotten is not globally recognised and that the competing interests between the right to privacy and freedom of expression are balanced differently across the world.

Ultimately, the CJEU found that where a search engine operator has granted a de-listing request of a data subject in an EU member state, there is no obligation under EU law for a search engine operator to be ordered to implement the de-listing on all versions of its search engine globally. The CJEU further noted that while EU law does not require de-referencing from all versions of a search engine, such a practice is not prohibited. A judicial authority of a member state remains competent to weigh up – in the light of national standards of protection of fundamental rights – a data subject's right to privacy and the protection of personal data concerning them, on the one hand, and the right to freedom of information, on the other, and, after weighing those rights against each other, to order the operator of that search engine to carry out a de-referencing concerning all versions of that search engine.

Intervening parties ARTICLE 19 and the Electronic Frontier Foundation welcomed the ruling:

> "This ruling is a victory for global freedom of expression. Courts or data regulators in the UK, France or Germany should not be able to determine the search results that internet users in America, India or Argentina get to see. The Court is right to state that the balance between privacy and free speech should be taken into account when deciding if websites should be de-listed – and also to recognise that this balance may vary around the world. It is not right that one country's data protection authorities can impose their interpretation on Internet users around the world."

Other cases have also recently been added to the body of case law on this issue. In *Hurbain v Belgium*, the ECtHR held that an order enforcing the right to be forgotten of a person involved in a road accident through anonymisation did not breach the publisher's freedom of expression. In *Biancardi v Italy*, it likewise held that an online publisher's failure to comply with a de-indexing request justified restricting the publisher's freedom of expression by allowing the request.

The careful navigation of balancing privacy rights against freedom of expression will continue to pose challenges as the digital landscape continues to evolve.[59]

4.3. The extra-territorial scope of the right to be forgotten

In many ways, the CJEU clarified the extra-territorial scope of the right to be forgotten. The CJEU has acknowledged that states are still entitled to develop the content of this right within their respective jurisdictions and are still at liberty to adopt different approaches when balancing the relevant rights and interests – provided that such an approach is compliant with international human rights norms.

4.4. Opportunities and risks

[59] For more on the importance of balancing these right see the Written Observations of ARTICLE 19 and Others (2017) *Google LLC v Commission Nationale de l'Information et des Libertés* (CNIL) (accessible here).

The right to be forgotten can provide important protections for privacy and can fulfil an important role in promoting agency and autonomy. State and non-state actors have far-reaching powers when it comes to the online personal information and identity of individuals. Allowing individuals to have some ownership of their personal information gives them a degree of control over their digital identities. Most online personal information has no bearing on public interest considerations and has far more intrinsic value to the individual than to society at large. The current jurisprudential and legislative developments in this regard have been sensitive to this, recognising the difference between what is of value to an individual, what is interesting to the public, and what is in the public interest.

There were concerns that an "overly expansive right to be forgotten will lead to censorship of the Internet because data subjects can force search engines or websites to erase personal data, which may rewrite history."[60] In some instances, it is permissible for individuals not to be indefinitely defined by their past. The *Google Spain* judgment provides some direction on this, where it recognised the need for relevant considerations to take place – such as the nature and sensitivity of the information, the public interest and the role played by the data subject in public life – when finding a fair balance between the right of the data subject and the interests of internet users.

Shortly after the *Google Spain* judgment, Google received an array of requests from people to have articles of their past removed from the search engine. Google's regular Transparency Reports provide some guidance on how it deals with requests, providing examples of some of the outcomes of requests for erasure. In 2017, for example, the report noted some responses to politician's requests stating "[w]e did not delist the URLs given his former status as a public figure", while another stated, "[w]e delisted 13 URLs as he did not appear to be currently engaged in political life and was a minor at the time." ARTICLE 19 explains that, from a child's rights perspective, binding children to negative aspects of their past can "impede their development and diminish their sense of self-worth."

There are legitimate benefits that accompany the right to be forgotten; however, there are also risks associated with the right, in particular around the enforcement of rights and the adverse effect this can have on the right to freedom of expression.[61] A lack of cogent regulatory safeguards can result in search engines becoming the "judge, jury, and executioner" of the right to be forgotten.[62] There are risks involved in conferring such a decision-making power on a private entity, particularly given the need to balance competing rights, an exercise traditionally reserved for courts.[63] The Electronic Frontier Foundation expressed concern that the "ambiguous responsibility upon search engines" will censor the internet.

[60] Michael L Rustad & Sanna Kulevska, 'Reconceptualizing the Right to Be Forgotten to Enable Transatlantic Data Flow*', 28 Harvard Journal of Law and Technology* 349 (2017) at 373 (accessible here).
[61] Id.
[62] Forde, 'Implications of the Right to be Forgotten' 17 *Tulane Journal of Technology and Intellectual Property* 83 (2015) at 113 -114 (accessible here). See further Lindsay 'The 'Right to be Forgotten' by Search Engines under Data Privacy Law: A Legal Analysis of the Costeja Ruling' 6 *Journal of Media Law (*2016) 159 at 173 – 174.
[63] Kuczerawy & Ausloos, 'From Notice-and-Takedown to Notice-and-Delist: Implementing Google Spain', 14 *Colorado Technology Law Journal* 219 (2016,) (accessible here).

Ensuring adequate safeguards in the right to be forgotten

Access Now has provided some guidance on ensuring clear safeguards for the implementation of the right to be forgotten:

- A right to de-list must be limited to the sole purpose of protecting personal data.
- Criteria for de-listing must be clearly defined in law to protect human rights.
- Competent judicial authorities should interpret standards for deciding what is de-listed.
- The right to de-list must be limited in scope and application.
- Search engines must be transparent about how they comply with de-listing requests.
- Users must have easy access to a remedy.

4.5. Conclusion

The right to be forgotten brings to the fore the tensions between the right to privacy and the right to freedom of expression and given the rapid pace at which digital space is changing, these tensions will likely persist. Provided public interest overrides are prioritised and adequate safeguards are put in place, there can be some degree of consonance.

5. MONITORING OBLIGATIONS OF SEARCH ENGINES AND PLATFORMS

5.1. Overview of monitoring obligations of search engines and platforms

The internet has been described as "the greatest tool in history for global access to information and expression".[64] But it is also a powerful tool for disinformation and hate speech which have, as captured in the Joint Letter from Special Rapporteurs and experts, "exacerbated societal and racial tensions, inciting attacks with deadly consequences around the world." The increase in the spread of disinformation and the rise of the internet being used for nefarious purposes has put non-state actors in a somewhat precarious position. The UN Human Rights Office of the High Commissioner notes that along with the many opportunities associated with the internet, there are growing threats of unlawful activities online. The ease with which malicious content can spread online has posed a dilemma for states and intermediaries. On the one hand, there is a need to mitigate online harms, but on the other, in order to do so, content must not be moderated in a manner that leads to censorship and free speech violations.[65] Intermediaries are now complying with state laws concerning content regulation and are also, in some instances, acting proactively to monitor content, either of their own volition or in order to escape liability.[66]

The 2018 Report by the UNSPR noted key concerns regarding content regulation:

[64] APC, 'Reorienting rules for rights: A summary of the report on online content regulation by the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression' (2018) (accessible here).
[65] Langvardt, 'Regulating Online Content Moderation' Georgetown Law Journal 106 (2018) at 1354-1359 (accessible here).
[66] APC, 'Content Regulation in the Digital Age Submission to the United Nations Special Rapporteur on the Right to Freedom of Opinion and Expression (2018) (accessible here).

> "States regularly require companies to restrict manifestly illegal content such as representations of child sexual abuse, direct and credible threats of harm and incitement to violence, presuming they also meet the conditions of legality and necessity. Some [s]tates go much further and rely on censorship and criminalization to shape the online regulatory environment."

Monitoring obligations for search engines and platforms are loosely understood as general obligations imposed on intermediaries to monitor all content and filter unwanted content.[67] Intermediaries faced with these obligations are expected to develop content recognition technologies or other automatic infringement assessment systems and essentially develop and utilise filtering systems.[68] In instances where there are strict monitoring obligations, monitoring will likely become the norm, opening intermediaries to automatic and direct liability.[69] Monitoring obligations raise concerns with respect of intermediary liability. It has been noted that:

> "Monitoring obligations drastically tilt the balance of the intermediary liability rules toward more restriction of speech, may hinder innovation and competition by increasing the costs of operating an online platform, and may exacerbate the broadly discussed problem of over-removal of lawful content from the Internet."[70]

Further to the above, there has been a trend, akin to that of the right to be forgotten, where states demand the global removal of content that violates domestic law.[71] Notwithstanding the recent findings of the CJEU, these demands might continue, as predicted by the UNSR in the 2018 Report, to have the chilling effect of allowing censorship across borders.

The imposition of monitoring obligations appears to have primarily been about copyright infringements. However, it is growing at an unprecedented rate, causing grave concern for free expression.[72] Judgments of the European Court of Human Rights (ECtHR) provide useful insight into the issues regarding online platforms and liability for users' comments.

5.2. Jurisprudential developments

The *Delfi v Estonia* matter was the first of the prominent cases to address the issue of content moderation and online media liability. An Estonian newspaper, Delfi, published an article that was critical of a ferry company. The article received 185 comments online, some of which were targeting a board member of the company, L, and were considered threatening and/or offensive. L requested that the comments be immediately taken down and claimed approximately €32,000 in compensation for non-pecuniary damages. Delfi agreed to remove the comments but refused to pay the damages. L approached the Harju County Court, bringing a civil claim against Delfi. The County Court found that the company could not be considered the publisher of the comments, and it did not have an obligation to monitor them. L appealed to the Tallinn Court of Appeal who remitted the matter back to the County Court for

[67] Frosio, 'From Horizontal to Vertical: an Intermediary Liability Earthquake in Europe' Centre for International Intellectual Property Studies Research Paper (2017) at 12 (accessible here).
[68] Id.
[69] Id.
[70] Stanford Law, 'Monitoring Obligations' (2017) (accessible here).
[71] See discussion above on the right to be forgotten, particularly the discussion on *Google LLC v Commission Nationale de l'Information et des Liberties (CNIL).*
[72] Frosio, 'The Death of 'No Monitoring Obligations' A Story of Untameable Monsters' *JIPITEC* (2017) (accessible here).

reconsideration, concluding that the lower court had erred in its finding about Delfi's liability. The matter eventually reached the Supreme Court, which found that there was a legal obligation to avoid causing damage to other persons and that Delfi should have prevented the clearly unlawful comments from being published. The Supreme Court noted that after the comments had been published, Delfi failed to remove them on its own initiative, although it must have been aware of their unlawfulness. Delfi's failure to act was found to be unlawful.

Delfi applied to the First Section of ECtHR, arguing that the imposition of liability for the comments violated its right to freedom of expression. The ECtHR was faced with the question of whether Delfi's obligation, as established by the domestic judicial authorities, to ensure that comments posted on its internet portal did not infringe the personality rights of third persons was in accordance with the right to freedom of expression. In order to resolve this question, the ECtHR developed a four-stage test:

- The context of the comments.
- The measures applied by Delfi in order to prevent or remove defamatory comments.
- The liability of the actual authors of the comments as an alternative to the applicant company's liability.
- The impacts of the restrictions imposed on Delfi in a democratic society.

The ECtHR found that the restriction on Delfi's right to freedom of expression was justified and proportionate, taking into consideration the following:

- The insulting and threatening nature of the comments which were posted in reaction to an article published by Delfi;
- The insufficiency of the measures taken by Delfi to avoid damage being caused to other parties' reputations and to ensure a realistic possibility that the authors of the comments will be held liable; and
- The moderate sanction imposed on Delfi.

Following this decision by the First Section, the matter was then referred to the Grand Chamber of the ECtHR. In 2015, the Grand Chamber affirmed the judgment of the First Section. In this regard, in the 2015 *Delfi v Estonia* judgement, the Grand Chamber noted:

> "[W]hile the Court acknowledges that important benefits can be derived from the Internet in the exercise of freedom of expression, it is also mindful that liability for defamatory or other types of unlawful speech must, in principle, be retained and constitute an effective remedy for violations of personality rights."

The Grand Chamber, in determining if freedom of expression had been infringed, considered the restriction was lawful, sought to achieve a legitimate aim and was necessary in a democratic society. Ultimately the Grand Chamber concluded that Delfi was liable for defamation as the publisher of the comments. The Grand Chamber found that "an active intermediary which provides a comments section cannot have absolute liability" and noted that "freedom of expression cannot be turned into an exercise in imposing duties."

While the Grand Chamber found that the liability against Delfi had been a justified and proportionate restriction on the news portal's freedom of expression, it noted, in its appendix that:

> "We trust that this is not the beginning (or the reinforcement and speeding up) of another chapter of silencing and that it will not restrict the democracy-enhancing potential of the new media. New technologies often overcome the most astute and stubborn politically or judicially imposed barriers. But history offers discouraging examples of censorial regulation of intermediaries with lasting effects."

Shortly after the Grand Chamber's *Delfi* judgment, the Fourth Section of the ECtHR considered whether a non-profit, self-regulatory body of intermediaries (MTE) and an internet news portal (Index) were liable for offensive comments posted on their websites in *Magyar Tartalomszolgáltatók Egyesülete v Hungary*. In 2010, the two parties published an article critical of two real estate agents. The article attracted some comments that the estate agents found to be false and offensive and which, they argued, infringed on their right to a good reputation. MTE and Index were held liable by the Hungarian courts for the comments. MTE and Index approached the ECtHR arguing that their right to freedom of expression had been violated.

The ECtHR noted that interferences with the freedom of expression must be "prescribed by law," have one or more legitimate aims, and be "necessary in a democratic society." The ECtHR applied the same four-stage test as it did in Delfi but differed from its finding in Delfi, concluding that there had been a violation of freedom of expression. The ECtHR found that:

- The comments triggered by the article can be regarded as going to a matter of public interest and while they were vulgar, they were not necessarily offensive, noting that style constitutes part of the communication as the form of expression and is protected together with the content of the expression.

- The conduct of MTE and Index in providing a platform for third parties to exercise their freedom of expression by posting comments is a journalistic activity of a particular nature. It was noted that it would be difficult to reconcile MTE and Index's liability with existing case law that cautions against the punishment of a journalist for assisting in the dissemination of statements made by another person.

- MTE and Index took certain general measures to prevent defamatory comments on their portals or to remove them.

The ECtHR found that there had been a violation of freedom of expression and concluded with the following:

> "…[I]n the case of *Delfi,* the Court found that if accompanied by effective procedures allowing for rapid response, the notice-and-take-down-system could function in many cases as an appropriate tool for balancing the rights and interests of all those involved. The Court sees no reason to hold that such a system could not have provided a viable avenue to protect the commercial reputation of the plaintiff. It is true that in cases where third-party user comments take the form of hate speech and direct threats to the physical integrity of individuals, the rights and interests of others and of society as a whole might entitle Contracting States to impose liability on Internet news portals if they failed to take measures

to remove clearly unlawful comments without delay, even without notice from the alleged victim or from third parties. However, the present case did not involve such utterances."

UNSR guidance on applying human rights standards to online content moderation

These guidelines and recommendations are based on the Guiding Principles on Business and Human Rights as well as established international law, norms, and practices. These can be used when engaging with state and non-state actors to ensure compliance with human rights standards when online content is being moderated. Below is an outline of some of the key recommendations:

1. **Human rights by default**: Companies should incorporate directly into their terms of service and community standards relevant principles of human rights law that ensure content-related actions will be guided by the same standards of legality, necessity and legitimacy that bind state regulation of expression.

2. **Legality**: Company rules routinely lack the clarity and specificity that would enable users to predict with reasonable certainty what content places them on the wrong side of the line. Companies should supplement their efforts to explain their rules in more detail with aggregate data illustrating trends in rule enforcement, and examples of actual cases or extensive, detailed hypotheticals that illustrate the nuances of interpretation and application of specific rules.

3. **Necessity and proportionality**: Companies should not only describe contentious and context-specific rules in more detail; they should also disclose data and examples that provide insight into the factors they assess in determining a violation, its severity and the action taken in response.

4. **Non-discrimination**: Meaningful guarantees of non-discrimination require companies to transcend formalistic approaches that treat all protected characteristics as equally vulnerable to abuse, harassment and other forms of censorship.

These guidelines and recommendations provide further guidance on the processes for company moderation and related activities:

1. **Prevention and mitigation**: Companies should adopt and then publicly disclose specific policies that "direct all business units, including local subsidiaries, to resolve any legal ambiguity in favour of respect for freedom of expression, privacy, and other human rights". Companies should also ensure that requests are in writing, cite specific and valid legal bases for restrictions and are issued by a valid government authority in an appropriate format.

2. **Transparency**: Best practices on how to provide such transparency should be developed. Companies should also provide specific examples as often as possible and should preserve records of requests made.

3. **Due diligence**: Companies should develop clear and specific criteria for identifying activities that trigger assessments and assessments should be ongoing and adaptive to changes in circumstances or operating context.

4. **Public input and engagement**: Companies should engage adequately with users and civil society, particularly in the global south, to consider the human rights impact of their activities from diverse perspectives.

5. **Rule-making transparency**: Companies should seek comments on their impact assessments from interested users and experts when introducing products and rule modifications. They should also clearly communicate to the public the rules and processes that produced them.

6. **Automation and human evaluation**: Company responsibilities to prevent and mitigate human rights impacts should take into account the significant limitations of automation and, at a minimum, technology developed to deal with considerations of scale should be rigorously audited and developed with broad user and civil society input.

7. **Notice and appeal**: Companies could work with one another and civil society to explore scalable solutions such as company-specific or industry-wide ombudsman programmes and the promotion of remedies for violations.

8. **Remedy**: Companies should institute robust remediation programmes, which may range from reinstatement and acknowledgement to settlement processes.

9. **User autonomy**: While content rules in closed groups should be consistent with baseline human rights standards, platforms should encourage such affinity-based groups given their value in protecting opinion, expanding space for vulnerable communities and allowing the testing of controversial or unpopular ideas.

It has been noted that there are some inconsistencies in the ECtHR's approach to online liability.[73] However, it does appear that the shift away from the *Delfi* reasoning was a shift in the right direction.[74] Ultimately, these cases have illustrated that even though freedom of expression is paramount, complete immunity is not always attainable, and there might be instances where intermediaries will be responsible for the moderation of content.[75]

5.3. Efforts to address content moderation at the global level

The UN Human Rights Office of the High Commissioner has noted:

[73] Fahy, 'The Chilling Effect of Liability for Online Reader Comments' European Human Rights Law Review (2017) (accessible at https://www.ivir.nl/publicaties/download/EHRLR_2017_4.pdf).
[74] Id at 3. See also Media Defence 'European Court clarifies intermediary liability standard' (2016) (accessible here).
[75] For substantive commentary on the impact of these cases on intermediary liability see Maroni, 'A Court's Gotta Do, What a Court's Gotta Do. An Analysis of the European Court of Human Rights and the Liability of Internet Intermediaries through Systems Theory' EUI Working Paper (2019) (accessible here).

> "One of the greatest threats to online free speech today is the murkiness of the rules . . . States circumvent human rights obligations by going directly to the companies, asking them to take down content or accounts without going through legal process, while companies often impose rules they have developed without public input and enforced with little clarity. We need to change these dynamics so that individuals have a clear sense of what rules govern and how they are being applied."

Alongside the considerable rights implications for the moderation of online content by intermediaries, there is a glaring lack of adequate rules, guidelines, procedures, and remedies in relation to the current practices of content moderation that are cause for concern.[76] It is clear that a human rights framework ought to guide the principles for company content moderation.

6. CONCLUSION

The growing power of private actors within the internet and technology sphere raises new questions with regard to the protection of freedom of expression in the modern age. Private actors have gained the ability to filter and control the flow of information to internet users, raising questions about net neutrality, and complex challenges with regard to enabling access to the internet and to information in developing countries, while maintaining the free and unhindered flow of information.

These powerful actors, along with online news publishers and a host of other internet intermediaries, have also become responsible for hosting huge quantities of information created and posted by regular users, raising questions about how responsibility should be apportioned for illegal or damaging content online. In particular, concerns have been raised that apportioning liability to intermediaries risk creating a digital ecosystem in which freedom of expression is routinely and structurally stymied because of fears of being held liable.

The right to privacy and the protection of personal information has come up against the free flow of information in the issue now known as 'the right to be forgotten,' which has begun to be dealt with at length in regional and domestic courts. This issue relates closely to that of the content moderation obligations of private platform providers and search engines, who must make influential decisions every day about what content will be allowed and what will be removed, with significant consequences for the right to freedom of expression in the digital age.

As a result, it is vital that mechanisms and processes for greater transparency and accountability over the decisions of these powerful, private actors be put in place to ensure alignment with international human rights law and standards on freedom of expression and access to information.

[76] ARTICLE 19, 'Social Media Councils: Consultation' (2019) (accessible here).