

Module 5

Trends in Censorship by Private Actors

*Advanced Modules
on Digital Rights and
Freedom of
Expression Online*



ISBN 978-0-9935214-1-6

Published by Media Legal Defence Initiative: www.mediadefence.org

This report was prepared with the assistance of ALT Advisory: <https://altadvisory.africa/>

This work is licenced under the Creative Commons Attribution-NonCommercial 4.0 International License. This means that you are free to share and adapt this work so long as you give appropriate credit, provide a link to the license, and indicate if changes were made. Any such sharing or adaptation must be for non-commercial purposes and must be made available under the same “share alike” terms. Full licence terms can be found at <http://creativecommons.org/licenses/by-ncsa/4.0/legalcode>.

Table of Contents

Introduction	1
Net Neutrality	2
<i>An overview of net neutrality</i>	2
<i>Net neutrality, development and human rights</i>	2
<i>Current challenges and debates</i>	3
<i>Practically engaging with net neutrality</i>	6
<i>Conclusion</i>	9
Intermediary Liability	9
<i>Internet intermediaries – an overview</i>	9
<i>Internet intermediary liability</i>	10
<i>Strict liability</i>	11
<i>Broad immunity model</i>	11
<i>Safe-harbor model</i>	12
<i>Applicable international human rights standards and current international best practices</i> .13	
<i>Conclusion</i>	16
Right To Be Forgotten	16
<i>Overview of the right to be forgotten</i>	16
<i>Evolution of the right to be forgotten</i>	17
<i>The extra-territorial scope of the right to be forgotten</i>	18
<i>Opportunities and risks</i>	19
<i>Conclusion</i>	20
Monitoring Obligations of Search Engines and Platforms	20
<i>Overview of monitoring obligations of search engines and platforms</i>	20
<i>Jurisprudential developments</i>	22
<i>Efforts to address content moderation at the global level</i>	24

MODULE 5

Trends in Censorship by Private Actors

- To provide an overview of the current ways in which non-state actors are facilitating online censorship.
 - To provide an overview of the fundamental international and regional legal principles that are implicated as a result of censorship.
 - To understand the implications of net neutrality.
 - To examine the misuse of intermediary liability to curb expression and access.
 - To understand the development and implications for the right to be forgotten.
 - To identify the monitoring obligations of search engines and platforms.
-

Introduction

It is now well established that “the same rights that people have offline must be protected online”.¹ However, in addition to increasing trends in censorship by states, increasing trends in censorship by non-state actors are threatening and infringing an array of rights, most notably the right to freedom of expression. Litigators and activists must now contend with not only state abuses of human rights, but also violations of digital rights by private actors.

According to the [2011 Report](#) of the United Nations Special Rapporteur on Freedom of Expression (UNSR), the “framework of international human rights law remains relevant today and equally applicable to new communication technologies such as the Internet.” This is particularly true for freedom of expression, as the UNSR explains: the “[i]nternet has become a key means by which individuals can exercise their right to freedom of opinion and expression.” With the rise in the use of the internet, there has also been a rise in new actors who have begun wielding significant power. Unfortunately, the use of this power is often opaque and problematic. Social media platforms and multinational online companies are, to a large extent, in control of the facilitation of people’s enjoyment of their human right online. Like many state actors, non-state actors are not always acting in accordance with the basic and fundamental principles of international human rights law.

This module grapples with some of the long-term threats to freedom of expression from non-state actors, as well as emergent threats. Alongside providing a brief overview of relevant

¹ UN Human Rights Council, ‘The promotion, protection and enjoyment of human rights on the Internet’ A/HRC/RES/20/8 (16 July 2012) (accessible at <https://daccess-ods.un.org/TMP/9602589.01119232.html>). See further UN Human Rights Council ‘Report of the Special Rapporteur on the rights to freedom of peaceful assembly and of association’ A/HRC/41/41 (17 May 2019) (“2019 Special Rapporteur Report”) (accessible at <https://undocs.org/A/HRC/41/41>).

topics, it provides practical guidance on how to ensure that fundamental rights and freedoms are respected, protected and promoted online.

Net Neutrality

An overview of net neutrality

In the [2017 Report](#) that focused on the Information and communications technology (ICT) sector, the UNSR explained that net neutrality requires that internet data is treated equally without undue interference.² 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.³ Essentially, this means that internet service providers (ISPs) must remain neutral and impartial when providing internet access. In this regard, ISPs cannot alter competition, or unduly interfere with or diminish opportunities for content providers. Additionally, and as explained by Center for Technology and Democracy in a report on the [Importance of Internet Neutrality to Protecting Human Rights Online](#), service providers cannot discriminate against or manipulate internet traffic on the basis of source, destination, content or associated application. For example, an ISP cannot block, slow down or alter access to service A or make it faster and easier to access service B.

Net neutrality fulfils an important role in the current digital climate and is important for ensuring that people can freely access information and impart ideas across our information society. It promotes diversity, pluralism and innovation. The Steering Committee on Media and Information Society of Council of Europe in its report on [Protecting Human Rights through Network Neutrality: Furthering Interest Users' Interests, Modernising Human Rights and Safeguarding the Open Internet](#) explained that net neutrality encourages internet users to freely elect how they utilise 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.”

Net neutrality, development and human rights

From a human rights perspective, net neutrality is undeniably a tool for the advancement of freedom of expression, and, accordingly, 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.⁴ Ensuring network neutrality is seen as central to the protection of fundamental human rights

² 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 at <https://www.mediadefence.org/resources/mldi-training-manual-digital-rights-and-freedom-expression-online>).

³ Carrillo, ‘Having Your Cake and Eating It Too? Zero-Rating, Net Neutrality, and International Law’ (2016) 19 *Stanford Technology Law Review* 364 at 367 (accessible at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2746447).

⁴ Id.

and an enabler to fair competition and innovation, as it promotes freedom and enhances network access.⁵ In the same [2017 Report](#), the UNSR noted:

“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 providers. The State’s positive duty to promote freedom of expression argues strongly for network neutrality in order to promote the widest possible non-discriminatory access to information.”

Yet, despite the demonstrable link between human rights and net neutrality and the clearly defined position of the UNSR, the past decade has shown there to be 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 individual’s freedom of expression online are on the rise. This will be outlined below.

Current challenges and debates

Presently there are two common methods of limiting net neutrality:

- The first entails the **blocking or throttling of content**. It is largely acknowledged that this form of restriction is contrary to international human rights norms. This can be done by both state and non-state actors. 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.”
- The second, which appears to be less drastic than blocking, involves the differential treatment of content through the exclusion of preferred content, by ensuring that other content has a zero-download cost.⁶ This is more commonly referred to as **zero-rating**. The [2017 Report](#) of the UNSR describes zero-rating as “the practice of not charging for the use of internet data associated with a particular application or service; other services or applications, meanwhile, are subject to metered costs.” States have responded differently to net neutrality and zero-rating, with some legislating strong protections for the former. Among the key issues of the debate regarding net neutrality and zero-rating is the digital divide between developed and developing nations.

There is an emergent trend that developed states are more likely to ban zero-rating completely. Canada, Norway, Slovenia and the Netherlands are some of the states that have

⁵ Audibert and Murray, ‘A Principled Approach to Network Neutrality’ (2016) *LSE Research Online* at 120 (accessible at http://eprints.lse.ac.uk/67362/7/Murray_Principled%20approach_2016.pdf).

⁶ 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* (Springer, 2016) (“Net Neutrality Compendium”) at 241.

prohibited service providers from differentiating between tariffs for internet access services.⁷ Developed countries generally have widespread access to the internet, as well as affordable mobile data. On the other hand, developing and emerging countries are faced with different challenges, including limited access, high data prices and widespread digital divides, causing some 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 for net neutrality in India

The net neutrality debate came to the fore with two zero-rated options being offered to Indian users in 2015 – Facebook’s ‘Internet.org’ and Bharti Airtel’s ‘Airtel Zero’. In February 2015, Facebook launched Internet.org with the stated intention of providing free basic internet services to people in India, but only to selected online content.⁸ At around the same time, 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.⁹

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. This process led to the TRAI releasing recommendations on the prohibition of discriminatory data services, which essentially prohibited ISPs from offering or charging discriminatory tariffs for data services on the basis of content. It is worth mentioning that amidst the upheaval around zero-rating, Mark Zuckerberg, the founder of Facebook, stated in a [video](#): “Some may argue for an extreme definition of net neutrality that says that it’s somehow wrong to offer any more services to support the unconnected, but a reasonable definition of net neutrality is more inclusive. Access equals opportunity. Net neutrality should not prevent access.”

Facebook argued that some access is better than no access. This was not well received by digital rights activities, who lobbied to introduce regulations safeguarding net neutrality. Within two years, the net neutrality landscape underwent significant changes:

⁷ Marsden in Net Neutrality Compendium at 248.

⁸ Carrillo above n 3 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 at <https://jsis.washington.edu/news/spotlight-indias-internet-facebooks-free-basics-basic-failure/>).

⁹ Marsden in Net Neutrality Compendium at 251.

- 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.
- In 2017, TRAI provided the Department of Technology with further recommendations regarding net neutrality.
- In 2018, the Indian Government pledged its commitment to the fundamental principles and concepts of net neutrality.
- In July 2018, India was heralded for adopted the [world’s strongest net neutrality norms](#).

The fight regarding net neutrality in the United States

Legislative and policy developments in the U.S. provide a useful case study into the nuances of the net neutrality principle and illustrate how politics and economics are at a crossroads with human rights. The Harvard Business Review, in an article on the [Tangled Web of Net Neutrality and Regulation](#), notes that “Despite being a simple idea, net neutrality has proven difficult to translate into U.S. policy. It sits uncomfortably at the intersection of highly technical internet architecture and equally complex principles of administrative law.”

In 2015, following a DC Circuit Court of Appeals ruling, the Federal Communications Commission (**FCC**) in the U.S. 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.¹⁰ Under the new administration, the view on net neutrality changed.

In 2017, the FCC voted to repeal the Open Internet Rules.¹¹ This decision was viewed as a negative step for many digital rights and free expression activists. [Access Now](#) captured some of the responses by open internet advocates and rights organisations:

- “This order brazenly prioritizes the profits of internet middlemen over the health of the internet ecosystem and the freedom of internet users. We’re very disappointed to see this abdication by the U.S. of its leadership in internet governance” - Aravind Ravi-Sulekha, Internet Freedom Foundation (India).
- “Today the country with the largest share of the global internet economy has entered into a dangerous experiment. By abolishing the rules that protect the innovation of its startups and the free speech of its citizens, the benefits of mankind’s greatest invention

¹⁰ 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 at <https://www.accessnow.org/net-neutrality-matters-human-rights-across-globe/>).

¹¹ See Washington Post, ‘The FCC just voted to repeal its net neutrality rules, in a sweeping act of deregulation’ (2017) (accessible at <https://www.washingtonpost.com/news/the-switch/wp/2017/12/14/the-fcc-is-expected-to-repeal-its-net-neutrality-rules-today-in-a-sweeping-act-of-deregulation/>). See further Electronic Frontier Foundation ‘Team Internet Is Far From Done: What’s Next For Net Neutrality and How You Can Help’ (2017) (accessible at <https://www.eff.org/deeplinks/2017/12/team-internet-far-done-whats-next-net-neutrality-and-how-you-can-help>).

— the internet — are put in jeopardy in exchange for short-term gains for a few telecoms companies. We hope this historic mistake will be corrected and eventually pave the way for real legislative protection of Net Neutrality in the United States.” — Thomas Lohninger, Executive Director, epicenter.works (Austria).

- “The internet must be free, open, and preserve the rights of all the users, without any kind of discrimination or repression or censorship of their rights. It was never intended as a tool to give power to IAPs or ISPs to be a ‘gatekeeper,’ privileging certain users or blocking others based on business or governmental interests. Let’s safeguard Net Neutrality!”— Housseem Kaabi, President, International Institute of Debate (Tunisia).
- “The ending of Net Neutrality in the U.S. could be the beginning of the end of the open, interoperable, free internet. It is now a question of how much, not if, freedom of expression online will be undermined around the world as a result of this short-sighted decision to enrich the entrenched near-monopolies who control internet access in the United States.” — Quinn McKew, Deputy Executive Director, ARTICLE 19 (United Kingdom).

In 2018, the FCC’s repeal of the net neutrality rules became official.¹² Net neutrality advocates challenged this decision, and, in 2019, the DC Circuit Court ruled in favour of the FCC and upheld its repeal of the 2015 Rules.¹³ In February 2020, despite attempts by various stakeholders, the DC Court of Appeals dismissed an appeal to reverse the repeal of the net neutrality rules.¹⁴

It is likely that the net neutrality debate will continue in the U.S., illustrating the stark contrast between those in favour of net neutrality and the economic interests of those who seek to curb it.

Practically engaging with net neutrality

As illustrated above, state and non-state actors often seek ways to depart from the principles of net neutrality and influence people’s freedom of expression online. In order to practically overcome the threats to net neutrality, there are two key considerations: first, the need to ensure adequate safeguards that preserve net neutrality; and second, the need to understand what limitations in relation to net neutrality are permissible. According to McDiarmid and Shears in the Net Neutrality Compendium:

“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

¹² New York Times, ‘Net Neutrality Has Officially Been Repealed. Here’s How That Could Affect You’ (2018) (accessible at <https://www.nytimes.com/2018/06/11/technology/net-neutrality-repeal.html>).

¹³ Washington Post, ‘Appeals Court Ruling Upholds FCC’s Cancelling of Net Neutrality Rules’ (2019) (accessible at <https://www.washingtonpost.com/technology/2019/10/01/appeals-court-upholds-trump-administrations-cancelling-net-neutrality-rules/>).

¹⁴ Engadget, ‘US Appeals Court Will Not Rule on Repealing Net Neutrality Laws’ (2020) (accessible at <https://www.engadget.com/2020/02/07/net-neutrality-us-appeals-court/>).

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.”¹⁵

As discussed above, states should preserve net neutrality in order 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.¹⁶

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:¹⁷

- 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.

Minimum standards and safeguards for network neutrality regulatory instruments:

The [Net Neutrality Compendium](#) in its Policy Statement on Network Neutrality further suggests the following safeguards for Network Neutrality regulatory instruments:

- **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:** ISPs should act in accordance with the principle of network neutrality. Any deviation from this principle may be considered as reasonable traffic management as long as it is necessary and proportionate to:
 - Preserve network security and integrity.

¹⁵ McDiarmid and Shears, ‘The Importance of Internet Neutrality to Protecting Human Rights Online’ in *Net Neutrality Compendium* at 31-32.

¹⁶ *Id* at 38.

¹⁷ *Id* 38-41.

- Mitigate the effects of temporary and exceptional congestion, primarily by means of protocol-agnostic measures or, when these measures do not prove practicable, by means of protocol-specific measures.
- Prioritise emergency services in the case of unforeseeable circumstances or force majeure.
- **Law enforcement:** None of the foregoing should prevent ISPs from giving force to a court order or a legal provision in accordance with human rights norms and international law.
- **Transparent traffic management:** ISPs should publish meaningful and transparent information on characteristics and conditions of the internet access services they offer, the connection speeds that are to be provided, and their traffic management practices, notably with regard to how internet access services may be affected by simultaneous usage 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 accordance with privacy and data protection obligations and subject to clear legal protections.
- **Implementation:** The competent national authorities should promote independent testing of internet traffic management practices, ensure the availability of internet access and evaluate the compatibility of internet access policies with the principle of network neutrality, as well as with the respect of human rights norms and international law. National authorities should publicly report their findings. Complaint procedures to address network neutrality violations should be available and violations should attract appropriate fines. All individuals and stakeholders should have the possibility to contribute to the detection, reporting and correction of violations of the principle of network neutrality.

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](#) 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.”

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.¹⁸ 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.

¹⁸ 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, emphasise state duties to ensure environments that enable business respect for human rights.
- **In relation to 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.

Conclusion

Developing countries continue to face challenges in relation to net neutrality and the suggestion that some access is better than no access. But the international human rights framework is clear, and it is necessary for civil society actors and human rights litigators to ensure that net neutrality is protected through lobbying states, submitting complaints to regulators, strategic litigation and public advocacy. The goal is equal opportunity in access.

Intermediary Liability

Internet intermediaries – an overview

An internet intermediary is a broad, constantly developing term. 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”. They fulfil a variety of functions, including 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.¹⁹ Examples of internet intermediaries range from ISPs and web hosting companies who provide the infrastructure, to search engines and social media platforms, who provide content and facilitate communication.²⁰ Simply put, “internet intermediaries are the pipes through which internet content is transmitted and the

¹⁹ Media Defence above n 2 at 6.

²⁰ ARTICLE 19, “Internet intermediaries: Dilemma of Liability” (2013) at 3 available at https://www.article19.org/data/files/Intermediaries_ENGLISH.pdf. See further Li, ‘Beyond Intermediary Liability: The Future of Information Platforms’ Yale Law School *Information Society Project* (2018) at 9 (accessible at https://law.yale.edu/sites/default/files/area/center/isp/documents/beyond_intermediary_liability_-_workshop_report.pdf).

storage spaces in which it is stored, and are therefore essential to the functioning of the internet.”²¹ 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”.²²

It is not difficult to create a link between internet intermediaries and the advancements of an array of human rights. As the 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.”

Internet intermediary liability

Given the roles intermediaries are playing in society, particularly in relation to the myriad of implicated rights, it is imperative to understand their legal liability. The [Association for Progressive Communications \(APC\)](#) explains that internet intermediary liability refers to the liability of intermediaries for illegal or harmful activities performed by users through their services, where they have an obligation to prevent the occurrence of unlawful or harmful activity by users of their services – failing which may lead to legal consequences such as orders to compel or criminal sanctions. In a [2012 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.

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.

While intermediary liability can be associated to 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

²¹ Id at 6.

²² Riordan, ‘The Liability of Internet Intermediaries’ DPhil thesis, Oxford University (2013) at 1 (accessible at https://ora.ox.ac.uk/objects/uuid:a593f15c-583f-4acf-a743-62ff0eca7bfe/download_file?file_format=pdf&safe_filename=THESIS02&type_of_work=Thesis).

expression and access. The legal liability of intermediaries has a direct impact on users' rights. 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, broad immunity model, and the safe-harbour model.

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](#) documents laws around the world that govern internet intermediaries and shape users' digital rights. It provides both basic and advanced tools to search for and visualise how legislation, decisions and public policies are evolving globally. It has captured the following in relation to China:

- 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.

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 for this model, which protects intermediaries from liability for illegal behaviour by users from liability 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.

Safe-harbor model

The safe-harbor model seemingly adopts a middle-ground approach. Otherwise known as conditional liability, this approach gives intermediaries immunity, provided they comply with certain requirements. Through this approach, as explained in the UNESCO Report, intermediaries are not required 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 take down procedures’, which can be content or issue-specific. There are mixed views on this; for some it is a fair middle-ground; for others, it is a necessary evil to guard against increased filtering or a complete change to the intermediary landscape.²³ 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.

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. An immediate response to takedown notices is necessary, failing which the immunity from liability is forfeited.

Concerns have been noted regarding South Africa’s framework, like most concerns around the safe harbour approach: 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.²⁴

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 where internet

²³ 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 at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3344213).

²⁴ See further Comninos, ‘Intermediary liability in South Africa’ (2012) (accessible at https://www.apc.org/sites/default/files/Intermediary_Liability_in_South_Africa-Comninos_06.12.12.pdf). See also Rens, ‘Failure of Due Process in ISP Liability and Takedown Procedures’ in *Global Censorship, Shifting Modes, Persisting Paradigms* (2015) (accessible at https://law.yale.edu/sites/default/files/area/center/isp/documents/a2k_global-censorship_2.pdf).

intermediaries are left to their own devices to determine what is good or what is 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, along with previous guides, 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.”²⁵ 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.

Applicable international human rights standards and current international best practices

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 evoked. 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 [Protect, Respect and Remedy’ 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.

In the 2019 [Joint Declaration: Challenges to Freedom of Expression in the Next Decade](#), it was observed that:

“private companies have responsibilities to respect human rights and remedy violations, and that addressing the challenges outlined above requires multi-stakeholder support and the active engagement of State actors, media outlets, intermediaries, civil society and the general public.”

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.

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 and narrow test, one that must:

- Be provided by law.

²⁵ Media Defence above n 2 at 28.

- Pursue a legitimate aim.
- Conform to the strict tests of necessity and proportionality.²⁶

Laws and 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 necessitated in jurisdictions that adopt the strict liability model and the safe-harbor model. The strict liability model can be easily tested and found to be compliant. The safe-harbor model presents a slightly more in-depth engagement in order to determine compliance, with the [Kenyan Copyright \(Amendment\) Bill, 2017](#), providing a useful example to illustrate the application of the applicable tests.

Kenyan Copyright (Amendment) Bill

The Bill proposes substantial amendments to the current Copyright Act, and of particular relevance is the proposal in relation to intermediary liability. A key feature of the Bill is the introduction of the safe-harbor approach. The Bill provides for “conduit” safe-harbors and “caching” safe harbors. The former, per section 35A(1)(a), protects 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 section 35A(1)(b), intermediaries are protected if their role is related to content storage that is “automatic, intermediate and temporary”. This protection is conditional upon the removal of content following a take-down notice.²⁷

There have been calls for clarity and improved notice-and-takedown procedures and concerns that the Bill falls short of international standards on freedom of expression. [ARTICLE 19](#) lists five problems with the Bill in terms of notice-and-takedown procedures:

- **Lack of proportionality:** criminal sanctions are imposed on intermediaries who fail to remove content. As discussed above, this causes intermediaries to lean toward censorship and blocking, which infringes freedom of expression.
- **Lack of clarity:** the procedures are vague and do not provide clarity on the issue of counter notices.
- **Lack of due process:** there is no mention of judicial review or appeal mechanisms. There is also no requirement notifying the content publisher of the alleged infringement. The 48-hour time frame for the removal of content is too short to allow for the submission of a counter notice.

²⁶ 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 at <https://www.osce.org/netfreedom-guidebook?download=true>)

²⁷ 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’ (2019) The African Journal of Information and Communication (accessible at https://journals.co.za/docserver/fulltext/afjic_n23_a5.pdf?expires=1581473231&id=id&acname=quest&checksum=1AD5DE6F4FD5EA3A0CB45F94F3335E67).

- **Lack of transparency:** there is no obligation to maintain records of takedown requests or provide access to such records.
- **Severe sanctions:** the harsh sanctions for false take down notices are disproportionate to the purpose of deterring such.

It is apparent that the necessity and proportionality legs of the test prove to be the sticking points in relation to this Bill. While the safe-harbor method might serve a legitimate aim, if the regulations that guide it are not clear, necessary and proportionate, then they unjustifiably limit the right to freedom of expression.

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 must not be required to be restricted without an order by 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.

These principles have been relied on to test state rules and to gauge whether the legal frameworks regarding intermediary liability are adequate. This was done most recently in 2019 in India by [the Centre for Internet and Society](#) who submitted a report to the Indian government comparing the Manila Principles to the draft Information Technology [Intermediary Guidelines (Amendment) Rules], 2018. These submissions provided useful guidance on the provisions that were unaligned with the Manila Principles by highlighting the areas that have the potential to infringe upon the right to freedom of expression. The submission further provided recommendations to assist the Indian government in ensuring the regulations are compliant. The submissions are a useful illustration of the significance of these principles, as well as a useful resource for others who seek to test domestic legislation against international best practices.

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, in encouraging innovation and creativity.²⁸ 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.”²⁹

Right To Be Forgotten

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 information taking into account the right to privacy weighed against public interest considerations.³⁰ This right was given prominence following the 2014 Court of Justice of the European Union (CJEU) judgement in [Google Spain SL v Agencia Española de Protección de Datos](#) (Google Spain case).³¹ 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 for 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, such as [Access Now](#) concerned.

²⁸ 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 at https://law.yale.edu/sites/default/files/area/center/isp/documents/new_controversies_in_intermediary_liability_law.pdf).

²⁹ Li, “Beyond Intermediary Liability: The Future of Information Platforms” Yale Law School *Information Society Project* (2018).

³⁰ See Media Defence above n 2 at 35.

³¹. For a fuller case note see Media Defence above n 2 at 35.

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.”³² Others were more circumspect, noting concerns that when information is delisted it affects other fundamental rights, including freedom of expression and the right to receive and impart information and ideas.³³

Evolution of the right to be forgotten

Following from the abovementioned judgment, the right to be forgotten has been recognised in domestic contexts,³⁴ regional legislation and again by the CJEU. In 2018, Google published its [Transparency Report](#) revealing that 43% of the 2.4 million requests from users to be “forgotten” have been successful. 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 of 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 the 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 jurisprudential developments 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.

By way of background, in 2015, the French Data Protection Agency (CNIL) requested Google to globally remove information concerning a data subject. Google refused and limited its

³² Cook, ‘The Right to be Forgotten: A Step in the Right Direction for Cyberspace Law and Policy’ (2015) 6 *Journal of Law, Technology & the Internet* 121 at 121-123 (accessible at <https://scholarlycommons.law.case.edu/jolti/vol6/iss1/8/>).

³³ Kulk and Borgesius, ‘Freedom of expression and ‘right to be forgotten’ cases in the Netherlands after Google Spain’ (2015) 2 *European Data Protection Law Review* 113 at 116 (accessible at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2652171). See also ARTICLE 19, ‘The “Right to be Forgotten”: Remembering Freedom of Expression’ (2016) (accessible at https://www.article19.org/data/files/The_right_to_be_forgotten_A5_EHH_HYPERLINKS.pdf).

³⁴ See Media Defence above n 2.

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 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 such as [ARTICLE 19](#) and the [Electronic Frontier Foundation](#) have welcomed the ruling of the CJEU:

“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.”

The careful navigation of balancing privacy rights against freedom of expression will continue to pose challenges as the digital landscape continues to evolve.³⁵

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. States are still entitled to, at their discretion, develop the content of this right within their respective jurisdictions, and are still at liberty, as acknowledged by the CJEU, to adopt different approaches when balancing the right to privacy and the protection of personal data, on the one hand, and the right of access to information of internet users on the other – provided such an approach is compliant with international human rights norms.

³⁵ 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 at <https://www.article19.org/wp-content/uploads/2017/12/Google-v-CNIL-A19-intervention-EN-11-12-17-FINAL-v2.pdf>).

Opportunities and risks

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 some control of their digital identities. The majority of online personal information has no bearing on public interest considerations and has far more intrinsic value to the individual than 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.”³⁶ 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 [2017 Transparency Report](#) provided some guidance on how it has dealt with requests, providing examples of some of the outcomes of requests for erasure. Some responses to politician’s requests stated “[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 enforcement of rights and the adverse effect this can have on the right to freedom of expression.³⁷ A lack of cogent regulatory safeguards can result in search engines becoming the “judge, jury and executioner” of the right to be forgotten.³⁸ 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.³⁹ The [Electronic Frontier Foundation](#) expressed concern that the “ambiguous responsibility upon search engines” will censor the internet.

³⁶ Michael L Rustad & Sanna Kulevska, “Reconceptualizing the Right to Be Forgotten to Enable Transatlantic Data Flow” (2015) 28 *Harvard Journal of Law and Technology* 349 at 373 (accessible at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2627383).

³⁷ *Id.*

³⁸ Forde, ‘Implications of the Right to be Forgotten’ (2015) 17 *Tulane Journal of Technology and Intellectual Property* 83 at 113 -114 (accessible at <https://journals.tulane.edu/TIP/article/view/2652>). See further Lindsay ‘The ‘Right to be Forgotten’ by Search Engines under Data Privacy Law: A Legal Analysis of the Costeja Ruling’ (2014) 6 *Journal of Media Law* 159 at 173 – 174.

³⁹ Kuczerawy & Ausloos, ‘From Notice-and-Takedown to Notice-and-Delist: Implementing Google Spain’ (2016) 14 *Colorado Technology Law Journal* 219 (accessible at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2669471).

Ensuring adequate safeguards when implementing 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 comprehensive data protection legislation to avoid interference with human rights.
- Competent judicial authorities should interpret standards for determining what is de-listed.
- The right to de-list must be limited in scope and application.
- Search engines must be transparent about when and how they comply with de-listing requests.
- Users must have easy access to a remedy.

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, it is likely that these tensions will persist. Provided public interest overrides are prioritised and adequate safeguards are put place, there can be some degree of consonance.

Monitoring Obligations of Search Engines and Platforms

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”.⁴⁰ 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 to 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, and in order to so, content must not be moderated in a manner that leads to censorship and free speech

⁴⁰ 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 at <https://www.apc.org/en/pubs/reorienting-rules-rights-summary-report-online-content-regulation-special-rapporteur-promotion>).

violations.⁴¹ 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.⁴²

The [2018 Report](#) by the UNSPR noted key concerns regarding content regulation:

“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.⁴³ 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.⁴⁴ In instances where there are strict monitoring obligations, it is likely that monitoring will become the norm, opening intermediaries to automatic and direct liability.⁴⁵ Monitoring obligations raise concerns in 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.”⁴⁶

Further to the above, there has been a trend, akin to that of the right to be forgotten, where states demand global removal of content that violates domestic law.⁴⁷ 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 in relation to copyright infringements. However, it is growing at an unprecedented rate, causing grave concern for

⁴¹ Langvardt, ‘Regulating Online Content Moderation’ *Georgetown Law Journal* 106 (2018) 1354 at 1354-1359 (accessible at <https://www.law.georgetown.edu/georgetown-law-journal/wp-content/uploads/sites/26/2018/07/Regulating-Online-Content-Moderation.pdf>).

⁴² 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 at <https://www.ohchr.org/Documents/Issues/Opinion/ContentRegulation/APC.pdf>).

⁴³ Frosio, ‘From Horizontal to Vertical: an Intermediary Liability Earthquake in Europe’ *Centre for International Intellectual Property Studies Research Paper* (2017) at 12 (accessible at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3009156).

⁴⁴ *Id.*

⁴⁵ *Id.*

⁴⁶ Stanford Law, ‘Monitoring Obligations’ (2017) (accessible at <https://wilmap.law.stanford.edu/topics/monitoring-obligations>).

⁴⁷ 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)*.

free expression.⁴⁸ Judgments of the European Court of Human Rights (**ECtHR**) provide useful insight into the issues regarding online platforms and liability for users' comments.

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. By way of background, an Estonian newspaper, Defli, 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 was 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. Defli 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 reconsideration, concluding that the lower court had erred in its finding in relations to 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 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 the 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 Defli;
- 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.

⁴⁸ Frosio, 'The Death of 'No Monitoring Obligations' A Story of Untameable Monsters' *JIPITEC* (2017) (accessible at https://www.jipitec.eu/issues/jipitec-8-3-2017/4621/JIPITEC_8_3_2017_199_Frosio).

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. As a reminder, here we provide a short summary of a censorial attempt that targeted intermediaries.”

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 which 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 that 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:

“However, in 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 the 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.”

It has been noted that there are some inconsistencies in the ECtHR's approach to online liability.⁴⁹ However, it does appear that the shift away from the *Delfi* reasoning was a shift in the right direction.⁵⁰ 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.⁵¹

Efforts to address content moderation at the global level

UN Human Rights Office of the High Commissioner has noted:

“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

⁴⁹ 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).

⁵⁰ Id at 3. See also Media Defence 'European Court clarifies intermediary liability standard' (2016) (accessible at <https://www.mediadefence.org/news/european-court-clarifies-intermediary-liability-standard>).

⁵¹ 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 at https://cadmus.eui.eu/bitstream/handle/1814/62005/RSCAS%202019_20.pdf?sequence=1&isAllowed=y).

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 the glaring lack of adequate rules, guidelines, procedures and remedies in relation to the current practices of content moderation that are cause for concern.⁵² It is clear that a human rights framework ought to be guiding the principles for company content moderation.

Guidance from the UNSR on ensuring compliance with human rights standards when online content is being moderated

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.

UNSR guidance on the processes for company moderation and related activities

These [guidelines and recommendations](#) provided further guidance on the processes for company moderation and related activities:

⁵² ARTICLE 19, ‘Social Media Councils: Consultation’ (2019) (accessible at <https://www.article19.org/resources/social-media-councils-consultation/>).

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 comment 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 acknowledgment to settlements related to reputational or other harms.
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.

The UNSR goes on to provide specific recommendations, imparting the urgent need for “radical transparency, meaningful accountability and a commitment to remedy in order to protect the ability of individuals to use online platforms as forums for free expression, access to information and engagement in public life”.