Wrongful moderation: regulation of internet intermediary service provider liability and freedom of expression

Klos, M.

Wrongful moderation

Regulation of internet intermediary service provider liability and freedom of expression

Michael Klos

Universiteit Leiden

Wrongful moderation

Regulation of internet intermediary service provider liability and freedom of expression

Proefschrift

ter verkrijging van
de graad van doctor aan de Universiteit Leiden,
op gezag van rector magnificus prof.dr.ir. H. Bijl,
volgens besluit van het college voor promoties
te verdedigen op woensdag 21 september 2022
klokke 10.00 uur

door

Michael Klos
geboren te Zoetermeer
in 1991

Promotor:	prof. dr. P.B. Cliteur
Co-promotor:	mr. dr. G. Molier
Promotiecommissie:	prof. dr. A. Ellian
	dr. A. Kuczerawy (KU Leuven, België)
	prof. dr. A.N. Guiora
	(University of Utah, Salt Lake City, USA)
	dr. M.R. Leiser
	prof. dr. B.R. Rijpkema

Table of Contents

Foreword

> Life before death. Strength before weakness. Journey before destination. That was their motto, and was the First Ideal of the Immortal Words.[1]

The quote above this chapter is one of the few references to a book of fiction in this thesis. This quote, however, is well-fitting for my process of writing a dissertation. Authoring a dissertation is not always easy and certainly not always fun. Doing research is (from time to time) hard, especially when a large portion of this research takes place during a global pandemic. During the four years of a PhD-trajectory, this aspiring doctor learned a lot about himself, his research subject, and life at the university. How hard writing a dissertation may be, it is impossible to stop. It is impossible to abandon the research you are fascinated about and leave the topics you love behind – even for a short holiday. Writing a dissertation may be, however, confronting. This aspiring doctor became aware of his strengths but mainly of his weaknesses. Sometimes it is easy to linger with these weaknesses instead of finding the strength to go forward. My firm conviction is that the only way to survive a PhD-trajectory is by viewing it as a journey worth venturing into for its own sake – rather than fixating on the ends: the ceremonial defence of the dissertation (which is, of course, a scary thing in itself).

While I am writing this foreword in February 2022, I am looking back on the last four years and what is to come. A journey it was – and still is. During my PhD, I have had the opportunity to dry run many of my ideas by presenting them at conferences. In addition, I published some of my ideas in edited volumes and journals. I have explored the international dimension of internet intermediary regulation in the edited volume *Democracy and Globalization: Legal and Political Analysis on the Eve of the 4th Industrial Revolution*,[2] how the European Union regulates internet companies in the edited volume *Reflections on democracy in the European Union*[3] and the relationship between internet companies and the open society in *The Open Society and its Closed Communities*.[4] In the *Nederlands Juristenblad* (*NJB*), I published my first contribution on 'wrongful moderation'.[5] The *Nederlands Tijdschrift voor de Mensenrechten* (*NTM*) published some of my first reflections on the European Union's Digital Services Act.[6] I am grateful that, for now, I can continue my research in the form of a COVID-19 disinformation project financed by the NWO.

Looking back on the journey of the last four years, many of these publications were followed by new developments, requiring me to rethink and update substantial portions of my arguments. Some early published material formed the basis for parts of this dissertation. For

---

[1] B. Sanderson, *The Way of Kings*, New York, Tor Books, 2010, p. 924.

[2] M. Klos, 'Westphalian Sovereignty and the 4th Industrial Revolution: In Search of Legitimate Governmental Control over Online Content', *in* C. Sieber-Gasser & A. Ghibellini (Eds.), *Democracy and Globalization: Legal and Political Analysis on the Eve of the 4th Industrial Revolution*, Cham, Springer International Publishing, 2021, doi:10.1007/978-3-030-69154-7, pp. 81-121.

[3] M. Klos, 'Tackling Online Freedom of Expression: the European Approach', *in* A. Ellian & R. Blommestijn (Eds.), *Reflections on democracy in the European Union*, The Hague, Eleven International Publishing, 2020, pp. 27-56.

[4] M. Klos, 'Closed Online Communities and Freedom of Speech', *in* A. Ellian & P. Cliteur (Eds.), *The Open Society and its Closed Communities*, The Hague, Eleven, 2021, pp. 173-215.

[5] M. Klos, "Wrongful moderation': Aansprakelijkheid van internetplatforms voor het beperken van de vrijheid van meningsuiting van gebruikers', *Nederlands Juristenblad*, 2020/2976.

[6] M. Klos, 'De Digital Services Act: implicaties voor het recht op vrijheid van meningsuiting van gebruikers van onlineplatforms', *NTM/NJCM-bull.*, 2021/13.

example, Chapters 3 and 4 are a much more expanded version of the argument made in *Tackling Online Freedom of Expression: The European Approach* (2020). When appropriate, a footnote is placed with references to earlier publications. The topic of this dissertation is a fast-moving target, meaning that it is fascinating to research. Many people are working on similar themes, and the developments follow each other quickly. This dissertation includes developments till at least September 2021.

In front of the reader lies a thesis that could easily have been ten times as thick. I have chosen not to. A dissertation is the first test of a doctoral student's ability for scholarly work. I hope that the reader sees this thesis as such proof because I do not consider it the final destination. In the coming years, I hope to contribute to (scholarly) debates on themes such as disinformation. This work forms an introduction to the legal dimension of the internet. This work deals with the freedom of expression on the Internet from a state perspective. Although internet companies have been increasingly seen as a threat to users' freedom of expression, I argue in this work that such discussions are not meaningful without discussing how states influence users' freedom of expression by regulating these companies. In many cases, it is not the company that is purely responsible for limiting users' freedom of expression but the state that bears responsibility indirectly.

Acknowledgements

The road to university was a long one. I had to pile up studies within the Dutch education system before I was allowed to start at university. It was thus not self-evident that I would be allowed to start a PhD program. I am thankful that my supervisors, Paul Cliteur and Gelijn Molier, were willing to offer me an academic and a true second home at Leiden University. I also wish to thank Afshin Ellian, chairman of the Department of Jurisprudence, for the many opportunities to contribute to the courses taught by our department and the possibility to work on multiple (academic) projects. I would also like to thank my close colleagues. I wish to thank Arie-Jan Kwak for his down-to-earthiness. Without his relativising words, I would have spent many weeks on those matters that do not matter. I would like to thank Bastiaan Rijpkema for his confidence in starting a project on COVID-19 disinformation with me and Sarah de Lange (UVA, Amsterdam). In the last months of completing my dissertation, my conversations with them have been particularly educating. I also wish to thank Tessa van Buchem. Her puns were always on time. I also wish to thank Erwin Dijkstra, Jip Stam, and Jorieke Manenschijn for the interesting conversations. Without my fellow PhD candidates, the journey would be a lonely one.

In particular, I must thank my wife, Nathalie Schnabl-Klos. She did not understand it all when I wanted to work every night and weekend to work on this dissertation. She was right. Without her support, I would never have completed this thesis. I would also like to thank my son, Christopher Albertus Klos, who came into my life in May 2021. He provided me with enough sleepless nights so that I could never worry about finishing my dissertation. I was simply too tired for that.

Introduction

A post with a wild conspiracy theory about COVID-19 vaccines causing infertility on LinkedIn. A photograph depicting Black Pete during the Dutch Sinterklaas celebration on Facebook. A video shot at a beach shows uncovered female nipples on Instagram. While these examples may not seem to have much in common, they are examples that are (or could become) subject to the regulation of the providers offering social media platforms.[7] In these cases, the contents of the post (the conspiracy theory), the photograph (Black Pete), and the video (a female nipple) are part of the regulation on these social media platforms. Such regulation (whether enacted by the state or themselves) enables platform providers to (for example) label or remove the content in question. Social media platforms of which YouTube, Facebook, Instagram, LinkedIn and TikTok form well-known examples can delete, alter, restrict, hide, or otherwise remedy the (perceived) harmful content of user-provided information. These interventions are called moderation: interventions to remedy the harmful effects of prohibited content – irrespective of who (the state or the platform) prohibited the content.[8] Next to moderation, social media platforms recommend, organise, and prioritise information based on its content. For example, social media platforms may prioritise popular information based on previous interactions with this content or recommend information that fits the user's profile. Such efforts are called curation.[9] As discussed in Paragraph 2.2, moderation and curation are examples of provider-enacted regulations that allow users to submit their information and retrieve information submitted by other users.[10]

Social media platforms are just one functionality these so-called internet intermediary service providers (hereafter: 'provider' or 'providers') offer. Providers can offer many more intermediary functions to allow users to publish their own information (so-called user-provided information).[11] Some of these providers offer services dedicated to a niche (such as Boardgame Geek for tabletop games) or towards a specific goal (such as Wikipedia as an online encyclopaedia), or for specific content (such as GitHub for computer code). Of course, some providers do not offer the possibility for users to provide information of their own. For example, Netflix, Disney+ or Amazon Prime offer a catalogue of series and movies without enabling users to provide their own. Music streaming services such as Spotify and Deezer follow a similar approach. Other services may have limited control over the content of user-provided information while they function as an intermediary. These providers only fulfil a technological role in transferring information from the user to other providers or between two other providers. A well-known

---

[7] For example, Meta, 'Adult Nudity and Sexual Activity', *Facebook Transparency Center,* 23 December 2021, available at facebook.com/communitystandards/adult_nudity_sexual_activity (retrieved on 14 February 2022); Meta, 'Hate Speech', *Facebook Transparency Center,* 24 November 2021, available at facebook.com/communitystandards/hate_speech (retrieved on 14 February 2022); Twitter, 'COVID-19 misleading information policy', *Twitter Help Center*, available at help.twitter.com/en/rules-and-policies/medical-misinformation-policy (retrieved on 15 February 2022).

[8] E. Goldman, 'Content Moderation Remedies', *Michigan Technology Law Review*, Vol. 28, No. 1, 2021, doi:10.36645/mtlr.28.1.content, pp. 5-6.

[9] Council of Europe, 'Content moderation: best practices towards effective legal and procedural frameworks for self-regulatory and co-regulatory mechanisms of content moderation' (Guidance Note adopted by the Steering Committee for Media and Information Society (CDMSI) at its 19th plenary meeting, 19-21 May 2021), Strasbourg, Council of Europe, 2021, available at rm.coe.int/0900001680a2cc18, p. 11.

[10] See Paragraph 1.1 for a more eleborate exposition of what an internet intermediary service provider is.

[11] See Paragraph 1.1 for a definition of user-provided information.

example of such a provider is a so-called internet service provider (hereafter: ISP) that offers an internet connection to users.

Many (but certainly not all) providers have an extensive influence on what users can and cannot provide to their service. For example, providers can set the rules for their service by enacting community standards. In addition, providers can 'hard-code' what users can and cannot do by limiting technological possibilities, such as only allowing videos and no photographs. Violations of the community standards may be remedied by moderation – often by removing the information with violating content. Such influence of the provider over user-provided information may lead to friction between the user and the provider. For example, the user may not have a similar view on the desirability of some content as the provider. In addition, the enforcement of standards may lead to friction. A human moderator may erroneously remove user-provided information because the moderator misses the irony or sarcasm in the message. The same argument is valid for automatic moderation systems that miss the nuance of a message. Some accuse providers that curate user-provided information of being biased towards some political content because some political content is argued to be shown more than others.[12]

Especially moderation efforts are critically scrutinised because of their impact on the users' freedom of expression rights.[13] Moderation may lead to removing user-provided information with content that violates the rules. Additionally, moderation may result in (temporary) suspensions of user accounts which means that the user (during this suspension) cannot use the service.[14] In other cases, a provider may terminate a user account after a more severe or repeated violation which means that the user cannot return.[15] Termination of user accounts wholly cuts off the user from a channel and thus severely impacts the users' possibility to reach their audience. For example, the ban of Donald Trump by major social media platforms is still effective at the time of writing.[16]

Moderation carried out by providers often comes with fundamental disagreement about the reach of freedom of expression rights on a privately held platform. However, providers enacting content moderation norms are only a part of the story. Providers may impose such norms

---

[12] For example, US Department of Justice, 'Department of Justice's Review of Section 230 of the Communications Decency Act of 1996', *US Department of Justice,* 2020, available at justice.gov/archives/ag/department-justice-s-review-section-230-communications-decency-act-1996 (retrieved on 15 February 2022).

[13] See, for example, Recital 41 and 42 of Commission Proposal COM(2020) 825 final of 15 December 2020 Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC (*Digital Services Act*), pp. 26-27.

[14] M.L. Jones, 'Silencing Bad Bots: Global, Legal and Political Questions for Mean Machine Communication', *Communication Law and Policy*, Vol. 23, No. 2, 2018, doi:10.1080/10811680.2018.1430418, p. 189; J. Mchangama, N. Alkiviadou & R. Mendiratta, 'Rushing to Judgment: Are Short Mandatory Takedown Limits for Online Hate Speech Compatible with The Freedom of Expression?', *The Future of Free Speech Project,* January 2021, available at futurefreespeech.com/rushing-to-judgment-are-short-mandatory-takedown-limits-for-online-hate-speech-compatible-with-the-freedom-of-expression (retrieved on 15 February 2022), p. 14.

[15] For example, Donald Trump, see M. Kruse, 'Can Donald Trump Survive 'Virtual Impeachment'?', *Politico,* 8 January 2021, available at politico.com/news/magazine/2021/01/08/donald-trump-capitol-insurrection-riot-impeachment-456352 (retrieved on 15 February 2022).

[16] However, the Oversight Board critcally reviewed Meta its practice on Facebook and Instagram. See Oversight Board, 'Case decision 2021-001-FB-FBR', *Oversight Board,* 5 May 2021, available at oversightboard.com/decision/FB-691QAMHJ (retrieved on 15 February 2022).

because they must align to local laws[17] or wish to adhere to human rights standards.[18] Providers, for example, may restrict hate speech because this is illegal in a substantial portion of jurisdictions or prohibit discriminatory content because of international human rights standards – even when such content is legal in some jurisdictions.[19] Providers, however, may also set such standards for reasons of their own. For example, advertisers may not wish to affiliate themselves with hateful content, or the provider may wish to foster civil discussions on their service by prohibiting insults. Irrespective of where these standards come from, such influence over user content makes providers, as Kate Klonick names them, 'the New Governors of online speech'.[20]

The policies enacted by these 'new governors' cannot necessarily count on the consent of the users governed by these standards, which raises questions about the legitimacy of such interventions by private actors. Some policies are likely accepted as necessary by a broad userbase,[21] while other policies can count on criticism.[22] For example, Meta (the owner of Facebook, Instagram, and WhatsApp) has a history of struggling with its policies. Its prohibition on depicting female nipples was (and is) met with heavy criticism.[23] While Facebook changed its policy over time, the female nipple is still treated differently from the male nipple.[24] Less controversial from an international perspective is Meta's ban on blackface as a part of the hate speech policy in 2020.[25] While this policy was internationally welcomed as a necessary step against racism, prohibiting blackface also restricted users from sharing images of Black Pete. Black Pete is (or better said: was) a fictional character making its appearances around the Dutch holiday known as Sinterklaas. This character is no longer welcome on Facebook.[26] Banning Black Pete by Meta and other providers was met with severe criticism by pro-Black Pete groups in the Netherlands.[27] However, the

---

[17] I borrowed 'alignment' as terminology for compliance with local regulation from Milton Mueller, see M. Mueller, *Will the Internet Fragment?*, Cambridge, Polity Press, 2017, p. 49.

[18] As is the case with hate speech prohibitions which are to some extent aligned to national legislation. Compare, for example, Council Framework Decision 2008/913/JHA of 28 November 2008 on combating certain forms and expressions of racism and xenophobia by means of criminal law, *OJ L 328*, *6.12.2008* (data.europa.eu/eli/dec_framw/2008/913/oj); Twitter, 'Hateful conduct policy', *Twitter Help Center*, available at help.twitter.com/en/rules-and-policies/hateful-conduct-policy (retrieved on 15 February 2022).

[19] For example, the 'Brussels Effect', see A. Bradford, 'The Brussels Effect', *Northwestern University Law Review*, Vol. 107, No. 1, 2012.

[20] K. Klonick, 'The New Governors: The People, Rules, and Processes Governing Online Speech', *Harvard Law Review*, Vol. 131, No. 6, 2018, p. 1603.

[21] It is inconceivable that a very large proportion of users would not agree to policies against, for example, child sexual abuse material.

[22] Some categories of material are politicised. The removal of such material (female nipples, images of self-harm or, for example, anorexia nervosa) can lead to a loss of legitimacy as such material is not allowed to be displayed, see T. Gillespie, *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions that Shape Social Media*, New Haven, Yale University Press, 2018, pp. 170-171.

[23] Gillespie, 2018, *Custodians of the Internet*, pp. 141-169.

[24] K. Klonick, 'Facebook Under Pressure', *Slate,* 13 September 2016, available at slate.com/technology/2016/09/facebook-erred-by-taking-down-the-napalm-girl-photo-what-happens-next.html (retrieved on 15 February 2022); Meta, 2021, 'Adult Nudity and Sexual Activity'.

[25] Meta, 2021, 'Hate Speech'.

[26] S. Hulsen, 'Het zwartepietbeleid van sociale media: 'YouTube trekt rookgordijn op'', *nu.nl,* 30 November 2019, available at nu.nl/tech/6014528/het-zwartepietbeleid-van-sociale-media-youtube-trekt-rookgordijn-op.html (retrieved on 15 February 2022).

[27] RTL Nieuws, 'Facebook verwijdert groepen met Zwarte Piet, oprichters zoeken alternatief', *RTL Nieuws,* 4 September 2020, available at rtlnieuws.nl/nieuws/nederland/artikel/5181622/facebook-zwarte-piet-blackface-

Oversight Board established by Meta to adjudicate upheld the policy after a Dutch user complained of removing content depicting Black Pete. The majority of the Oversight Board pointed out that '[m]odified "Piet" traditions that have abandoned the use of blackface are […] not affected by the Hate Speech Community Standard'.[28] The Board argued that "it is consistent with Facebook's human rights responsibilities to adopt operational rules and procedures that promote equality and non-discrimination."[29] The Board, briefed by international experts, balances freedom of expression rights against the right of non-discrimination.

Completely voluntary community guidelines freely enacted by the provider in question are not this dissertation's main interest. Instead, this dissertation focuses on the state's role in enacting such standards. Meta chose to prohibit the female nipple on Facebook and enact hate speech regulations prohibiting depictions of blackface. While these policies may (rightfully or not) limit freedom of expression rights, they result from private decision-making. In contrast, other providers allow the female nipple or may not prohibit blackface under hate speech policies.

However, it is necessary to be cautious when concluding that private companies can set the rules as they wish. While the standards in the policies rely on private law instruments such as contracts, these standards may result from governmental coercion instead of free private decision-making. Even when policymakers frame these policies as the result of self-regulation, they may not merely result from free choice.[30] Some policies are requested, demanded, or even legally required by governmental actors. For example, during the pandemic in 2020-2022, governmental actors sought to influence providers' policies by requesting them to act against disinformation and conspiracy theories undermining the government's response to the global pandemic.[31]

Thus, state actors may directly regulate providers to remove illegal or unlawful content of user-provided information, but they could also ask providers to regulate harmful content. States are, in this conduct, guided by (national) doctrine about the liability of (offline) information intermediaries (such as publishers or editors of books and newspapers), but also by liability regimes tailored explicitly to and enacted for these providers. The responsibility of providers for the illegal content of the information of their users is, for example, cast in legislation. However, there is also the possibility of non-legislative regulation by (non-binding) codes of conduct. Of interest to this

---

beheerders-racisme (retrieved on 15 February 2022); P. Sabel, 'Toezichtsraad: Zwarte Piet is raciaal stereotype en wordt terecht geweerd van Facebook en Instagram ', *De Volkskrant,* 13 April 2021, available at volkskrant.nl/nieuws-achtergrond/toezichtsraad-zwarte-piet-is-raciaal-stereotype-en-wordt-terecht-geweerd-van-facebook-en-instagram~b03f7cee (retrieved on 15 February 2022).

[28] Oversight Board, 'Case decision 2021-002-FB-UA', *Oversight Board,* 13 April 2021, available at oversightboard.com/decision/FB-S6NRTDAJ (retrieved on 15 February 2022).

[29] Oversight Board, 2021, 'Case decision 2021-002-FB-UA'.

[30] D. Citron, 'Extremist Speech, Compelled Conformity, and Censorship Creep', *Notre Dame Law Review*, Vol. 93, No. 3, 2018, p. 1070.

[31] Joint Communication JOIN(2020) 8 final of the European Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions of 6 June 2020 Tackling COVID-19 disinformation - Getting the facts right; European Commission, 'Code of Practice on Disinformation', *European Commission,* 2 December 2021, available at digital-strategy.ec.europa.eu/en/policies/code-practice-disinformation (retrieved on 14 February 2022); World Health Organization, 'Managing the COVID-19 infodemic: Promoting healthy behaviours and mitigating the harm from misinformation and disinformation', *World Health Organization,* 23 September 2020, available at who.int/news-room/detail/23-09-2020-managing-the-covid-19-infodemic-promoting-healthy-behaviours-and-mitigating-the-harm-from-misinformation-and-disinformation (retrieved on 15 February 2022).

dissertation is how these regulatory (legislative and non-legislative) instruments provide a legal incentive to providers to overregulate or underregulate user-provided information for illegal, unlawful, or otherwise considered harmful content.

As understood in this dissertation, a legal incentive means that the regulation "persuades parties to engage in certain conduct."[32] State legislation, for example, may provide a legal incentive to moderate illegal hate speech by imposing liability on providers when they fail to do so. Regulation, however, may also (accidental or purposely) persuade providers to do more or less than that. For example, state regulation requiring providers who engage in *some* moderation to remove *all* illegal hate speech may persuade the provider to cease all moderation to prevent liability. However, such state regulation may equally introduce the risk that the provider does much more than only moderating hate speech. Liability for illegal hate speech content may persuade the provider to remove all user-provided information that remotely touches upon hate speech content.[33] In sum, in this dissertation, a legal incentive that persuades providers to do less than the regulation targets is named underregulation. Overregulation, in contrast, occurs when the provider moderates content out of fear of legal liability – even when this content is not a part of the regulation targeting the provider. Underregulation means that the provider does not regulate the content explicitly targeted by the regulation but changes its conduct to circumvent liability. However, overregulation and underregulation do not occur when providers change their conduct for other reasons.

Providers are subject to different liability regimes. The liability regimes of the European Union (hereafter: EU) and the United States of America (hereafter: US) are the most noteworthy because of their regulatory importance.[34] In the EU, the e-Commerce Directive (2000)[35] and the coming Digital Services Act[36] (hereafter: DSA) offer the EU approach toward providers' liability for the content of user-provided information. However, the e-Commerce Directive does not contain provisions for when the provider becomes liable but only when the provider can rely on an exception from liability. As discussed in the coming chapters, the EU approach toward provider liability can be characterised as a conditional liability (or conditional immunity) regime because the provider's liability depends upon fulfilling specific criteria.[37] The US approach towards provider liability laid down in Section 230 of the Communications Decency Act[38] offers a different approach towards provider liability by exempting providers from liability for the content of user-provided information and moderation efforts. Scholars view the EU approach often as less favourable

---

[32] Wex, 'Incentive', *Legal Information Institute*, available at law.cornell.edu/wex/incentive (retrieved on 18 January 2022).

[33] Citron, 'Extremist Speech, Compelled Conformity, and Censorship Creep', *Notre Dame Law Review*, 2018, pp. 1052-1055; E. Goldman, 'The Complicated Story of FOSTA and Section 230', *First Amendment Law Review*, Vol. 17, 2019 (available at firstamendmentlawreview.org/volume-17), pp. 288-289.

[34] For example, the so-called "Brussels Effect", see Bradford, 'The Brussels Effect', *Northwestern University Law Review*, 2012; A. Bradford, *The Brussels Effect: How the European Union Rules the World*, New York, Oxford University Press, 2020.

[35] Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market (*Directive on electronic commerce*), *OJ L 178*, *17.7.2000* (data.europa.eu/eli/dir/2000/31/oj).

[36] Commission Proposal COM(2020) 825 final (*Digital Services Act*).

[37] See, especially, Paragraph 4.1.

[38] §230. Protection for private blocking and screening of offensive material, 47 USCA § 230 (West 2018, Westlaw Next through PL 116-91).

towards freedom of expression rights because it is argued to provide an incentive to overregulate illegal or unlawful content of user-provided information.[39] Because the exemptions to liability are knowledge-based, providers are agued to be given an incentive to remove content when they doubt its illegal or unlawful nature. Removal is the only way they are sure they are still exempted from liability, raising freedom of expression issues.[40] On the other hand, the Section 230 approach towards provider liability is more favourable for providers but not necessary for the freedom of expression rights of the users of the service. Immunity for the provider does not mean immunity for the user. Immunity, in other words, does not require providers to take into account the rights of users.[41]

The EU approach, however, differs from the US approach. The possibility of enacting content-based restrictions is severely limited in the US due to the broad constitutional protection of freedom of expression rights.[42] In the EU, freedom of expression rights does not necessarily limit content-based restrictions. Therefore, in the EU, it is possible to regulate terrorist content, hate speech, and even disinformation. The possibility of enacting new regulation allows governmental actors to threaten regulation when the provider fails to self-regulate user-provided information with content that does not qualify as illegal or unlawful but as harmful.[43] At the same time, providers that wish to self-regulate harmful but legal content may be exposed to legal action of their users when the liability regime does not offer an exemption for liability that comes from moderation.[44]

Research question

Therefore, the research question that is central to this dissertation is:

> To what extent do the different liability regimes (laid down in the e-Commerce Directive in the EU and Section 230 in the US) applicable to providers provide a legal incentive to overregulate or underregulate user-provided information by moderating or curating information based on the fact that its content can be considered illegal, unlawful, or harmful.

---

[39] J. Kosseff, *The Twenty-Six Words That Created the Internet*, Ithaca, Cornell University Press, 2019, p. 153; J. van Hoboken & D. Keller, 'Design Principles for Intermediary Liability Laws', *Transatlantic High Level Working Group on Content Moderation Online and Freedom of Expression*, 2019, available at ivir.nl/nl/twg/publications-transatlantic-working-group, p. 4; D. Keller, 'Empirical Evidence of "Over-Removal" by Internet Companies Under Intermediary Liability Laws', *The Center for Internet and Society,* 8 May 2020, available at cyberlaw.stanford.edu/blog/2015/10/empirical-evidence-over-removal-internet-companies-under-intermediary-liability-laws (retrieved on 15 February 2022).

[40] Article 14 of Directive 2000/31/EC (*Directive on electronic commerce*).

[41] D. Keller, 'Who Do You Sue? State and Platform Hybrid Power over Online Speech', *Aegis Paper Series* No. 1902, Hoover Working Group on National Security, Technology, and Law, 2019, available at lawfareblog.com/who-do-you-sue-state-and-platform-hybrid-power-over-online-speech.

[42] D. Keller, 'Six Constitutional Hurdles for Platform Speech Regulation', *The Center for Internet and Society,* 22 January 2021, available at cyberlaw.stanford.edu/blog/2021/01/six-constitutional-hurdles-platform-speech-regulation-0 (retrieved on 15 February 2022).

[43] Citron, 'Extremist Speech, Compelled Conformity, and Censorship Creep', *Notre Dame Law Review*, 2018, pp. 1046-1047; European Commission, 'Disinformation: EU assesses the Code of Practice and publishes platform reports on coronavirus related disinformation', *European Commission,* 10 September 2020, available at ec.europa.eu/commission/presscorner/detail/en/ip_20_1568 (retrieved on 14 February 2022).

[44] See, for example, Rb. Noord-Holland (vzr.), 6 October 2021, ECLI:NL:RBNHO:2021:8539 (*Kamerlid/LinkedIn*).

This dissertation consists of three parts that contribute to answering this question. The first part discusses the background and theory of regulation of providers that function as internet intermediary service providers. The second part discusses the liability regime applicable to providers in the US and the European context (mainly the EU). In the third part, these liability regimes are critically reviewed: to what extent do these liability regimes incentive overregulation and underregulation in the light of applicable freedom of expression rights?

Methodology and roadmap

Part 1

The first two chapters analyse how user-provided information can be regulated by regulating providers that offer a service for such information. Chapter 1 focuses on the concepts related to providers, and Chapter 2 focuses on the conceptualisation of internet content regulation and internet intermediary regulation. The first part of this dissertation, in sum, thus introduces the necessary vocabulary, technological, and theoretical background to understand how the state relates to providers.

Chapter 1

The first chapter discusses what is pivotal to fulfilling an intermediary function. Therefore, the different intermediary functions providers can fulfil are discussed. The first part of this chapter discusses what an internet intermediary service provider is, how a provider differs from traditional "offline" intermediaries (such as newspapers), and the defining chrematistics of these providers. Because the internet is (and was) treated differently by lawmakers, this chapter provides a concise introduction to how EU and US lawmakers regulate(d) providers.

The second part of this chapter discusses how regulation of providers evolved in the US and the EU based on Goldman's "three waves" of regulation.[45] The chapter continues by analysing how regulation of providers relates to their technological capabilities, legal obligations, and functional relationship with the content of user-provided information. The relationship of the different providers to the technological building blocks of the internet (for example, their relation to the physical cable infrastructure, physical servers, and software) reveals their technological capabilities when regulating the content of user-provided information. Besides, these providers rely on legal concepts set out in legislation and case law. These legal concepts codify assumptions and expectations of how providers can intervene in the content of user-provided information. This part is legal conceptual since it identifies the (legal) presumptions in legislation and case law regarding the relationship between providers and user-provided information. Next to the technological limitations and the legal framing of providers, the third dimension, the functional dimension, reveals the providers' relationship to the content of user-provided information. This functional dimension thus shows how providers are presumed to be involved in the content of user-provided information and how this relates to user expectations. For example, a user may not expect an e-mail service provider to alter the content of the e-mails sent by the user. These dimensions result in a typology of providers that clarify how they relate to internet content

[45] E. Goldman, 'The Third Wave of Internet Exceptionalism', *in* B. Szoka & A. Marcus (Eds.), *The Next Digital Decade: Essays On The Future Of The Internet*, Washington, D.C., TechFreedom, 2010.

regulation in the following chapters. In addition, this typology uncovers some normative presumptions about the most suitable providers for carrying out content-based restrictions.

Chapter 2

The second chapter builds upon the first chapter by discussing what regulatory instruments are available to make providers responsible for the content of user-provided information. This chapter provides a legal and theoretical overview of how the regulation of providers could influence the content of user-provided information. This chapter first sets out the targets of state regulation and how states allot legal responsibilities between these different actors. Pivotal to this chapter is the distribution of legal liability of the content of user-provided information between the provider and the user of the service. As shown in Chapter 2, a vast body of literature deals with the liability of providers. Imposing liability, however, has to be distinguished from regulating providers. Liability and regulation conceptually differ on a few points. The liability of providers focuses on when a provider becomes legally liable for the content of user-provided information. Regulation, thus, may see to making a provider legally liable for illegal or unlawful content. Many instruments, however, may regulate providers without making them legally liable.[46]

Regulation may mean standard-setting but also "control by rules".[47] A second distinction is between regulation and legislation. Enacting legislation is reserved to the legislator, while not exclusively the state can set standards or impose control by rules. For regulation that does not qualify as legislation, this may originate from many state actors and non-state actors. While it is possible to contest the distinction between regulation and legislation, I understand regulation as an overarching concept for all types of rule-setting.[48] Legislation, in this sense, is just one possibility to regulate providers, while liability arising from enacted legislation is one of the instruments. The second chapter of this dissertation discusses internet content regulation as rule-based control over the content of user-provided information. In other words, regulation is all rule-guided decisions regarding the content of user-provided information irrespective of who is responsible for the regulation. When the provider is made responsible for carrying out such regulation, it is called internet intermediary regulation. Internet intermediary regulation encompasses all rule-guided decisions over the providers' conduct regarding the content of user-provided information. In contrast, internet content regulation encompasses all regulations irrespective of whether the provider that functions as an intermediary is directly or indirectly involved.

After discussing what actors form suitable points of regulation in theory and practice, the instruments these actors can deploy to carry out state-induced content-based regulation are discussed in the second part of this chapter. How providers address illegal, unlawful, or harmful content of user-provided information opens the door to the third question of the second chapter: the available remedies that could follow on such rule violation. Mainly this sees to the distinction between moderation and curation. While it is possible to view moderation as a specific form of curation, this chapter distinguishes between moderation and curation. This distinction, in part, follows the distinction in content moderation and recommender systems in the proposal for the

---

[46] European Commission, 2021, 'Code of Practice on Disinformation'.

[47] Lexico, 'Meaning of regulate in English', *Lexico*, available at lexico.com/definition/regulate (retrieved on 15 February 2022).

[48] N. Kosti, D. Levi-Faur & G. Mor, 'Legislation and regulation: three analytical distinctions', *The Theory and Practice of Legislation*, Vol. 7, No. 3, 2019, doi:10.1080/20508840.2019.1736369, pp. 170-171 and 175.

new DSA by the EC.[49] Curation, however, is broader than automatic recommendations by recommender systems and may also apply to manual recommendations. For now, it is essential to note that moderation results in remedies imposed on user-provided information because of its content or the account of the user responsible for the content.[50] In contrast, curation encompasses interventions that cannot be qualified as a remedy following a rule violation. Because of its impact on the accessibility of user-provided information, this chapter distinguishes moderation and curation.

The third part of the second chapter distinguishes the different remedies that may follow on moderation efforts of providers. This chapter provides an overview of the remedies in (mainly) EU legislation. Unlike the other topics discussed in this chapter, the literature on content moderation remedies is very limited. As Goldman notes, the remedies that may follow after a rule violation is easy to surpass while studying, for example, the legitimacy of (the process of) content moderation.[51]

This second chapter's fourth and last part deals with the international dimension of regulating providers. The regulation of actors, moderation and curation, and the remedies following moderation all have an international dimension. The last part of this chapter deals with the difference between state regulation and provider-imposed regulation in the international dimension of internet content regulation. Next to this, this part deals with the question of how a jurisdiction may aim to regulate user-provided information with a focus on the EU and the US.

In short, the central question is whether state regulation provides a legal incentive to providers to overregulate or underregulate user-provided information. Of course, the height of the fine or the damages paid when the provider is liable is, as noted, a (or perhaps the decisive) factor of concern. The height of the damages arising from civil liability is often not explicitly regulated in internet intermediary liability regimes. The potential heights of the damages the court may award are, thus, unpredictable. For (relatively) new administrative law instruments deployed in the EU,[52] the fines are often tied to the global turnover of providers, making it easier to predict what these fines could mean for the providers' conduct. However, this dissertation is limited to legal liability as a legal incentive, assuming that liability forms an incentive to change the conduct to avoid liability. What the height of the fine or the (potentially) awarded damages would mean for the provider's conduct is outside this dissertation's scope.

Part 2

The third and fourth chapter discusses how the EU and the US regulate providers that offer a service for users to provide information of their own. The focus is on regulation that is specific to this intermediary function. The US and the EU are home to numerous providers. In terms of successful regulation, scholars view the EU as one of the most influential regulators of providers that offer an internet intermediary service.[53] Because the US regulatory regime is a few years older

[49] Article 2(o) and (p) of Commission Proposal COM(2020) 825 final (*Digital Services Act*), pp. 44-45.

[50] Goldman, 'Content Moderation Remedies', *Michigan Technology Law Review*, 2021, pp. 5-6 and 25-31.

[51] Goldman, 'Content Moderation Remedies', *Michigan Technology Law Review*, 2021, pp. 8-9.

[52] Article 18(3) of Regulation (EU) 2021/784 of the European Parliament and of the Council of 29 April 2021 on addressing the dissemination of terrorist content online, *OJ L 172*, *17.5.2021* (data.europa.eu/eli/reg/2021/784/oj).

[53] Known as the "Brussels Effect", see Bradford, 'The Brussels Effect', *Northwestern University Law Review*, 2012; Bradford, 2020, *The Brussels Effect*.

than the EU regime, I first discuss the US and then discuss the European context and, more specifically, the EU regime.

Chapter 3

For the US, this dissertation focuses on the 47 USCA § 230 (Section 230 of the Communications Decency Act)[54] and (to a lesser extent) 17 USCA § 512 (the Digital Millennium Copyright Act).[55] Section 230 of the Communications Decency Act of 1996 (hereafter: CDA) forms the internet intermediary regulation in the US, while the DMCA forms a major exception on Section 230 for violations of intellectual property law.[56] Regarding the case law relevant for this dissertation, I rely on (helpful) overviews of the most relevant case law for Section 230 provided by the Electronic Frontier Foundation and Jeff Kosseff.[57] Regarding the DMCA and other statutes that complement or diverge from Section 230, I only discuss the main rule without providing an extensive overview of the related case law.

Chapter 4

For the EU, this dissertation mainly focuses on Articles 14 and 15 of the e-Commerce Directive,[58] which forms the heart of the EU internet intermediary regulatory regime for providers that can regulate the content of user-provided information.[59] As far as diverging from or complementing the general rules laid down in the e-Commerce Directive, this dissertation also includes Regulations, Directives, and other policy instruments.[60] Next to these instruments, I include the relevant interpretations offered by the European Court of Justice (hereafter: ECJ) as far as it concerns these articles. While with respect to Section 230, the body of case law is extensive, requiring a selection of the most influential judgements, the ECJ interpreted Articles 14 and 15 in a few cases. National legislation that transposed this Directive or national case law is only illustrative. The regime laid down in the e-Commerce Directive is intended to be complemented and partly replaced by the DSA. The European Commission (hereafter: EC) published its proposal

---

[54] 47 USCA § 230 (West 2018, Westlaw Next through PL 116-91).

[55] §512. Limitations on liability relating to material online, 17 USCA § 512 (West 2010, Westlaw Next through PL 116-179).

[56] See Kosseff, 2019, *The Twenty-Six Words That Created the Internet*, p. 97; Par. 5.66 of F. Wilman, *The Responsibility of Online Intermediaries for Illegal User Content in the EU and the US*, Cheltenham, Edward Elgar Publishing, 2020, doi:10.4337/9781839104831, pp. 166-167.

[57] J. Kosseff, 'Resources', *Jeff Kosseff*, available at jeffkosseff.com/resources (retrieved on 15 February 2022); Electronic Frontier Foundation, 'CDA 230: Key Legal Cases', *Electronic Frontier Foundation*, available at eff.org/issues/cda230/legal (retrieved on 14 February 2022).

[58] Directive 2000/31/EC (*Directive on electronic commerce*).

[59] Par. 1.18-1.19, 2.11 and 3.01 of Wilman, 2020, *The Responsibility of Online Intermediaries for Illegal User Content in the EU and the US*, pp. 8-9, 15-16 and 56.

[60] An example of a Directive is Directive (EU) 2018/1808 of the European Parliament and of the Council of 14 November 2018 amending Directive 2010/13/EU on the coordination of certain provisions laid down by law, regulation or administrative action in Member States concerning the provision of audiovisual media services (Audiovisual Media Services Directive) in view of changing market realities, *OJ L 303*, *28.11.2018* (data.europa.eu/eli/dir/2018/1808/oj). An example of Regulation is Regulation (EU) 2021/784. An example of other policy instruments forms Communication COM(2018)236 final of the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions of 26 April 2018 Tackling Online Disinformation: A European Approach.

for the DSA in December 2020. While the DSA could (and will) be amended, this dissertation relies on the text adopted by the Commission in its December 2020 proposal.[61]

As noted, this dissertation discusses the regulation of and liability for the content of user-provided information of providers. Regarding the European Court of Human Rights (hereafter: ECtHR) cases, I only discuss article 10 ECHR cases (freedom of expression) directly relevant to providers of internet intermediary services. I consider case law directly relevant when its complements or diverges from the general rule set out in the EU e-Commerce Directive and the relevant case law.[62] I only consider ECtHR case law that directly or indirectly affects providers.

Part 3

Chapter 5

This dissertation's fifth and concluding chapter uses the concepts discussed in Part 1 to analyse the (proposals for) legislation applicable to providers discussed in Part 2. While Part 1 discusses how the providers themselves regulate content, this chapter uses these concepts to understand state regulation. For this purpose, I identify three factors derived from this dissertation's first parts. The first factor is ambiguity (both the target of regulation and its scope). The second factor is the instruments, and the third factor is the remedies. As shown in Part 1, ambiguity may cause providers to overregulate or underregulate when it is unclear how state regulation applies to them. Some instruments are more likely to cause overregulation than others. When these instruments are required by law, these instruments may cause overregulation. The same applies to remedies: when the law requires removal, such legislation may cause overregulation.

Based on the characteristics of the providers as discussed in Chapter 1 and the liability regimes set out in Chapter 2, the different liability regimes discussed in Part 2 are grouped, analysed, and discussed. Chapter 5 discusses strict liability, conditional liability (or immunity), (full) immunity regimes, and soft law regulatory regimes.[63] Discussed is to what extent these liability regimes offer an incentive for overregulation or underregulation in terms of ambiguity, chosen instruments, and what remedies these regimes require.

---

[61] Commission Proposal COM(2020) 825 final (*Digital Services Act*).

[62] *Delfi AS v. Estonia*, no. 64569/09, 10 October 2013, ECLI:CE:ECHR:2013:1010JUD006456909; *Delfi AS v. Estonia* [GC], no. 64569/09, ECHR 2015-II, 16 June 2015, ECLI:CE:ECHR:2015:0616JUD006456909; *Magyar Tartalomszolgáltatók Egyesülete and Index.hu Zrt v. Hungary*, no. 22947/13, 2 February 2016, ECLI:CE:ECHR:2016:0202JUD002294713. To a lesser extent also *Appleby and Others v. the United Kingdom*, no. 44306/98, ECHR 2003-VI, 6 May 2003, ECLI:CE:ECHR:2003:0506JUD004430698.

[63] Gillespie, 2018, *Custodians of the Internet*, p. 33.

Part 1: A Conceptualisation of Internet Intermediary Service Provider Liability

1 A (legal) gallery of internet intermediary regulation

Introduction

One of the most persistent myths is that privately-held providers are exclusively to blame for 'censoring' users while the sovereignty of the territorial state regarding internet content regulation is eroding.[64] The opposite is true: state regulation on 'the internet' has increased since the 1990s. While not 100% successful, this regulation increase can hardly be seen as an erosion of sovereignty. States can regulate providers and are increasingly doing so. For example, the US and the EU successfully imposed regulations exempting intermediaries of (some) liability for the content user-provided information.[65] However, constitutional law and human rights standards form a limitation on state regulation of providers.[66] Regulation of providers is thus not technologically impossible or legally unrealistic.

While the content categories subjected to regulation may have changed, governmental pressure on providers to regulate user-provided information is hardly new. Two of the first challenges for the territorial state were preventing minors from encountering internet pornography and combatting annoying spam.[67] Later the focus shifted to fighting illegal content such as sexual child abuse material, (illegal) gambling, and computer-related criminality.[68] Over the years, the list of categories subjected to regulation has grown. In 2018, new legislation removed the immunity for (civil) liability claims based on sex trafficking law in the US.[69] In 2021 the EU adopted new regulations targeting online terrorist content.[70] Regulation of user-provided information is thus not merely the result of providers prohibiting specific content by imposing self-regulation. To regulate internet content, lawmakers enacted real laws backed by real fines. Many of these laws aim to regulate providers directly.[71]

---

[64] In the Netherlands, this position is represented by constitutional law scholars and by governmental advisory bodies, see for example, R. Passchier, *Artificiële intelligentie en de rechtsstaat: Over verschuivende overheidsmacht, Big Tech en de noodzaak van constitutioneel onderhoud*, Den Haag, Boom juridisch, 2021, p. 81; Adviesraad Internationale Vraagstukken, 'Regulering van online content: Naar een herijking van het Nederlandse internetbeleid (AIV-advies 113)', *Adviesraad Internationale Vraagstukken,* 2020, available at adviesraadinternationalevraagstukken.nl/documenten/publicaties/2020/06/24/regulering-van-online-content (retrieved on 14 February 2022), p. 11.

[65] 47 USCA § 230 (West 2018, Westlaw Next through PL 116-91); Directive 2000/31/EC (*Directive on electronic commerce*).

[66] For example, Russia and Turkey are fairly successful in enforcing the law, see A. Kolodyazhnyy, A. Marrow & A. Osborn, 'Russia says Twitter complying with demand to remove 'banned content'', *Reuters,* 30 April 2021, available at reuters.com/technology/russia-says-twitter-is-complying-with-demand-remove-banned-content-2021-04-30 (retrieved on 15 February 2022); C. Caglayan, et al., 'YouTube says to appoint Turkey representative in line with new law', *Reuters,* 16 December 2020, available at reuters.com/article/us-turkey-socialmedia-youtube-idUSKBN28Q1T2 (retrieved on 14 February 2022). In the US, for example, the First Amendment prohibits many possible regulatory approaches, see Keller, 2021, 'Six Constitutional Hurdles for Platform Speech Regulation'.

[67] L. Lessig, *Code Version 2.0*, New York, Basic Books, 2006, p. 245.

[68] P. van Eecke, 'Online service providers and liability: A plea for a balanced approach', *Common Market Law Review*, Vol. 48, No. 5, 2011, p. 1461.

[69] Allow States and Victims to Fight Online Sex Trafficking Act of 2017 (FOSTA-SESTA), H.R. 1865, 115th Cong. (2018 through PL 115-164); Goldman, 'The Complicated Story of FOSTA and Section 230', *First Amendment Law Review*, 2019, pp. 282-284.

[70] Regulation (EU) 2021/784.

[71] Network Enforcement Act 2017 (*Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken*); Regulation (EU) 2021/784.

Regulation may not only originate from legislation enacted by the traditional territorial state. Directives and regulations proposed by the EC and adopted by the European Parliament (hereafter: EP) and the European Council (hereafter: EUCO) lay the foundation for regulation for providers.[72] Especially noteworthy is that the EC also seeks to regulate internet intermediaries without proposing legislation by, for example, concluding legally non-binding codes with providers.[73] While these codes have no legal binding force, they may be used by a judge interpreting open norms laid down in (national) legislation in a court case.[74] Violations of a voluntary code may also lead to reputational costs for the service provider.[75] Besides the possible costs of not following regional codes, these codes offer providers and regulators advantages. Such regional regulation may lead to a higher level of compliance which is attractive for the regulator, while it also may lower compliance costs for the provider.[76]

Next to regulation targeting new content categories (such as terrorist content), new regulation adds obligations on how providers should address these different content categories. For example, new regulation imposes obligations on providers in handling erroneous removal of user-provided information.[77] Providers are made legally responsible for taking down illegal or unlawful content and protecting users' freedom of expression rights. In the EU, the ambitious proposal for the DSA published in December 2020 forms an example of such regulation.[78]

Regulation of providers takes on new forms. While providers were granted exemptions for legal liability in the 1990s, internet intermediary regulation in the 2020s seeks to codify newly found legal responsibilities for these providers.[79] During the Arab spring, the view was that social media networks were an opportunity for democratic reform.[80] In 2021 this positive view changed in critique. Election disinformation and conspiracy theories leading to the violent insurrection in the US Capitol on 6 January 2021 are just two examples of such criticism.[81] Real-world events such as

---

[72] Of which is the most notable Directive 2000/31/EC (*Directive on electronic commerce*). There is a proposal to update the Directive with the Digital Services Act, see Commission Proposal COM(2020) 825 final (*Digital Services Act*).

[73] See on the status of these codes V. Mak, *Legal Pluralism in European Contract Law*, Oxford, Oxford University Press, 2020, doi:10.1093/oso/9780198854487.001.0001, pp. 131-135.

[74] For example, Rb. Amsterdam (vzr.), 9 September 2020, ECLI:NL:RBAMS:2020:4435, Rec. 4.4-4.5 and 4.11 (*YouTube*).

[75] Bradford, 2020, *The Brussels Effect*, p. 161. See for monitoring by the EC, European Commission, 2021, 'Code of Practice on Disinformation'; European Commission, 'The EU Code of conduct on countering illegal hate speech online', *European Commission*, available at ec.europa.eu/info/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/eu-code-conduct-countering-illegal-hate-speech-online_en (retrieved on 14 February 2022).

[76] Bradford, 2020, *The Brussels Effect*, pp. 162-166.

[77] Klos, "Wrongful moderation': Aansprakelijkheid van internetplatforms voor het beperken van de vrijheid van meningsuiting van gebruikers', *Nederlands Juristenblad*, 2020/2976.

[78] Commission Proposal COM(2020) 825 final (*Digital Services Act*); F. Wilman, 'Het voorstel voor de Digital Services Act: Op zoek naar nieuw evenwicht in regulering van onlinediensten met betrekking tot informatie van gebruikers', *Nederlands tijdschrift voor Europees recht*, No. 1-2, 2021, doi:10.5553/NtER/138241202021027102002, p. 34; Klos, 'De Digital Services Act: implicaties voor het recht op vrijheid van meningsuiting van gebruikers van onlineplatforms', *NTM/NJCM-bull.*, 2021/13, pp. 137-138.

[79] Commission Proposal COM(2020) 825 final (*Digital Services Act*).

[80] E. Morozov, *To Save Everything, Click Here*, London, Penguin, 2014, pp. 127-128.

[81] T. Nguyen & M. Scott, "Hashtags come to life': How online extremists fueled Wednesday's Capitol Hill insurrection', *Politico,* 8 January 2021, available at politico.com/news/2021/01/07/right-wing-extremism-capitol-hill-insurrection-456184 (retrieved on 8 January 2021).

the migration crisis (for example, in the EU),[82] election interference (in various regions),[83] live streams of terrorist attacks (addressed by, for example, the EU),[84] and the COVID-19 pandemic (worldwide)[85] fuelled or strengthened the (perceived) need for new regulation.

Regulating the providers that offer internet services is not uncomplicated. Every form of internet regulation may have unintended and unpredictable side effects. These side-effects may have massive consequences for human rights.[86] While it is easy to enact legislation to require providers to be more responsible for illegal or unlawful content of user-provided information, such legislation may not always have this effect. Besides, overregulating the internet may hamper innovation and thus the economic and societal benefits that the internet could bring.[87] The territorial state, shortly put, had (and still has) to find a mode of regulation that does remedy harmful effects that may come from the usage of the internet while safeguarding the (potential) economic and social benefits (including the possibility to exercise freedom of expression rights).

The first chapter, thus, provides an answer to the question to what extent it is necessary to distinguish regulation between (different) online and offline information intermediaries to prevent overregulation and underregulation based on the content of the information. First, I set out the differences between online and offline information intermediaries to answer this question. After setting out these differences, the discussion turns to the three waves of regulation of providers to discuss the differences in regulation between online and offline providers. After this legislative overview, this chapter discusses the differences in regulation of providers based on their technological capabilities, legal obligations, and functional involvement with the content of user-provided information. In short, this chapter thus distinguishes between

1. Who is the provider of an intermediary service?
2. What is the provided intermediary service?
3. What are the specific activities, roles, and functions that the provider of an intermediary service fulfils?

1.1 Offline information intermediaries and internet intermediary service providers

Before discussing how providers relate to internet content regulation, it is first necessary to discuss what an intermediary – or more specifically, an information intermediary – is. The online dictionary

---

[82] Recitals 53 and 57 of European Parliament resolution of 13 December 2016 on the situation of fundamental rights in the European Union in 2015 (2016/2009(INI)), *OJ C 238*, *6.7.2018*, pp. 17-18; European Commission, 'Code of Conduct on Countering Illegal Hate Speech Online', *European Commission,* 30 June 2016, available at ec.europa.eu/info/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/eu-code-conduct-countering-illegal-hate-speech-online_en (retrieved on 14 February 2022), p. 1.

[83] European Commission, 2021, 'Code of Practice on Disinformation'.

[84] L. Kayali, 'Europe's struggle against viral terrorist content', *Politico,* 21 May 2019, available at politico.eu/article/how-europe-plans-to-fight-christchurch-style-viral-content-its-complicated-fake-news-social-media-facebook-twitter-eu-terrorism (retrieved on 15 February 2022); Regulation (EU) 2021/784.

[85] World Health Organization, 2020, 'Managing the COVID-19 infodemic: Promoting healthy behaviours and mitigating the harm from misinformation and disinformation'; Joint Communication JOIN(2020) 8 final.

[86] For example, the risk that more content is removed than strictly necessary. Balkin refers to this as collateral censorship, see J. Balkin, 'Free Speech is a Triangle', *Columbia Law Review*, Vol. 118, No. 7, 2018, pp. 2016-2017. Another risk is that internet intermediaries refrain from (voluntary) moderation because of the risk of liability, see A. Kuczerawy, 'The EU Commission on voluntary monitoring: Good Samaritan 2.0 or Good Samaritan 0.5?', *CiTiP Blog,* 14 April 2018, available at law.kuleuven.be/citip/blog/the-eu-commission-on-voluntary-monitoring-good-samaritan-2-0-or-good-samaritan-0-5 (retrieved on 15 February 2022).

[87] As noted in Recital 60 of Directive 2000/31/EC (*Directive on electronic commerce*).

Lexico defines 'intermediary' as: "A person who acts as a link between people in order to try and bring about an agreement; a mediator."[88] Intermediate has its origin in contracting *inter* and *medius*, translated as *between* and *middle*.[89] An intermediary, thus, is expected to fulfil a role as a mediator between two (or more) parties. As will be shown, the internet knows its fair share of intermediaries. While these intermediaries are vastly different in size and function, they have, as Perset states, in common that they

> bring together or facilitate transactions between third parties on the Internet. They give access to, host, transmit and index content, products and services originated by third parties on the Internet or provide Internet-based services to third parties.[90]

Providers that function as intermediaries facilitate transactions in the broadest sense of the word. Providers function primarily as intermediaries that facilitate sharing and encountering (user-provided) information on the internet. However, how these providers mediate is different from how traditional offline information intermediaries mediate. While the dictionary definition of intermediary may suggest otherwise, not all providers try to conclude agreements between users.

Intermediaries mediate between authors and readers in the traditional (offline) information intermediary industry. These information intermediaries play a decisive role in deciding what information is published and what publications are not. The characterisation of intermediaries as gatekeepers comes precisely from this role.[91] Such a gatekeeping role is not without consequences. When an intermediary has control over the content of a publication, this also creates legal responsibilities. Whether a newspaper will print a piece is the editor's decision. Printing a libellous article without sufficient fact-checking may render the newspaper liable for its content.[92]

Other intermediaries are less involved with the actual content of a publication. For example, a printer does not proofread all documents for illegal content before printing. A bookshop owner or newsstand may make hard choices regarding what is put on the scarce shelf space but not read all books or newspapers before putting them out for sale. However, this more distant role does not mean that a bookseller is entirely exempted from liability for the book's content. Knowledge of the illegal or unlawful content of the book may expose the bookseller to liability.[93] Publishers and editors may be exempted from liability for the content of books to avoid preventive (self)censorship out of fear of liability. Such exceptions, however, do not mean that

---

[88] Lexico, 'Meaning of intermediary in English', *Lexico*, available at lexico.com/definition/intermediary (retrieved on 15 February 2022).

[89] Lexico, 'Meaning of intermediate in English', *Lexico*, available at lexico.com/definition/intermediate (retrieved on 15 February 2022).

[90] K. Perset, 'The Economic and Social Role of Internet Intermediaries', *OECD Digital Economy Papers* No. 117, Paris, OECD, 2010, doi:10.1787/20716826, p. 9.

[91] J. Oster, *Media Freedom as a Fundamental Right*, Cambridge, Cambridge University Press 2015, doi:10.1017/CBO9781316162736, p. 62.

[92] See, for example, *Lindon, Otchakovsky-Laurens and July v. France* [GC], no. 21279/02, 36448/02, § 65-67, ECHR 2007-IV, 22 October 2007, ECLI:CE:ECHR:2007:1022JUD002127902; *Khawar v. Globe Intern., Inc.*, 965 P.2d 696, 704-708 (Cal. S.C. 1998); *Globe Intern., Inc. v. Khawar*, 119 S.Ct. 1760 (1999).

[93] See, for example, *Smith v. People of the State of California*, 80 S.Ct. 215, 216-220 (1959); HR, 14 February 2017, ECLI:NL:HR:2017:220 (concl. P.C. Vegter), Rec. 2.1 and 3.4, *Nederlandse Jurisprudentie* 2017/259, m.nt. E.J. Dommering; R. Blommestijn & M. Klos, 'Een giftige paddenstoel voor de vrijheid van meningsuiting: Bol.com en het verbieden van 'foute' boeken', *Nederlands Juristenblad*, 2020/1209.

they have no legal responsibilities at all.[94] The general principle is simple: increasing control over and involvement in the content of a publication comes with (legal) responsibilities.

Information intermediaries are unmissable for producers and consumers of such information. Precisely this position makes information intermediaries a potent regulator over what information finds its way to consumers and what information does not. The internet is not different with respect to the reliance on information intermediaries. Users browsing the internet may feel that there are no gatekeepers – that there are no gates. A user posting on a social media platform perhaps may not realize how many intermediaries are involved. The user first has an internet connection facilitated by an internet service provider. Without this connection posting to an internet platform would be impossible. The website facilitating user-provided information is the second intermediary. The website and the provider all use intermediary services of their own. As shown, not all these intermediaries offer control points over user content. Not all intermediaries fulfil a gatekeeping role or have the (technological and legal) possibilities or responsibilities to intervene in what content can find its way on the internet.

Providers of internet information intermediary services are different from traditional information intermediaries. Posting on the internet does not require a printer to print a tweet before its content is available. There is no bookstore with limited (shelving) space for a limited selection of social media posts. Providers that offer social media functionalities do not exercise ex-ante editorial control over user-provided information often.[95] Social media platforms perhaps exist by the grace of allowing user-provided information.[96] That users can post everything they like does not mean that the providers of these platforms do not perform a governing or gatekeeping role.[97] Providers can and do intervene in the content of user-provided information. Sometimes such interventions are even referred to as a form of (private) 'censorship'.[98] As Lessig notes, censorship is a hefty term to use in the context of legitimate speech regulation. Lessig refers to legitimate 'censorship' as "speech regulation".[99] Due to the negative connotation of censorship, in this dissertation, censorship is merely used in the context of clearly illegitimate governmental interventions on freedom of expression rights.

---

[94] In the Netherlands, for example, the publisher and printer cannot be prosecuted for "crimes committed by the printing press" as long as they state the name and place of residence of the person who ordered the printing on the print or reveal the person when the examining magistrate request to do this, see Article 53 and 54 of Wetboek van Strafrecht (Dutch Criminal Code).

[95] In the Netherlands, some online news papers that offer such functionalities pre-screended user comments at one time in their history, such as nu.nl.

[96] According to the OECD which relies on the terminology user-created or generated content this concerns public accessible, non-professional content which has some creative effort, see OECD, 'Participative Web and User-Created Content', *OECD,* 2007, available at oecd-ilibrary.org/science-and-technology/participative-web-and-user-created-content_9789264037472-en, doi:10.1787/9789264037472-en, pp. 17-18.

[97] F.T. Wu, 'Collateral Censorship and the Limits of Intermediary Immunity', *Notre Dame Law Review*, Vol. 87, No. 1, 2011, pp. 298-300; E.B. Laidlaw, *Regulating Speech in Cyberspace: Gatekeepers, Human Rights and Corporate Responsibility*, Cambridge, Cambridge University Press, 2015, doi:10.1017/CBO9781107278721, pp. 52-56; Par. 6.21 of Wilman, 2020, *The Responsibility of Online Intermediaries for Illegal User Content in the EU and the US*, p. 177. Klonick, 'The New Governors: The People, Rules, and Processes Governing Online Speech', *Harvard Law Review*, 2018.

[98] M.K. Land, 'Against Privatized Censorship: Proposals for Responsible Delegation', *Virginia Journal of International Law*, Vol. 60, No. 2, 2020, p. 46.

[99] Lessig, 2006, *Code Version 2.0*, p. 254.

A second relationship in which providers mediate is between the service user (probably including you) and governments that seek to regulate the user.[100] States seek to regulate the content of user-provided information not by imposing fines or punishments on the creator of the information with illegal content but by regulating the providers that offer the service. Balkin contrasts this "new-school speech regulation" with "old-school speech regulation", in which governments directly regulate the responsible party as the creator of the information with illegal content.[101] According to Balkin, "new-school speech regulation" is characterised by states "attempting to coerce or co-opt private owners of digital infrastructure to regulate the speech of private actors."[102] Because the state depends on the providers to carry out state regulations, the provider plays a mediating role between the state and the user.

How providers shape their roles and what users expect from them is different from traditional intermediaries. Nobody frowns when a newspaper edits a reader-submitted piece for an opinion page (as long as the line of thought is maintained). Newspapers even select between different contributions offered to them. Similar interventions on, for example, Facebook are unthinkable. Newspapers edit; providers of social media platforms do not. This difference, however, does not mean that internet intermediaries do not intervene in user content at all. Providers do fulfil roles that come close to classic editorial functions: moderation and curation. While moderation usually sees to remedying extremes of the content of user-provided information or other user behaviour,[103] curation encompasses selecting and organising user-provided information based on its content.[104] The distinction between moderation and curation is not always easy to make and is certainly not recognised by every scholar.[105] As Paragraph 2.2 shows, curation may sometimes even take the form of moderation. For this paragraph, it is necessary to consider that moderation involves remedies imposed after establishing a rule violation. In contrast, curation encompasses recommendations made to (groups of) users based on the information's content and the users' characteristics.

All providers that offer a platform for user-provided information have some moderation in place. According to Gillespie, moderation "is essential, constitutional, definitional" for platform services.[106] Providers offering platform services typically allow users to submit information and offer social tools to encounter and interact with such user-provided information.[107] Because providers offer a service to upload user-provided information, they can intervene in the content of this information. This intervention may see to editing the content of the information but also

---

[100] R. MacKinnon, *Consent of the Networked: The Worldwide Struggle For Internet Freedom*, New York, Basic Books, 2013 [2012], p. 9.

[101] Balkin, 'Free Speech is a Triangle', *Columbia Law Review*, 2018, pp. 2015-2016.

[102] Balkin, 'Free Speech is a Triangle', *Columbia Law Review*, 2018, p. 2016.

[103] Lexico, 'Meaning of moderation in English', *Lexico*, available at lexico.com/definition/moderation (retrieved on 15 February 2022).

[104] Lexico, 'Meaning of curate in English', *Lexico*, available at lexico.com/definition/curate (retrieved on 15 February 2022).

[105] The distinction between 'hard' moderation and 'soft' curation is not easy to make. For example, Daphne Keller views removal and ranking as a subset of curation activities, see Keller, 2019, 'Who Do You Sue? State and Platform Hybrid Power over Online Speech'.

[106] Gillespie, 2018, *Custodians of the Internet*, p. 21.

[107] See for Gillespie's definition of platform, Gillespie, 2018, *Custodians of the Internet*, p. 18 and 21. The definition provided here and by Gillespie overlaps with the definition of 'online platform' as proposed in the Digital Services Act, see Article 2(h) of Commission Proposal COM(2020) 825 final (*Digital Services Act*), p. 45.

lead to complete removal or limiting its accessibility. Moderation sees to activities relating to "detection, review, and enforcement"[108] of the platform's guidelines which is hard or impossible when there is no direct access to and control over user-provided information. Providers that offer a service consisting of platform functionalities without moderation are a rarity. According to Gillespie, moderation is "the commodity" offered by these services. Moderation makes a platform usable. Without moderation, illegal or undesirable content would swamp the platform.[109] Therefore, even platforms that promise almost unrestricted freedom of expression have some moderation.[110] While a broad range of remedies is available, moderation efforts typically lead to keeping information up (no violation) or removing information (after a violation) based on its content.[111] The rationality behind removal and the possible alternative remedies is part of the discussion in Chapter 2 of this dissertation.

Moderation is the first, most defining activity of these internet intermediary service providers. The second activity undertaken by these providers is curation. Curation comes close to classic editorial functions. All platform services include some moderation. In contrast, not all services offer curation.[112] As set out, curation encompasses selecting and organising user-provided information. Curation is different from moderation since curation usually is not used to remedy the platform's policy violations. Curation encompasses decisions about what, how and when information with specific content is shown to users. Many providers use so-called 'recommender systems'[113] – automatic systems to recommend user-provided information to other users.[114] While filtering and recommending relevant content is a classic intermediary function, the difference is that providers are hardly active or conscious. Instead, predictions form the basis for recommendations of what user-provided information may be relevant for those who receive these recommendations.[115] Providers, however, often lack in-depth knowledge of why automatic systems make a specific recommendation because the process is complex. For example, Twitter noted in some countries that its algorithm amplified posts by right-wing politicians more than user-provided information posted by left-wing politicians. The reason for this difference in amplification? Twitter could not tell for sure.[116] Recommender systems (automatic systems that recommend user-provided information to other users), for example, take into account previous interactions with other information and what other users clicked on that have similar profiles to the user that receives the recommendations. There are signs that these recommender systems, for example, recommend user-provided information with extremist content after viewing content that

---

[108] Gillespie, 2018, *Custodians of the Internet*, p. 21.

[109] Gillespie, 2018, *Custodians of the Internet*, p. 207.

[110] Parler, 'Community Guidelines', *Parler,* 2 November 2021, available at parler.com/documents/guidelines.pdf (retrieved on 15 February 2022).

[111] Goldman, 'Content Moderation Remedies', *Michigan Technology Law Review*, 2021, pp. 4-6.

[112] Of course, this position could be contested as well. The argument can be made that platforms that do not select or organise user content but simply provide a chronological timeline which is also a form of curation.

[113] Article 2(o) of Commission Proposal COM(2020) 825 final (*Digital Services Act*), p. 45.

[114] Gillespie, 2018, *Custodians of the Internet*, p. 196.

[115] Legally, how internet intermediaries are involved in user content, may matter for their liability, see Judgement of the Court (Grand Chamber) of 12 July 2011 in *C-324/09, L'Oréal SA and Others v. eBay International AG and Others*, ECLI:EU:C:2011:474, in particular Rec. 116.

[116] R. Chowdhury & L. Belli, 'Examining algorithmic amplification of political content on Twitter', *Twitter Blog,* 21 October 2021, available at blog.twitter.com/en_us/topics/company/2021/rml-politicalcontent (retrieved on 14 February 2022); Chowdhury & Belli, 2021, 'Examining algorithmic amplification of political content on Twitter'.

only relates lightly to such extremist content.[117] Automatic content curation is thus not a stamp of approval of the provider vowing for the authenticity, originality, or factuality of the content of information in question. The providers may have other goals in recommending user-provided information to other users.[118] However, amplifying information with specific content is not (always) grounded in a conscious decision.

In its gatekeeping role, a provider could exercise broad discretion. Providers are merely required to moderate illegal or unlawful content. Providers, however, could moderate additional categories of content on top of illegal content.[119] Providers that curate even have a broader discretion in promoting and demoting user-provided information. In doing so, some providers assert that they are concerned with upholding their users' freedom of expression rights.[120] However, there is little transparency about how providers regulate user-provided information.[121] Next to a lack of transparency, only a few (in the EU) to almost none (in the US) legal remedies exist for users to oppose moderation efforts the user deems unfair.[122]

A distinction between ex-ante and ex-post regulation is helpful in this respect. Ex-ante regulation encompasses interventions before admittance; ex-post regulation to inventions on user-provided information already admitted to the service. While ex-ante regulation of user-provided information comparable to traditional media is still technologically possible, many providers refrain from such ex-ante control because it is hard to scale. Instead, they choose ex-post moderation of user-provided information. Providers choose such moderation since the liability regime for user-provided information differs from traditional media. Newspapers exercising editorial control also bear legal responsibility for what they publish. For providers, this is not necessarily the case.[123] Traditional media usually are liable for publishing illegal content. In contrast, some additional conditions must be satisfied for internet providers (at least) before the provider can be held liable. There may be good reasons for such exceptionalism.[124] For example, state regulation imposing liability on providers for the content of user-provided information may lead

---

[117] Research suggests that this is the case with YouTube, see J. Whittaker, et al., 'Recommender systems and the amplification of extremist content', *Internet Policy Review*, Vol. 10, No. 2, 2021, doi:10.14763/2021.2.1565 pp. 12-13 and 15-16.

[118] Balkin, 'Free Speech is a Triangle', *Columbia Law Review*, 2018, pp. 2047-2048.

[119] See, for example, Article 14 of Directive 2000/31/EC (*Directive on electronic commerce*). A similar argument is made by Keller, 2019, 'Who Do You Sue? State and Platform Hybrid Power over Online Speech', p. 26.

[120] See, for example, Meta, 'Mark Zuckerberg Stands for Voice and Free Expression', *Meta Newsroom,* 17 October 2019, available at about.fb.com/news/2019/10/mark-zuckerberg-stands-for-voice-and-free-expression (retrieved on 14 February 2022).

[121] W. Benedek & M.C. Kettemann, *Freedom of Expression and the Internet*, Strasbourg, Council of Europe Publishing, 2020, pp. 87-89; Klonick, 'The New Governors: The People, Rules, and Processes Governing Online Speech', *Harvard Law Review*, 2018, pp. 1665-1666.

[122] D.K. Citron, 'Fix Section 230 and hold tech companies to account', *Wired,* 6 May 2021, available at wired.co.uk/article/section-230-social-media (retrieved on 14 February 2022); Keller, 2019, 'Who Do You Sue? State and Platform Hybrid Power over Online Speech', p. 2; Council of Europe, 2021, 'Content moderation: best practices towards effective legal and procedural frameworks for self-regulatory and co-regulatory mechanisms of content moderation', pp. 31-33.

[123] Most noteworthy, 47 USCA § 230(c) (West 2018, Westlaw Next through PL 116-91); Article 14 of Directive 2000/31/EC (*Directive on electronic commerce*).

[124] The term 'exceptionalism' is derived from Goldman, 2010, 'The Third Wave of Internet Exceptionalism'.

to a 'chilling effect'[125] on users' freedom of expression rights.[126] Because providers have a legal incentive to engage in excessive moderation of illegal content, this may lead to a chilling effect encompassing the removal of grey-area content that is not illegal or unlawful.

Regulation that imposes liability on providers is not the only reason providers regulate user-provided information. For example, political pressure to act against online terrorist content may be a vestibule to legislation backed by fines, causing providers to self-regulate in advance.[127] However, also non-state pressure may influence the policies of providers. Users voting with their feet or calling for a boycott may have such effects.[128] Besides, providers may copy the policies that apply to other services, leading to the formation of what douek calls "content cartels".[129] A distinction between regulation from other types of influence is necessary. Regulation means using rules that aim to alter the provider's conduct. Mere influence does not encompass such rules. However, for this dissertation, government actors signalling that they will enact regulation will also be counted towards regulation because it aims to control the conduct of providers by enacting rules.[130]

While pressure on providers is not always successful,[131] it is undeniable that providers operate in a highly regulated landscape. Such regulation directly or indirectly targets user-provided information and thus affects the user who provided the information to the service. Because the provider is the primary target of and responsible for carrying out regulation, how such regulation views the user's role is secondary. Therefore, it is necessary to discuss the landscape in which these providers function. Why did providers become the primary target for regulation? Why do governments not invest in better regulating users instead? How does such regulation relate to the technological features of these providers? What are the legal categories used to understand these

---

[125] *Lexico* defines 'chilling effect' as

> A discouraging or deterring effect on the behaviour of an individual or group, especially the inhibition of the exercise of a constitutional right, such as freedom of speech, through fear of legal action.

See, Lexico, 'Meaning of chilling effect in English', *Lexico*, available at lexico.com/definition/chilling_effect (retrieved on 15 February 2022).

[126] Wu, 'Collateral Censorship and the Limits of Intermediary Immunity', *Notre Dame Law Review*, 2011, pp. 304-308.

[127] Citron, 'Extremist Speech, Compelled Conformity, and Censorship Creep', *Notre Dame Law Review*, 2018, pp. 1037-1038.

[128] However, the cost of leaving or changing platforms may be too high, see Gillespie, 2018, *Custodians of the Internet*, p. 177. This is caused by network effects: the value of a network (or platform) is tied to its amount of users, see M. Yemini, 'The New Irony of Free Speech', *Columbia Science and Technology Law Review*, Vol. 201, No. 1, 2018, pp. 181-182. Of course, this brings intermediaries in a position of enormous power, see F. Pasquale, 'Platform Neutrality: Enhancing Freedom of Expression in Spheres of Private Power', *Theoretical Inquiries in Law*, Vol. 17, No. 2, 2016, doi:10.1515/til-2016-0018, p. 496.

[129] e. douek, 'The Rise of Content Cartels', *Knight Columbia,* 11 February 2020, available at knightcolumbia.org/content/the-rise-of-content-cartels (retrieved on 14 February 2022), pp. 18-19.

[130] C. Angelopoulos, et al., 'Study of fundamental rights limitations for online enforcement through self-regulation Institute', *Institute for Information Law (IViR),* 2015, available at hdl.handle.net/11245.1/7317bf21-e50c-4fea-b882-3d819e0da93a, pp. 56-57.

[131] D. Rushe & Associated Press, 'Mark Zuckerberg: advertisers' boycott of Facebook will end 'soon enough'', *The Guardian,* 2 July 2020, available at theguardian.com/technology/2020/jul/02/mark-zuckerberg-advertisers-boycott-facebook-back-soon-enough (retrieved on 15 February 2022).

providers? How does this relate to the roles and functions these providers fulfil? These questions are central to this first chapter.

1.2 Drafting the laws of the internet

In the 1990s, legislators had to deal with a thorny question: to what extent should providers be liable for the content of user-provided information? Enacted legislation offers an 'exceptionalist'[132] position to these providers, which treats providers differently than their offline counterparts. This exceptionalism arose because the providers of internet intermediary services are viewed differently from their offline counterparts, which justifies a different legal treatment. New proposals for legislation are (necessary) built upon these pre-existing statutes: there is always some path dependency. Earlier choices with respect to the legal regimes influence new regulations.[133] EU proposal for the new DSA builds upon the e-Commerce Directive enacted in 2000.[134] In the US, new proposals for legislation must somehow relate to the 1996 enacted Section 230 of the CDA.[135] The discussion of this legislation takes place in Part 2. For now, it is necessary to remark that lawmakers seek to increase the regulatory burden for providers, making them responsible for distinct content categories of user-provided information and (perceived) harms that come from the existence of these services.

While lawmakers increasingly add new obligations for providers in legislation, this does not mean that a critical stance from the state was absent before. According to Goldman, internet regulation came in three waves of "exceptionalism", which means that the internet is in all these waves treated differently from other, mainly traditional – offline – media.[136] In the first wave, regulation favoured providers over offline media; in the second wave, regulation became stricter for providers of internet intermediary services. The third wave came with a more nuanced view of internet intermediary regulation by differentiating regulation between providers based on their characteristics.[137]

1.2.1 The first wave: exceptionalist statutes that form the foundation

On 24 May 1995, the New York Supreme Court ruled in a defamation case against such a provider in *Stratton Oakmont v. Prodigy*. Prodigy exploited an online bulletin board – an early predecessor of social media platforms. Such a bulletin board allowed users to publish comments. In the 'Money Talk'-section, comments were posted on the subject of the brokerage house Stratton Oakmont. Some of these comments were defamatory.[138] Stratton Oakmont sued Prodigy for damages and asked for a court order to remove the defamatory comments.[139] Whether Prodigy qualified as a distributor or publisher of the defamatory comments was pivotal for the liability of Prodigy. As Kosseff notes, the distributor/publisher distinction was decisive at the time. As a distributor,

---

[132] Goldman, 2010, 'The Third Wave of Internet Exceptionalism'.

[133] For a brief description of this phenomenon, see L.B. Solum, 'Legal Theory Lexicon: Path Dependency', *Legal Theory Blog,* 2 September 2018, available at lsolum.typepad.com/legaltheory/2018/09/legal-theory-lexicon-path-dependency.html (retrieved on 15 February 2022).

[134] Commission Proposal COM(2020) 825 final (*Digital Services Act*), pp. 1-3.

[135] FOSTA-SESTA, H.R. 1865, 115th Cong. (2018 through PL 115-164).

[136] Goldman, 2010, 'The Third Wave of Internet Exceptionalism', p. 165.

[137] Goldman, 2010, 'The Third Wave of Internet Exceptionalism', pp. 165-167.

[138] *Stratton Oakmont, Inc. v. Prodigy Services Co.*, 23 Media L Rep 1794 (N.Y. Sup. Ct. 1995). Kosseff, 2019, *The Twenty-Six Words That Created the Internet*, pp. 45-48.

[139] Kosseff, 2019, *The Twenty-Six Words That Created the Internet*, p. 48.

Prodigy would only be liable when scienter could be proven. As a publisher, Prodigy would be liable no matter what.[140]

How was Prodigy involved in the user-provided information? Prodigy, the court established, 1) set the rules on the bulletin board, 2) reserved the right to delete posts in violation of these rules, and 3) manually reviewed the bulletin board for violating the rules but abandoned this practice due to the large volumes of comments. Instead, Prodigy replaced manual review with automatic scanning software. This software only scanned for words that were on a so-called blocklist. However, this automatic filtering software could not evaluate whether the comment's meaning was defamatory. For such comments, Prodigy relied on a notice and takedown procedure. Prodigy would remove defamatory comments after it received a notification.[141] Based on how Prodigy handled user-provided information, the Court concluded that Prodigy exercised editorial control over the comments and that Prodigy thus could be considered a publisher of these comments.[142]

The *Prodigy* ruling led to a discussion in the US House of Representatives. Two members of the House, Christopher Cox and Ron Wyden, expressed their concern that the *Stratton Oakmont* ruling would lead providers to refrain from moderation out of fear of liability. Another risk was that it would cause providers to shut down their services. Therefore, Cox and Wyden proposed an amendment to the Communications Decency Act, which was adopted and codified into law as Section 230.[143] Section 230 aimed to exempt providers from liability as a "publisher or speaker" for the content of user-provided information[144] and wished to encourage voluntary moderation by shielding providers from liability for voluntary "good faith" moderation.[145] These protections will be discussed more extensively in Chapter 3. As Kosseff notes, Section 230 offers much broader protection to providers than traditional media.[146] For example, under Section 230, whether a provider has knowledge (scienter) of defamatory content of user-provided information is irrelevant.[147] The reasons, the text of the statute, and its effects on the internet make Section 230 truly an exceptionalist statute.[148] Paragraph 3.1 discusses the different exceptions and nuances of Section 230 protections. For now, it is only necessary to keep in mind that Section 230 offered an

---

[140] This is the standard laid down in *Cubby, Inc. v. CompuServe, Inc.*, 776 F.Supp. 135, 139-142 (S.D. New York 1991); Kosseff, 2019, *The Twenty-Six Words That Created the Internet*. Under Section 230, the publisher/distributor distinction, however, has lost its meaning since distributing could be seen as a subclass of publishing for the purpose of Section 230, see *Zeran v. America Online, Inc.*, 129 F.3d 327, 331-334 (3rd Cir. 1997).

[141] Kosseff, 2019, *The Twenty-Six Words That Created the Internet*, pp. 49-51.

[142] Kosseff, 2019, *The Twenty-Six Words That Created the Internet*, p. 52. *Stratton Oakmont, Inc. v. Prodigy Services Co.*, 23 Media L Rep 1794 (N.Y. Sup. Ct. 1995).

[143] Kosseff, 2019, *The Twenty-Six Words That Created the Internet*, pp. 60-66; D. Citron & B. Wittes, 'The Internet Will Not Break: Denying Bad Samaritans § 230 Immunity', *Fordham Law Review*, Vol. 86, No. 2, 2017, pp. 405-406; 47 USCA § 230 (West 2018, Westlaw Next through PL 116-91).

[144] 47 USCA § 230(c)(1) (West 2018, Westlaw Next through PL 116-91).

[145] 47 USCA § 230(c)(2) (West 2018, Westlaw Next through PL 116-91).

[146] Kosseff, 2019, *The Twenty-Six Words That Created the Internet*, p. 65.

[147] E. Goldman, 'Why Section 230 Is Better Than the First Amendment', *Notre Dame Law Review*, Vol. 95, No. 1, 2019 (available at scholarship.law.nd.edu/ndlr_online/vol95/iss1/3), p. 38.

[148] E. Goldman, 'An Overview of the United States' Section 230 Internet Immunity', *in* G. Frosio (Ed.) *The Oxford Handbook of Online Intermediary Liability*, Oxford, Oxford University Press, 2020, doi:10.1093/oxfordhb/9780198837138.013.8, pp. 162-165.

exemption for liability from user-provided information to providers that do not equally apply to traditional intermediaries.

The history of how the e-Commerce Directive of 2000 made it into the EU lawbooks is different and less exciting. The Directive, as its name suggests, deals with electronic commerce. Only a portion of the Directive discusses the legal exception of liability for providers.[149] Section 230 would (mainly) get its EU equivalent in Article 14, which limits the liability for hosting service providers,[150] and Article 15, which prohibits member states from imposing a general obligation on internet intermediaries to monitor for user-provided information with illegal or unlawful content.[151] As discussed in Paragraph 4.1, the protections offered by the e-Commerce Directive are not absolute and certainly not unconditional. While there could be a freedom of expression incentive behind these provisions, the original proposal dating from 1998 suggests that economic motives were of primary concern to the drafters of the Directive. Elimination of internal market barriers is a requirement to fully profit from the innovation provided by providers of internet services. One of these barriers was that internet intermediaries had to adhere to the different liability regimes enacted by the member states of the EU.[152] The meaning of the Directive for the internal market still has a prominent place in the Directive.[153] However, unlike the 1998 proposal, the 2000 Directive also refers to users' freedom of expression rights in the recitals.[154] Recitals, setting out the purpose and goals of the Directive, may gain legal meaning in court proceedings in interpreting the provisions laid down in the Directive.[155] For example, as Recital 46 notes, interventions on user-provided information required under the Directive "has to be undertaken in the observance of the principle of freedom of expression".[156] The e-Commerce Directive offered some exemptions from liability for the content user-provided information. Exemptions that are, again, not offered to traditional information intermediaries – even when they blindly copy-paste the content of the information.

Goldman describes the first wave of exceptionalist regulation as 'Internet Utopianism'. Goldman's primary focus as a US legal scholar is on US Section 230. By granting such an exemption from liability, providers of internet intermediary services were (largely) left unregulated. Meaning that providers were not actively required to combat user-provided information containing illegal or unlawful content.[157] These statutes may be different from those that apply to traditional

---

149 'Liability of intermediary service providers' is the title of the fourth section of Directive 2000/31/EC.

150 Hosting services are not liable for user content as long they 1) have not knowledge or are not aware of the illegal content 2) act expeditiously when they gain such knowledge or awareness and remove or disable access to this content, see Article 14 of Directive 2000/31/EC (*Directive on electronic commerce*). But, see also Article 12 and 13 of Directive 2000/31/EC (*Directive on electronic commerce*).

151 Article 15 of Directive 2000/31/EC (*Directive on electronic commerce*).

152 Recital 2, 4-5 and more specific 16 of Proposal COM(1998) 586 final of 23 December 1998 for a European Parliament and Council Directive on certain legal aspects of electronic commerce in the internal market, *OJ C 30, 5.2.1999*.

153 Recital 2-3 and 5-7 of Directive 2000/31/EC (*Directive on electronic commerce*).

154 Recital 9 and 46 of Directive 2000/31/EC (*Directive on electronic commerce*).

155 Van Eecke, 'Online service providers and liability: A plea for a balanced approach', *Common Market Law Review*, 2011, pp. 1467-1468.

156 Recital 46 of Directive 2000/31/EC (*Directive on electronic commerce*).

157 Goldman, 2010, 'The Third Wave of Internet Exceptionalism', p. 165.

intermediaries, to which often more strict liability regimes apply.[158] US Section 230 and the safe harbours in the Directive are the products of a time of internet optimism. This optimism can be characterised as utopian regarding what the internet would bring, combined with the (incorrect) claim that the internet was unregulatable.[159] One of the voices of this utopian thinking was Barlow. In 'A Declaration of the Independence of Cyberspace' Barlow argued that 'Cyberspace' is and should be independent of the territorial state and its government:

> We are forming our own Social Contract. This governance will arise according to the conditions of our world, not yours. Our world is different.[160]

Barlow's view is a form of norm duality: what is prohibited offline is not necessarily prohibited on the internet. However, this utopian thinking never found its way to the US or EU lawbooks. As Lawrence Lessig notices in *Code Version 2.0* published in 2006, the internet and state regulations differed from when the first edition came out in 1999. Lessig:

> the dominant idea among those who raved about cyberspace then was that cyberspace was beyond the reach of real-space regulation. Governments couldn't touch life online. And hence, life online would be different, and separate, from the dynamic of life offline.[161]

How governments understood the internet in 1999 changed radically in 2006. In 2006 it was clear that the territorial state was interested in regulating the internet and could do so. Internet exceptionalism and the cyberlibertarian ideal are thus not one-on-one related.[162] Neither the US nor the EU in the 1990s accepted the unregulatability of the internet. Instead, the US and the EU seemed warry that a lack of exceptionalism would hamper innovation and harm freedom of expression. Not because the internet was beyond the reach of the law, but because of the fear that existing legislation could hinder innovation.[163]

While governments increasingly exert sovereignty over the internet,[164] the exceptionalist legislation enacted in the 1990s is still in place. [165] The exceptionalist laws became a monumental part of the internet intermediary regulation landscape. Scholars, providers, and legislators became fond of this legislation. The following paragraphs show that it is difficult to change such fundamental legislation.

---

[158] For example, in the Netherlands a provider is better protected than a book seller for criminal prosecution for group defamation, see Blommestijn & Klos, 'Een giftige paddenstoel voor de vrijheid van meningsuiting: Bol.com en het verbieden van 'foute' boeken', *Nederlands Juristenblad*, 2020/1209.

[159] Lessig, 2006, *Code Version 2.0*, p. 3.

[160] J. Barlow, 'A Declaration of the Independence of Cyberspace', *Electronic Frontier Foundation,* 8 February 1996, available at eff.org/nl/cyberspace-independence (retrieved on 14 February 2022).

[161] Lessig, 2006, *Code Version 2.0*, p. ix.

[162] H.B. Holland, 'In Defense of Online Intermediary Immunity: Facilitating Communities of Modified Exceptionalism', *University of Kansas Law Review*, Vol. 56, No. 2, 2008, doi:10.17161/1808.19996, p. 376.

[163] Recital 2-3 and 5-7 of Directive 2000/31/EC (*Directive on electronic commerce*); 47 USCA § 230(b) (West 2018, Westlaw Next through PL 116-91).

[164] For example, the EC keeps emphasising that conduct that is illegal offline is also illegal online which is an expression that EU norms also apply to the internet, see European Commission, 'Europe fit for the Digital Age: Commission proposes new rules for digital platforms', *European Commission,* 15 December 2020, available at ec.europa.eu/commission/presscorner/detail/en/ip_20_2347 (retrieved on 14 February 2022).

[165] For example, 47 USCA § 230 (West 2018, Westlaw Next through PL 116-91); Article 12-15 of Directive 2000/31/EC (*Directive on electronic commerce*).

1.2.2 The second wave: internet paranoia

While received in the early 1990s with optimism and utopian thinking, a more critical perspective emerged in the late 1990s. The second wave – which Goldman dubbed "Internet Paranoia" – meant that online activities were more strictly regulated than similar offline activities.[166] While the assumption was that internet regulation of the state was easy to circumvent, governments could regulate the internet by regulating the providers offering services on the internet.[167] The second wave of exceptionalism showed that undesirable conduct on the internet was only possible to regulate by regulating the providers.[168] Providers (often against their will) may enable users to engage in illegal or unlawful conduct by offering their services. This conduct is challenging for the state to address without the provider's help.[169] In other words, the state requires a point of control to regulate internet content successfully. The territorial state relies on providers such as Google, Microsoft, Facebook, and Amazon. These providers offer the means to carry out regulation and thus form targets for regulation themselves.[170]

Internet paranoia is never wholly abandoned. While it has a negative connotation, internet paranoia does not mean (necessarily) that there is an overreaction from governmental actors. Internet paranoia merely means that the treatment of internet providers diverges from the treatment of offline media. For example, during the COVID-19 pandemic, governments treated providers differently from traditional offline media with an unprecedented sense of urgency. While COVID-19 mis- and disinformation could also be spread by offline media, the governmental focus was mainly on providers of internet intermediary services. COVID-19 mis- and disinformation thus mark a new 'peak' of internet paranoia. The World Health Assembly declared that an "infodemic" was taking place, "particularly in the digital sphere, as well as the proliferation of malicious cyber-activities that undermine the public health response".[171] The EC repeated this in words and policy.[172] Despite this language, it is necessary to remark that misinformation and disinformation are not necessarily illegal. User-provided information that qualifies as disinformation or misinformation under providers' policies may even fall under the protection of freedom of expression rights.[173] Chapters 3 and 4 discuss that freedom of expression rights

---

166 Goldman, 2010, 'The Third Wave of Internet Exceptionalism', pp. 165-166.

167 J. Balkin, 'Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation', *U.C. Davis Law Review*, Vol. 51, No. 3, 2018, p. 1187.

168 Goldman does not connect 'Internet Paranoia' directly to internet intermediary regulation but only to a different treatment of similar conduct on the internet to offline conduct (for example, online hunting was criminalised while offline hunting was not). However, the different examples offered by Goldman see to regulating this conduct trough regulating internet intermediaries, see Goldman, 2010, 'The Third Wave of Internet Exceptionalism', pp. 165-166.

169 For example, Yahoo! was held responsible for allowing users from the US to sell Nazi paraphilia to users in France. After Yahoo! changed its policy, the case was dismissed as no longer relevant in the United States, see *Yahoo! Inc. v. La Ligue Contre Le Racisme et l'antisémitisme (LICRA)*, 433 F.3d 1199 (9th Cir. 2006).

170 Balkin, 'Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation', *U.C. Davis Law Review*, 2018, p. 1175.

171 The Seventy-third World Health Assembly, 'Resolution WHA73.1: COVID-19 response', *World Health Organization,* 19 May 2020, available at apps.who.int/gb/ebwha/pdf_files/WHA73/A73_R1-en.pdf (retrieved on 15 February 2022); World Health Organization, 2020, 'Managing the COVID-19 infodemic: Promoting healthy behaviours and mitigating the harm from misinformation and disinformation'.

172 Joint Communication JOIN(2020) 8 final, pp. 1 and 8-10.

173 J. van Hoboken, et al., 'Het juridisch kader voor de verspreiding van desinformatie via internetdiensten en de regulering van politieke advertenties', Amsterdam, IVIR, 2019, available at ivir.nl/publicaties/download/Rapport_desinformatie_december2019.pdf, pp. 18-19.

primarily work between the state and its citizens, meaning that state interference in their citizens' freedom of expression rights is restricted. That said, this does not mean that disinformation is not harmful. The state might have a legitimate interest in regulating COVID-19 mis- and disinformation.

The obligation for the state to respect citizens' freedom of expression rights seems to lose importance when regulating user-provided information through providers. US President Biden said in an interview that online platforms did not do enough against COVID-19 misinformation, which led to the qualification that they are "killing people".[174] While Biden retracted this statement a few days later,[175] the signal was clear: providers must step up their game in combating COVID-19 misinformation. Such non-legislative pressure to regulate speech is known as 'jawboning' – a practice that, according to Genevieve Lakier, may raise First Amendment issues in the US.[176] Not only is informal pressure put on providers to regulate COVID-19 misinformation, but there is also a legislative proposal that seeks to exempt service providers from Section 230 immunity for COVID-19 misinformation.[177]

To be clear: internet paranoia does not mean that (potential) legislative responses form an exaggeration. COVID-19 disinformation is harmful and may pose a real threat – how significant a threat will become apparent in the future. "Paranoia", instead, refers to the difference in treatment compared to traditional media. When it comes to harmful (not necessarily illegal) disinformation, providers face a stricter approach than traditional media.[178] There are no proposals to make television networks liable for the spread of COVID-19 misinformation. There are no legislative proposals to introduce new governmental oversight over television networks spreading mis- or disinformation. In this respect, providers of internet intermediary services are treated differently from traditional media. However, this new internet paranoia does not lead to abolishing exemptions for liability of the content of user-provided information. Instead, these exceptionalist laws – at their core – seem to survive new proposals – they even might be reinforced.[179] However,

---

[174] D. Judd, M. Vazquez & D. O'Sullivan, 'Biden says platforms like Facebook are 'killing people' with Covid misinformation', *CNN,* 17 July 2021, available at edition.cnn.com/2021/07/16/politics/biden-facebook-covid-19/index.html (retrieved on 15 February 2022).

[175] B. Klein, M. Vazquez & K. Collins, 'Biden backs away from his claim that Facebook is 'killing people' by allowing Covid misinformation', *CNN,* 20 July 2021, available at edition.cnn.com/2021/07/19/politics/joe-biden-facebook/index.html (retrieved on 15 February 2022).

[176] G. Lakier, 'The Trump Lawsuits, the Biden Administration's Misinformation Advisory and the Thorny First Amendment Problem of Jawboning', *Lawfare,* 26 July 2021, available at lawfareblog.com/trump-lawsuits-biden-administrations-misinformation-advisory-and-thorny-first-amendment-problem (retrieved on 15 February 2022).

[177] A bill to amend the Communications Act of 1934 to provide that, under certain circumstances, an interactive computer service provider that allows for the proliferation of health misinformation through that service shall be treated as the publisher or speaker of that misinformation, and for other purposes (Health Misinformation Act of 2021), S. 2448, 117th Cong. (2021).

[178] For example, in the Netherlands the Rathenau Institute devoted an entire report to online moral excesses that, according to the report, should be responded to with exceptionalist measures aimed only at internet intermediaries. Evidently, moral transgressions on the Internet are so serious that they deserve their own approach, without going as far as criminalisation that also extends to offline media, see M. van Huijstee, et al., 'Online ontspoord: Een verkenning van schadelijk en immoreel gedrag op het internet in Nederland', *Rathenau Instituut,* 7 July 2021, available at rathenau.nl/nl/digitaal-samenleven/online-ontspoord (retrieved on 15 February 2022).

[179] For example, the Digital Services Act-proposal contains an explicit exemption from liability arising from voluntary actions in 'detecting, identifying and removing, or disabling of access to, illegal content', see Article 6 of Commission Proposal COM(2020) 825 final (*Digital Services Act*), p. 47.

informal governmental pressure and formal legislation that makes internet intermediaries liable for user content may affect these laws contrary to their initial meaning.[180] While providers are not legally obligated to screen user-provided information for unlawful, illegal and (certainly not) harmful content, they may feel pressured to do so. However, these effects, at least until now, do not pose a real threat to the exceptionalist statutes.

1.2.3 Exceptional exceptionalism: a gallery of statutes

Both 'Internet Utopianism' and 'Internet Paranoia' express an exceptional state regulation stance towards the internet. The internet is treated differently from traditional, offline media. In the case of utopianism, the regulation of providers is more favourable in comparison to offline intermediaries. With internet paranoia, this is the other way around. Providers, in this case, are regulated stricter than offline intermediaries that deal with similar types of content or conduct. As shown, regulation of providers still knows its fair share of utopianism and paranoia. Utopianism and paranoia, however, are complemented with a more nuanced view of regulation. According to Goldman, this 'Exceptionalism Proliferation', which forms the third wave, can be characterised as the differentiation of regulation between types of providers.[181] Goldman points out, as an example, that social media networks are regulated differently from other websites.[182]

While Goldman noticed the 'Exceptionalism Proliferation' in 2010,[183] this may still be the default in regulating providers in 2021. At least, this is the case in the EU. While the EC emphasises that "[w]hat is illegal offline is also illegal online",[184] an exceptionalist approach is chosen in decisions over policy instruments to address user-provided information with illegal or unlawful content.[185] A video-sharing platform service has different obligations than other audiovisual media services[186] According to the proposal for the DSA, very large platforms should be regulated differently than more small-scale services.[187] Because providers differ, the standards that apply to different providers also differ. For example, providers are regulated based on the size and functionalities they offer. With the proposal for the DSA, the EC made this differentiation between both size and functionalities more than explicit.[188]

---

[180] For example, FOSTA-SESTA, H.R. 1865, 115th Cong. (2018 through PL 115-164). See Goldman, 'The Complicated Story of FOSTA and Section 230', *First Amendment Law Review*, 2019, pp. 288-289.

[181] Goldman, 2010, 'The Third Wave of Internet Exceptionalism', pp. 166-167.

[182] Goldman, 2010, 'The Third Wave of Internet Exceptionalism', p. 167.

[183] Goldman, 2010, 'The Third Wave of Internet Exceptionalism', p. 167.

[184] Communication COM(2017)555 final of the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions of 28 September 2017 Tackling Illegal Content Online Towards an enhanced responsibility of online platforms, p. 2.

[185] European Commission, 2016, 'Code of Conduct on Countering Illegal Hate Speech Online'; Communication COM(2018)236 final; Regulation (EU) 2021/784.

[186] See, Chapter IXA of Directive 2010/13/EU of the European Parliament and of the Council of 10 March 2010 on the coordination of certain provisions laid down by law, regulation or administrative action in Member States concerning the provision of audiovisual media services (*Audiovisual Media Services Directive*), *OJ L 95*, *15.4.2010* (data.europa.eu/eli/dir/2010/13/oj); As amended by Article 1(23) of Directive (EU) 2018/1808.

[187] Article 25 of Commission Proposal COM(2020) 825 final (*Digital Services Act*), p. 59.

[188] For example, 'very large online platforms' are regulated differently than online platforms that do not qualify as 'very large', see Article 25 of Commission Proposal COM(2020) 825 final (*Digital Services Act*). In the case of the Regulation on addressing the dissemination of terrorist content online, the size of the hosting service provider does not matter for the requirements under the Regulation. However, Member States seeking to impose a penalty, are required to take into account the size of the provider, see Article 18(2)(f) of Regulation (EU) 2021/784.

In the US, content-based regulation of providers leads to admissibility concerns under the First Amendment.[189] The impossibility of such legislation does not mean that providers can do as they wish. For example, providers could be held morally responsible for user-provided information, which may be a powerful incentive for providers to change their conduct to prevent other legislation that is not content-based.[190] While content-based restrictions are the subject of First Amendment scrutiny, this does not mean that legislation with such restrictions does not make it to the books. There are exceptionalist statutes proposed and adopted at the state level. For example, a Florida Senate bill made it into law in 2021.[191] Since 1 July 2021,[192] social media platforms, for example, "may not willfully deplatform a candidate for office who is known by the social media platform to be a candidate".[193] Providers of social media platforms that fail to comply with this legislation expose themselves to a fine of "$250,000 per day for a candidate for statewide office and $25,000 per day for a candidate for other offices."[194]

However, the Florida bill does not apply to social media platforms with less than 100 million global users every month or less than $100 million in annual revenue. In addition, this legislation also does not apply to "a company that owns and operates a theme park or entertainment complex".[195] Distinguishing between providers with and without a theme park in Florida is 'exceptionalism' (but this time for providers that also own an offline theme park) in its strangest form. Whether the legislation is enforced is questionable. On 30 June 2021, the United States District Court of the Northern District of Florida granted a preliminary injunction against the Florida social media bill, which is

> subject to strict scrutiny because it discriminates on its face among otherwise-identical speakers: between social-media providers that do or do not meet the legislation's size requirements and

---

189 Keller, 2021, 'Six Constitutional Hurdles for Platform Speech Regulation'.

190 B. Sander, 'Democratic Disruption in the Age of Social Media: Between Marketized and Structural Conceptions of Human Rights Law', *European Journal of International Law*, 2021, doi:10.1093/ejil/chab022, p. 17; Land, 'Against Privatized Censorship: Proposals for Responsible Delegation', *Virginia Journal of International Law*, 2020, pp. 387-388. As douek notes, such informal pressure is not always put on intermediaries publicly, see douek, 2020, 'The Rise of Content Cartels', p. 20. Such pressure, according to Keller occurs in the US and the EU. In the US governmental pressure on internet intermediaries to regulate content may violate the First Amendment, see Keller, 2019, 'Who Do You Sue? State and Platform Hybrid Power over Online Speech', pp. 6-7. As Balkin summarises such '[j]awboning sends the message that infrastructure providers should be patriotic and cooperate with the government, rather than getting on the bad side of government officials.' see Balkin, 'Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation', *U.C. Davis Law Review*, 2018, p. 1179.

191 2021 Fla. Sess. Law Serv. Ch. 2021-32 (SB 7072) (West).

192 At least, that was the plan which is put on hold now a preliminary injunction against the law is granted, see *NetChoice, LLC v. Moody*, 2021 WL 2690876 (N.D. Florida 2021); S. Morrison, 'Florida's social media free speech law has been blocked for likely violating free speech laws', *Vox Recode,* 1 July 2021, available at vox.com/recode/2021/7/1/22558980/florida-social-media-law-injunction-desantis (retrieved on 15 February 2022). However, appeal has been filed, see M. Masnick, 'Florida Man Governor Wastes More Florida Taxpayer Money Appealing Ruling About His Unconstitutional Social Media Law', *techdirt,* 13 July 2021, available at techdirt.com/articles/20210713/09513247161/florida-man-governor-wastes-more-florida-taxpayer-money-appealing-ruling-about-his-unconstitutional-social-media-law.shtml (retrieved on 15 February 2022).

193 106.072. Social media deplatforming of political candidates, Fla. Stat. Ann § 106.072(2) (West 2021, Westlaw Next).

194 Fla. Stat. Ann § 106.072(3) (West 2021, Westlaw Next).

195 501.2041. Unlawful acts and practices by social media platforms, Fla. Stat. Ann § 501.2041(1)(g) (West 2021, Westlaw Next).

> are or are not under common ownership with a theme park. The legislation does not survive strict scrutiny. Parts also are expressly pre-empted by federal law.[196]

An appeal is filed against the decision of the District Court. Commentators, however, do not hold their breath. Masnick, for example, criticises this appeal as "yet more of a waste of Florida taxpayer money on a frivolous legal battle".[197]

This brief overview showed how various providers are regulated exceptionally. Exceptionally compared to offline intermediaries and between the different functionalities and sizes providers of internet intermediary services. Through this exceptional regulation, providers are, for example, made responsible for policing hate speech,[198] online terrorist content,[199] and disinformation.[200] For sex trafficking content,[201] copyright violations,[202] and possible much more when new (pending) legislation is adopted.[203] However, regulating providers may lead to issues that are (again) exceptional to internet intermediaries. While the exceptionalist statutes aimed to prevent regulation from hurting innovation, economic progress, and the exercise of freedom of expression rights, new regulation may have unintended side effects. Setting out the landscape in which providers function helps to understand how regulation of providers works.[204] As the following paragraphs show, it is hard to get a clear overview of the regulatory landscape for providers. These regulations all relate in a certain way to how providers, technologically, legally, and functionally fulfil their roles as providers. The three dimensions are the topic of discussion in the next paragraph.

1.3 Carving internet intermediary regulation: three dimensions

As discussed in the previous paragraph, internet intermediary regulation differentiates between different intermediary functions. Regulation may, for example, consider the service provider's size, the monetary success of the service provider, or the actual roles and functions the provider fulfils in the internet intermediary landscape. As discussed, this exceptionalism is tied to the service provider's capabilities and, thus, how the provider relates to user-provided information. The provider relates to this user-provided information in three ways.

First, a service has a technological relationship to user-provided information. Providers differ in the technological capabilities to regulate the content of user-provided information. Due to the internet's design, some providers (for example, providers of social media platforms) are better equipped than others (for example, a telecom provider offering internet access services) to

---

196 *NetChoice, LLC v. Moody*, 2021 WL 2690876 (N.D. Florida 2021).

197 Masnick, 2021, 'Florida Man Governor Wastes More Florida Taxpayer Money Appealing Ruling About His Unconstitutional Social Media Law'.

198 European Commission, 2016, 'Code of Conduct on Countering Illegal Hate Speech Online'.

199 Regulation (EU) 2021/784.

200 European Commission, 2021, 'Code of Practice on Disinformation'.

201 FOSTA-SESTA, H.R. 1865, 115th Cong. (2018 through PL 115-164).

202 Article 17(3) of Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC, *OJ L 130*, *17.5.2019* (data.europa.eu/eli/dir/2019/790/oj).

203 For example, Article 14 of Commission Proposal COM(2020) 825 final (*Digital Services Act*). Article 14(1) contains the obligation to 'put mechanisms in place to allow any individual or entity to notify them of the presence on their service of specific items of information that the individual or entity considers to be illegal content.'

204 G. Dinwoodie, 'Who are Internet Intermediaries?', *in* G. Frosio (Ed.) *Oxford Handbook of Online Intermediary Liability*, Oxford, Oxford University Press, 2020, doi:10.1093/oxfordhb/9780198837138.013.2.

regulate the content of user-provided information. Second, providers are, as already shown, grasped in legal categories laid down in case law and statutes. Providers thus differ in their legal relationship with user-provided information. Some providers, for example, may become liable for the content of user-provided information, while others have an exception from such legal liability. The technological dimension is (to a large extent) a given. The legal dimension, however, is not. The technological dimension, to some extent, guides the legal categories applicable to providers. However, the legal dimension also consists of normative claims and policy goals. The technical dimension thus limits legal concepts. The legal concepts, however, seek to influence the functional dimension. The functional dimension (how providers choose to offer their services) may also influence the legal categories because legislators may consider new types of services when drafting new legislation (internet marketplaces and app stores).

The functional dimension is not a one-way street. Providers offer capabilities to users. Users are enabled to use the service in pre-defined ways. For example, an e-mail service may allow text input but forbid (large) attachments. A chat service may prohibit users from taking screenshots of chats or images as a privacy feature. A microblogging service may limit the number of characters or words in one post. These limitations are technological – it is hard to circumvent them. The functional dimension also has a legal (or better: moral) dimension. Users of these services trust the provider to refrain from conduct contrary to the user's goal in using this service. The provider trusts that the user does not misuse the service. The functional dimension is built upon the technological and legal dimension while it adds (unspoken) assumptions about what the provider and the user can expect from each other.

1.3.1 Internet intermediaries: the technological dimension

This chapter primarily deals with providers that offer internet intermediary services with functionalities related to storing, indexing, ranking, and recommending user-provided information. So-called hosting service providers are in the best position to intervene in the content of user-provided information because they have direct access to the information stored by them. So-called social media platforms are often targeted by state regulation because of this hosting role they fulfil. In contrast, an internet service provider (ISP) that provides internet access is normally exempted from such regulation because they lack such a hosting role.[205] A broader range of providers and related functions are subject to discussion to understand the differences in technological capabilities between providers and the different services they offer. Therefore, first, I discuss ISPs to contrast these providers with hosting service providers.

Internet service providers: a gateway to the internet

ISPs function as a gateway to the internet. Without ISPs, it is impossible to access any other service on the internet. ISPs may offer an internet connection through a landline (via the telephone line, coax cable or fibre optic cable) or a wireless connection (most commonly 4G or 5G internet access). Because ISPs depend on a physical infrastructure within the state's territory, they could be targeted directly by state regulation. The territorial state is not dependent on compliance or the

[205] An EU example is the proposal for the DSA. Section 2 (hosting) and Section 3 (very large online platforms) of Chapter 3 of the DSA do not concern ISPs, see Commission Proposal COM(2020) 825 final (*Digital Services Act*), p. 51 and 75. In the US, the DMCA distinguishes between different intermediary functions, compare 17 USCA § 512(a) and (c) (West 2010, Westlaw Next through PL 116-179).

help of a company abroad. The territorial state could even cut the cord in the most extreme case. When the service offered by the ISP is down, there is no access to the internet.[206]

According to Felix Wu, internet intermediary functions – broadly taken – are not technologically different from their pre-internet counterparts. For example, an ISP and traditional telephone company offer a similar service: access to a network.[207] In the case of a telephone provider, this consists of offering a connection to other telephones, while an ISP provides access to a network of other computers.[208] The only difference is the potential use of an internet connection. While a telephone service provider offers a service that allows the user to make a telephone call to another user, the potential usage of an internet connection is practically unlimited. A telephone provider gives access to others by offering voice communication; an ISP offers access to an incredible number of different services.

When it comes to ISP imposing regulations themselves, one of the concerns is that they may restrict access to other services for commercial reasons.[209] While both telephone service providers and ISPs could have commercial goals in restricting the usage of their services or physical infrastructure, for an ISP, such restrictions may be easier to monetise.[210] Technically, ISPs could, for example, restrict access to so-called *tube* services such as YouTube or prohibit videoconferencing services unless users subscribe to a more expensive service plan. As will be shown, modern ISPs have the technical capabilities to discriminate between services by filtering, restricting, and even blocking access to services. An ISP requiring a premium plan before video services can be accessed may seem far-fetched to US and European users, mainly because of so-called network neutrality regulations, which force ISPs to treat traffic equally.[211] However, in a slimmed-down and adapted form, ISPs sometimes favour some (types of) services over others. Zero-rating, for example, exempts the usage of a (group of) service(s) from counting towards monthly data limits. ISPs may favour one or more services by calculating data usage for music streaming applications, while video streaming applications would not count to the monthly restrictions.[212] Another form of favouring services over others is so-called 'paid prioritisation'. Paid prioritisation means that other providers could pay the ISP for faster connections for their users to their service. For example, an ISP may limit the bandwidth used for a specific service to a bitrate that equals HD quality. A streaming service could pay an ISP to remove these limitations to allow

---

[206] Klos, 2021, 'Westphalian Sovereignty and the 4th Industrial Revolution: In Search of Legitimate Governmental Control over Online Content', pp. 110-111.

[207] Wu, 'Collateral Censorship and the Limits of Intermediary Immunity', *Notre Dame Law Review*, 2011, p. 313.

[208] Britannica, 'Internet service provider', *Encyclopedia Britannica,* 13 March 2018, available at britannica.com/technology/Internet-service-provider (retrieved on 14 February 2022).

[209] T. Wu, *The Curse of Bigness: Antitrust in the New Gilded Age*, New York, Columbia Global Reports, 2018, pp. 94-97.

[210] Some issues are remarkably similar. Telephone companies, for example, prohibited users to attach own, not approved, equipment which is very similar to some restrictions ISPs impose, see T. Wu, 'Network Neutrality, Broadband Discrimination', *Journal on Telecommunications & High Technology Law*, Vol. 2, 2003, pp. 157 and 159-162; T. Wu, *The Master Switch: The Rise and Fall of Information Empires*, New York, Vintage Books, 2011, pp. 101-114.

[211] Wu, 'Network Neutrality, Broadband Discrimination', *Journal on Telecommunications & High Technology Law*, 2003, pp. 141-142 and 145-146.

[212] Body of European Regulators for Electronic Communications, 'What is zero-rating?', *Body of European Regulators for Electronic Communications*, available at berec.europa.eu/eng/netneutrality/zero_rating/ (retrieved on 14 February 2022).

users to stream in 4K quality.[213] Such a difference in speed (and thus quality) may nudge users to the faster services, while services that cannot pay the ISP fees may see users leave the service.

While network neutrality rules restrict paid prioritisation and zero-rating, this does not withhold ISPs from seeking the limits of this legislation. The same is true for other means of monetisation. In October 2021, ISPs reopened the debate about whether they could impose limitations on third-party services because of the success of the Netflix series "Squid Game". Due to its popularity, ISPs saw an increase in traffic which they want to pass on the costs to Netflix.[214]

ISPs, as a gateway to the internet, are bound by network neutrality rules. These rules require services to treat all traffic equally. However, not only ISPs themselves may impose usage limitations on their network users. Sometimes states seek to regulate providers (and sometimes especially ISPs) to prohibit specific content categories. The limitations discussed in this paragraph were about service-based restrictions. How would ISPs – and other providers – carry out content-based restrictions on information?

Distinguishing between content and service-based restrictions

ISP-imposed restrictions are (mostly) aimed at complete service. Content-based restrictions by an ISP based on the actual content of the information generally do a poor job. Because of how ISPs function, service-level restrictions are the primary way ISPs can influence the spread of content categories. ISPs, for example, could block access to a service by adding the domain name to a blocklist. A user typing in an internet address (a domain name) in their browser usually results in the ISP translating the domain name to a numerical IP address which allows the browser to find the location of the service on the internet. An ISP could prevent this and thus block access to the service. Such blockades, however, are circumventable by directly entering the IP address. Next to domain name blocking, ISPs can also completely block access to an IP address, which offers a more severe restriction.[215] A third option is that an ISP limits the usage of specific protocols. Closing certain "ports" disables the usage of services that use these ports. By such a limitation, an ISP can, for example, restrict access to video conferencing software.[216]

Service-level restrictions have some severe downsides. In the first place, allowing an ISP to discriminate between services raises competition questions. According to Tim Wu, network discrimination may hamper competitive innovation by raising financial barriers for new competitors. A new service may not have a chance when competitors have an advantage because of zero-rating or paid prioritisation. Network neutrality, requiring ISPs to treat all traffic equally, in opposition, stimulates competition since there are no barriers for new providers to reach potential users.[217] In the end, an internet without net neutrality thus may lead to fewer services and

---

213 K. Trendacosta, 'Busting Two Myths About Paid Prioritization', *Electronic Frontier Foundation,* 16 April 2018, available at eff.org/deeplinks/2018/04/busting-two-myths-about-paid-prioritization (retrieved on 15 February 2022).

214 M. Sweney, 'Squid Game's success reopens who pays debate over rising internet traffic', *The Guardian,* 10 October 2021, available at theguardian.com/business/2021/oct/10/squid-games-success-reopens-debate-over-who-should-pay-for-rising-internet-traffic-netflix (retrieved on 15 February 2022).

215 As, for example, was proposed in the content of The Piratebay, see HR, 13 November 2015, ECLI:NL:HR:2015:3307, *Nederlandse Jurisprudentie* 2018/110, m.nt. P.B. Hugenholtz.

216 Wu, 'Network Neutrality, Broadband Discrimination', *Journal on Telecommunications & High Technology Law*, 2003, p. 165.

217 Wu, 'Network Neutrality, Broadband Discrimination', *Journal on Telecommunications & High Technology Law*, 2003, pp. 141-142 and 145-146.

less diversity between these services, which may indirectly harm the freedom of expression rights of (potential) users.

There are some direct concerns for freedom of expression rights as well. When an ISP limits access to services, this would also affect the information made available through this service. When the ISP intends to restrict access to a specific instance of information by disabling access to the whole service, this is the textbook example of an overbroad restriction. Not all information on the service may contain illegal or otherwise prohibited content. The ISP thus also restrict legal and lawful information by restricting access to the complete service.[218] For example, the Committee of Ministers of the Council of Europe relates network neutrality directly to access to the internet as a right protected under freedom of expression rights. Allowing ISPs to restrict access to services may also (indirectly) limit access to information.[219]

At first, ISPs were a popular target for state regulation.[220] ISPs have a visible presence within jurisdictions in the form of a physical infrastructure that offers a point of contact for state regulation.[221] While content regulation through ISPs may be highly effective, the downside for freedom of expression rights of such bucket shot regulation is acknowledged.[222] Let alone some exceptions,[223] the EU and US have regulations that exempt providers of internet intermediary

---

[218] In the context of governmental regulation, Balkin warns for collateral censorship, see Balkin, 'Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation', *U.C. Davis Law Review*, 2018, pp. 1176-1177. However, there are little reasons to believe that ISPs conducting in such regulation themselves would not amount to similar effects. As Jonathan Zittrain puts it: 'ISPs can serve as Internet police, not only cordoning off areas from view when acting as hosts of content, but also more broadly restricting access to particular networked entities with whom their customers wish to communicate-thus determining what those customers can see, wherever it might be online.', J. Zittrain, 'Internet Points of Control', *Boston College Law Review*, Vol. 44, No. 2, 2003, p. 655.

[219] Paragraph 3 and 4 of Committee of Ministers, 'Declaration of the Committee of Ministers on network neutrality (Adopted by the Committee of Ministers on 29 September 2010 at the 1094th meeting of the Ministers' Deputies)', *Council of Europe,* 29 September 2010, available at rm.coe.int/09000016805ce58f (retrieved on 14 February 2022). See also Committee of Ministers, 'Recommendation CM/Rec(2016)1 of the Committee of Ministers to member States on protecting and promoting the right to freedom of expression and the right to private life with regard to network neutrality (Adopted by the Committee of Ministers on 13 January 2016, at the 1244th meeting of the Ministers' Deputies)', *Council of Europe,* 13 January 2016, available at rm.coe.int/09000016805c1e59 (retrieved on 14 February 2022).

[220] At least in Germany, France and Great Britain, see J. Goldsmith & T. Wu, *Who Controls the Internet: Illusions of a Borderless World*, New York, Oxford University Press, 2008, p. 73.

[221] Goldsmith & Wu, 2008, *Who Controls the Internet*, pp. 73-74; Zittrain, 'Internet Points of Control', *Boston College Law Review*, 2003, pp. 672-673.

[222] For example by the ECtHR, see *Ahmet Yıldırım v. Turkey*, no. 3111/10, § 66, ECHR 2012-VI, 18 December 2012, ECLI:CE:ECHR:2012:1218JUD000311110; *Cengiz and Others v. Turkey*, no. 48226/10 and 14027/11, § 64, ECHR 2015-VIII, 1 December 2015, ECLI:CE:ECHR:2015:1201JUD004822610; *Kablis v. Russia*, no. 48310/16 and 59663/17, § 94, 30 April 2019, ECLI:CE:ECHR:2019:0430JUD004831016; *Engels v. Russia*, no. 61919/16, § 33, 23 June 2020, ECLI:CE:ECHR:2020:0623JUD006191916; *Vladimir Kharitonov v. Russia*, no. 10795/14, § 38, 23 June 2020, ECLI:CE:ECHR:2020:0623JUD001079514; *OOO Flavus and Others v. Russia*, no. 12468/15, 23489/15 and 19074/16, § 36-39, 23 June 2020, ECLI:CE:ECHR:2020:0623JUD001246815.

[223] ISPs may be required by court order to end or prevent specific infringements, see Article 12(3) of Directive 2000/31/EC (*Directive on electronic commerce*). In Canada a new proposal for legislation targeting online harms, ISPs are explicitly targeted, see Government of Canada, 'Consultation closed: The Government's proposed approach to address harmful content online - Discussion guide', *Government of Canada,* 29 July 2021, available at canada.ca/en/canadian-heritage/campaigns/harmful-online-content/discussion-guide.html (retrieved on 15 February 2022); Government of Canada, 'Consultation closed: The Government's proposed approach to address

services from legal obligations to regulate user-provided information containing illegal or unlawful content directly or indirectly.[224] ISPs in the EU and the US (at least on the state level) are subjected to network neutrality regulations to prevent ISPs from imposing restrictions on what services may use their network.[225] Such net neutrality regulations changes who controls the network. Typically, de provider decides. Network neutrality regulation shifts this to the users. The user of a service and the service provider decide what is requested and transmitted – not the ISP.[226]

ISPs and other internet intermediary services must be distinguished for network neutrality regulation. Network neutrality regulation does not apply to all types of internet intermediary services. Only providers that offer internet access services (ISPs) must adhere to network neutrality regulations.[227] For example, social media platforms do not qualify as access providers and thus are not legally required to uphold network neutrality. The rationality behind this distinction is technological: where ISPs restricting the usage of an internet connection can have severe consequences in terms of user access to internet services, gatekeeping by social media platforms does not have a similar effect. Restrictions imposed by these providers do not extend to the whole

---

harmful content online - Discussion guide - Technical paper', *Government of Canada,* 29 July 2021, available at canada.ca/en/canadian-heritage/campaigns/harmful-online-content/technical-paper.html (retrieved on 15 February 2022). See also, M. Geist, 'Picking Up Where Bill C-10 Left Off: The Canadian Government's Non-Consultation on Online Harms Legislation', *Michael Geist,* 30 July 2021, available at michaelgeist.ca/2021/07/onlineharmsnonconsult (retrieved on 14 February 2022).

[224] For the EU, see Article 12 and 15 of Directive 2000/31/EC (*Directive on electronic commerce*). For the US, see 47 USCA § 230(c)(1) (West 2018, Westlaw Next through PL 116-91); 17 USCA § 512(a) (West 2010, Westlaw Next through PL 116-179).

[225] For the EU-context see, Article 3 of Regulation (EU) 2015/2120 of the European Parliament and of the Council of 25 November 2015 laying down measures concerning open internet access and amending Directive 2002/22/EC on universal service and users' rights relating to electronic communications networks and services and Regulation (EU) No 531/2012 on roaming on public mobile communications networks within the Union, *OJ L 310*, *26.10.2015* (data.europa.eu/eli/reg/2015/2120/oj). In the US, network neutrality was put under pressure by the Trump Administration, see J. Kastrenakes, 'Trump's new FCC chief is Ajit Pai, and he wants to destroy net neutrality', *The Verge,* 23 July 2017, available at theverge.com/2017/1/23/14338522/fcc-chairman-ajit-pai-donald-trump-appointment (retrieved on 15 February 2022). In December 2017, the FCC adopted a policy that largely departed from the core principles of net neutrality J. Kastrenakes, 'The FCC just killed net neutrality', *The Verge,* 14 December 2017, available at theverge.com/2017/12/14/16776154/fcc-net-neutrality-vote-results-rules-repealed (retrieved on 15 February 2022). However, President Biden signed an executive order in July 2021 in which the FCC is asked to restore net neutrality provisions, see R. Lawler & A. Robertson, 'Biden signs executive order targeting right to repair, ISPs, net neutrality, and more', *The Verge,* 9 July 2021, available at theverge.com/2021/7/9/22569869/biden-executive-order-right-to-repair-isps-net-neutrality (retrieved on 15 February 2022).

[226] A. Bridy, 'Remediating Social Media: A Layer-Conscious Approach', *Boston University Journal of Science and Technology Law*, Vol. 24, No. 2, 2018, p. 201.

[227] In the EU, it concerns services "that provides access to the internet, and thereby connectivity to virtually all end points of the internet", see Article 2(2) and 3 of Regulation (EU) 2015/2120. In the US, this is more complicated now an ISP does not necessarily fit in existing definitions. Therefore, the FCC argued that it has the power to classify ISPs under 'telecommunications services' which allows the FCC to impose network neutrality regulation, see K.A. Ruane, 'Net Neutrality: Selected Legal Issues Raised by the FCC's 2015 Open Internet Order', *in* D. Lambert (Ed.) *Net Neutrality and the FCC: Legal Issues and Matters of Debate*, New York, Nova Science Publishers, 2015, pp. 4-10. In 2017 the FCC reclassified ISPs which prohibits the FCC for imposing network neutrality regulation, see Kastrenakes, 2017, 'The FCC just killed net neutrality'. The pendulum, however, may swing back now the FCC is asked to reclassify ISPs as 'telecommunication services', see Lawler & Robertson, 2021, 'Biden signs executive order targeting right to repair, ISPs, net neutrality, and more'. Due to its technical nature and little meaning for content regulation these statutes and regulations will not be discussed at large.

of the internet. Users may still access the service, and the content of the information offered – only not through this specific platform. The user can directly type in this link in the browser and visit the website. ISPs and social media platforms are thus different in this respect. Imposing network neutrality to providers offering social media platform services would be unnecessary. Equally, it would be silly to make an ISP liable as a distributor of illegal content due to their lack of technological control over the actual content of the information transmitted.

The different providers offer functionalities on these different layers, giving them various degrees of control over the content of user-provided information. How providers relate to user-provided information can be best understood by dividing the internet into so-called layers. As Bridy notes, state regulation encourages or discourages providers functioning on these layers from enacting content-based restrictions.[228]

The OSI model and the layers of the internet

Providers both rely on and offer technological functions to the internet. The Open Systems Interconnection (OSI) model helps understand how these providers function by dividing the internet infrastructure into layers. The OSI model distinguishes between seven layers that function independently from each other. Independent means that the different layers do not have access to what happens in the other layers. However, the higher layers do require the existence of the lower layers. Without cables making up the physical layer (Layer 1), no data link layer enables data transmission (Layer 2).[229] Layering offers standardisation which enables all kinds of devices to communicate. The standardisation in layers allows users to choose what devices they connect to the network. Next, this standardisation allows the development of all kinds of applications that use the network.[230]

The seven layers (the physical layer, the data link layer, the network layer, the transport layer, the session layer, the presentation layer, and the application layer)[231] do not require an elaborate discussion. A simplified OSI model of three layers suffices to explain the technological capabilities of the different providers regarding user-provided information. Riordan groups the seven layers into the physical, network, and application layers. The physical layer includes all hardware (severs and cables). The network layer includes all computer code that enables computers to communicate and transmit information to other computers. The application layer includes all code related to offering an application service.[232] The user can interact with the service and see the content of the information. In other words, the content of information becomes visible on the application layer.

In the OSI model, there is not one layer that specifically enables regulating the content of user-provided information. Therefore, layering itself does not provide the possibility to impose

---

[228] Bridy, 'Remediating Social Media: A Layer-Conscious Approach', *Boston University Journal of Science and Technology Law*, 2018, p. 205.

[229] Wikipedia, 'OSI model', *Wikipedia,* 15 February 2022, available at en.wikipedia.org/wiki/OSI_model (retrieved on 15 February 2022). When it comes to explaining technology, there is no better resource available than Wikipedia.

[230] J.B. Speta, 'A Common Carrier Approach to Internet Interconnection', *Federal Communications Law Journal*, Vol. 54, No. 2, 2002, pp. 246-247.

[231] Wikipedia, 2022, 'OSI model'.

[232] J. Riordan, *The Liability of Internet Intermediaries*, Oxford, Oxford University Press, 2016, pp. 33-34 and 36-37.

such regulation.[233] While it would be possible to add an independent layer that directly enables content regulation, its effects would be negligible. Since all layers function independently, there is no technological obligation to use such a layer. Technically requiring this layer would cause pre-existing devices and services to lose or break functionalities while limiting newly developed ones to the content regulation layer's capabilities.[234]

With or without a content regulation layer, there will be differences between providers in terms of capabilities in carrying out content-based regulation. Most providers that (can) regulate the content of user-provided information are application layer services. Bridy refers to this layer as the "human-experiential layer" because the application layer is what users see.[235] This concentration of regulation on the application layer is provided by how the internet infrastructure functions. Because providers functioning on the lower layers do not have access to the content of the higher layers, it is technically not feasible to regulate the actual content of the information transmitted on the network layer.[236] For this reason, Balkin argues that the highest layer of the OSI model, the so-called application layer, is the most suitable layer to carry out regulation of user-provided information. Besides, these providers often provide so-called edge services.[237]

These edge services operate (as the name suggests) at the edges of the internet. As understood in this contribution, these edge services are closest to the user. The user understands that the edge service provider offers the service to the user.[238] The edge services are also best-known to different users. Facebook, for example, is a social media network (or platform) functioning at the edge of the internet. Other examples of edge services fitting this definition are Google, Netflix, and Amazon. As noted, defining these edge services is that these services have the most direct contact with the user who uses the service. Consequently, in the users' view, the responsibility for regulating user-provided information lies with the edge service providers.[239] When a user's post is removed on Facebook, nobody suspects that any other provider is responsible for this intervention other than Facebook. Not the hosting service, the payment service, or the ISPs are looked at when user-provided information is regulated, but the platform to which the user provided its information.

---

[233] An so-called 'Identity Layer' that offers the possibility to users to verify their identity could offer some control, see Lessig, 2006, *Code Version 2.0*, pp. 50-52. Riordan adds on top of the OSI model layers an eight layer which contains the actual content in a human readable format: the content layer, see Par. 2.30 of Riordan, 2016, *The Liability of Internet Intermediaries*, p. 34.

[234] Lessig, 2006, *Code Version 2.0*, p. 145. Of course, different internet intermediary services may depend on such a layer making it impossible to use this service without such a layer. When I discuss a 'content regulation layer', this may also mean an 'identity layer', since regulation depends on 'who did what where', see Lessig, 2006, *Code Version 2.0*, p. 54. Of course, it would be possible to regulate ISPs to prohibit connections that do not use layers that allow for content regulation.

[235] Bridy, 'Remediating Social Media: A Layer-Conscious Approach', *Boston University Journal of Science and Technology Law*, 2018, p. 205.

[236] Note 119 of Balkin, 'Free Speech is a Triangle', *Columbia Law Review*, 2018, p. 2037.

[237] Note 119 of Balkin, 'Free Speech is a Triangle', *Columbia Law Review*, 2018, p. 2037.

[238] This definition does not follow the normal technological definition of edge service.

[239] Note 119 of Balkin, 'Free Speech is a Triangle', *Columbia Law Review*, 2018, p. 2037; Ruane, 2015, 'Net Neutrality: Selected Legal Issues Raised by the FCC's 2015 Open Internet Order', p. 4. Edge services (or providers) can also be referred to as 'content and application' providers, see Yemini, 'The New Irony of Free Speech', *Columbia Science and Technology Law Review*, 2018, p. 149.

Non-edge services should refrain from imposing content-based restrictions because of the potential collateral effects. As functioning on the network layer service, the ISP does not have direct access to the content of the transmitted information. The actual content of information only shows at the application layer, allowing manipulation of the content.[240] ISPs that engage in content-based regulation violate their technological neutrality and usually restrict complete services. Because of the lack of control over the information and the risk of overregulation, ISPs should maintain neutrality. This concept is also known as the end-to-end principle, meaning that the internet's core should only be preoccupied with transferring bits and bytes. Functional interventions on the information provided to the service should occur at the edge of the internet.[241]

As Bridy puts it, "the core of the network is agnostic about the type of data it carries, and it treats all the data it carries the in same way."[242] The end-to-end principle and the related principle of neutrality allow developers to build all types of services and programmes without requiring the providers that offer these core functionalities of the internet to adjust their services.[243] Of course, the technological design of the internet and the normative propositions underpinning this design are not a given. According to Lessig, the regulatability of the internet could be increased by "complement[ing] the core with technology that adds regulability".[244] However, Lessig feels more for a second option that "regulates applications that connect to the core" and not the core itself.[245] In other words, Lessig argues that regulation should take place on what Lessig refers to as the "application space".[246] These services (comparable with the application layer and edge service providers) have a similar level of control over user-provided information as traditional information intermediaries.[247] The providers functioning on the application layer are, according to Riordan, the most suitable target for regulation. These services "exercise the most direct control over application content."[248] As noted, the providers that offer application layer services often offer their services at the edges of the internet.[249]

There are also non-edge service providers that have technological control over the content of information available on edge services. For example, edge services that do not possess hosting capabilities (required to store information) depend on other providers. While these hosting service providers could be technologically able to control specific instances of information, they do not function as edge services when other providers are dependent on these hosting services. Besides, these hosting services normally do not have a direct relationship with the user that provided the

---

[240] Bridy, 'Remediating Social Media: A Layer-Conscious Approach', *Boston University Journal of Science and Technology Law*, 2018, p. 205.

[241] Bridy, 'Remediating Social Media: A Layer-Conscious Approach', *Boston University Journal of Science and Technology Law*, 2018, p. 199.

[242] Bridy, 'Remediating Social Media: A Layer-Conscious Approach', *Boston University Journal of Science and Technology Law*, 2018, p. 200.

[243] Bridy, 'Remediating Social Media: A Layer-Conscious Approach', *Boston University Journal of Science and Technology Law*, 2018, pp. 200-201.

[244] Lessig, 2006, *Code Version 2.0*, p. 145.

[245] Lessig, 2006, *Code Version 2.0*, p. 145.

[246] Lessig, 2006, *Code Version 2.0*, p. 145.

[247] However, the consequences of removal on intermediary services on the internet more far-reaching than when a traditional intermediary would because removal of content applies to a platform and not just one medium, see Wu, 'Collateral Censorship and the Limits of Intermediary Immunity', *Notre Dame Law Review*, 2011, p. 314.

[248] Par. 2.57 of Riordan, 2016, *The Liability of Internet Intermediaries*, p. 40.

[249] Par. 2.34 of Riordan, 2016, *The Liability of Internet Intermediaries*, p. 35.

information. The user who uses the service (usually) does not know that the hosting service can technically regulate the content of the information that the user provided to another service operated by another provider. The same is true for so-called caching providers who provide services to speed up access to other services by caching (maintaining copies of the information and software necessary to provide the service) as close to the (potential) user as possible, which increases the availability of the service. While they are technically one of the closest services to the user, they do not qualify as an edge service provider as meant in this paragraph. When functioning as a caching service, the provider of this service depends on the edge provider to provide the information. Like the ISP, the expectation is that a caching provider refrains from content-based regulation.[250] As a rule of thumb, I propose thus that content-based regulation of (user-provided) information should occur on the service that is most recognisable for the user as the regulator.

In sum, the technological dimension of internet intermediary service providers leads to the conclusion that regulation of the content of (user-provided) information should only be enacted on the application layer and at the edge of the internet.[251] Goldman argues that providers that are (legally or functionally) restricted in their available remedies should not be imposed with content-based regulation by regulators since this would likely lead to overregulation.[252] As argued by Balkin, regulating providers on the physical or network layer may have significant adverse effects.[253] Providers offering application layer services are the most suitable targets for content-based regulation of user-provided content.[254] Legislation in the US and the EU reflects this in the legislation enacted to regulate providers: ISPs cannot impose content-based restrictions while application layer services have a much broader discretion.[255] The following paragraph discusses the legal categories used to regulate providers.

---

[250] For example, to remove something from Google's search engine cache part of the (normal) procedure is to first contact the content provider, see Laidlaw, 2015, *Regulating Speech in Cyberspace*, p. 217. Caching services engaging in content regulation may lead to severe freedom of expression rights restrictions since content regulation by caching services normally leads to a termination of services, see Balkin, 'Free Speech is a Triangle', *Columbia Law Review*, 2018, pp. 2038-2039. The passive role of caching providers is also reflected in legislation, see for example, Recital 42 and Article 13 of Directive 2000/31/EC (*Directive on electronic commerce*); Van Eecke, 'Online service providers and liability: A plea for a balanced approach', *Common Market Law Review*, 2011, pp. 1462-1463 and 1482; 17 USCA § 512(b) (West 2010, Westlaw Next through PL 116-179).

[251] Basically, both layering and the end-to-end design of the internet seeks to maintain "[…] 'end-to-end' functionality: that application control is remitted to the computers at the ends of the network and the network transmission and inter-networking protocols are kept as simple as possible.", see Speta, 'A Common Carrier Approach to Internet Interconnection', *Federal Communications Law Journal*, 2002, p. 246.

[252] Goldman, 'Content Moderation Remedies', *Michigan Technology Law Review*, 2021, pp. 49-50.

[253] Note 119 of Balkin, 'Free Speech is a Triangle', *Columbia Law Review*, 2018, p. 2037.

[254] Par. 2.57 of Riordan, 2016, *The Liability of Internet Intermediaries*, p. 40.

[255] Bridy, 'Remediating Social Media: A Layer-Conscious Approach', *Boston University Journal of Science and Technology Law*, 2018, p. 209.

1.3.2 Internet intermediaries: the legal dimension

The legal concepts founded in the 1990s[256] and in the early 2000s[257] to address the roles fulfilled by providers may not be fully applicable to the roles fulfilled by providers in the 2020s. For example, the e-Commerce Directive of 2000, harmonising in which circumstances providers may (at least) be exempted from liability for the content of user-provided information in the EU, distinguishes between three separate roles providers can fulfil.[258] In contrast, the general rule of the US liability regime laid down in Section 230 of the Communications Decency Act of 1996 applies to all "interactive computer services",[259] which is much broader than the three roles distinguished in the EU.

None of the legislation mentioned here relies on "internet intermediary" as a legal concept. Nor is there a (legal) definition of internet intermediary to be found in legislation. According to Dinwoodie, concepts such as "interactive computer services" and "hosting service provider" may partly form a 'proxy' for the concept of an internet intermediary.[260] Dinwoodie argues that, as a concept, "internet intermediary" leaves little room to emphasize the differences between different intermediary roles. There is not simply one type of provider. In addition, such terminology neglects that providers may fulfil many different intermediary functions. Because of this multitude of functions, one provider is (potentially) subjected to multiple regulatory and thus liability regimes.[261] As discussed in the introduction of this chapter, this paragraph sees to generic legislation that applies to providers that deal with user-provided information in the broadest sense. Of course, exemptions on this generic legislation exist as a *lex specialis*,[262] as discussed in Chapters 3 and 4 of this dissertation.

Internet intermediary services and the general safe harbour regime of the EU

The e-Commerce Directive relies on the broader "Information Society services", which are 1) "normally provided for remuneration", 2) "at a distance", 3) "by electronic means", and 4) "at the individual request of a recipient of services".[263] Of course, also non-internet intermediary services fall within the definition of "Information Society services".[264] The exemptions from liability for user-provided information for intermediary services laid down in articles 12 (mere conduit), 13

---

256 Of course, the US statutes and the EU directive listed here are updated over time but still rely on concepts originating from the 90's, see 47 USCA § 230 (West 2018, Westlaw Next through PL 116-91); 17 USCA § 512 (West 2010, Westlaw Next through PL 116-179). Directive (EU) 2015/1535 of the European Parliament and of the Council of 9 September 2015 laying down a procedure for the provision of information in the field of technical regulations and of rules on Information Society services, *OJ L 241*, *17.9.2015* (data.europa.eu/eli/dir/2015/1535/oj).

257 Directive 2000/31/EC (*Directive on electronic commerce*).

258 'Mere conduit', 'caching' and 'hosting', see Article 12 to 14 of Directive 2000/31/EC (*Directive on electronic commerce*).

259 47 USCA § 230(c) and (f) (West 2018, Westlaw Next through PL 116-91).

260 Dinwoodie, 2020, 'Who are Internet Intermediaries?', p. 38 and 41.

261 Dinwoodie, 2020, 'Who are Internet Intermediaries?', pp. 47-48.

262 In the US see, for example, protection of intellectual property law on the internet as laid down in 17 USCA § 512 (West 2010, Westlaw Next through PL 116-179). Intellectual property law is exempted from protection by Section 230, see 47 USCA § 230(e)(2) (West 2018, Westlaw Next through PL 116-91). A similar exemption can be found in the EU, see Recital 65 and Article 17(3) of Directive (EU) 2019/790.

263 Article 1(1)(b) of Directive (EU) 2015/1535.

264 For example, a webshop is not necessarly an internet intermediary service providers.

(caching) and 14 (hosting) read in conjunction with article 15 (the prohibition to impose a general obligation to monitor) of the Directive form the focal point of this discussion.

As noted, many well-known providers perform activities that transcend the three roles of the Directive, which raises the question of whether these activities fall within the safe harbour of the Directive. To recall, in the EU, providers fulfilling one of the three roles distinguished in the Directive can profit from a 'safe harbour' which shields the intermediary from liability for information provided or requested by a user as long as they fulfil a set of criteria. These criteria thus vary between the different intermediary roles.[265] Unlike mere conduit and caching providers, hosting service providers can regulate user-provided information by permanently removing or restricting access to information. In debates over increased regulation of providers, this mainly concerns hosting service providers. As the EC notes, "[i]llegal content on online platforms can proliferate especially through online services that allow upload of third party content."[266] In a later recommendation, the EC emphasised that "[p]roviders of hosting services play a particularly important role in tackling illegal content online, as they store information provided by and at the request of their users and give other users access thereto, often on a large scale."[267]

The safe harbours offered by the Directive are not absolute – they only apply to providers of intermediary services that fit the definition and uphold the requirements regarding user-provided information. For hosting services, this requirement is that "the provider does not have actual knowledge of illegal activity or information and, as regards claims for damages, is not aware of facts or circumstances from which the illegal activity or information is apparent".[268] When the provider gains knowledge of or becomes aware of the illegal or unlawful content of the user-provided information, then the provider must "acts expeditiously to remove or to disable access to the information".[269] These requirements will be discussed more extensively in Chapter 4. For now, it is sufficient to note that the safe harbour does not apply to providers that 1) have knowledge/awareness of the illegal or unlawful content of user-provided information and 2) do not act expeditiously to disable access to this information. Article 14 only applies to "an information society service is provided that consists of the storage of information provided by a recipient of the service".[270] For mere conduit services ("transmission in a communication network of information provided by a recipient of the service, or the provision of access to a communication network")[271] and caching services ("consists of the transmission in a communication network of information provided by a recipient of the service")[272] different requirements apply.

It is not easy to demarcate between the different roles – both factually and legally. For example, the ECJ in 2010 confused scholars and providers with its *Google France* ruling. The ECJ ruled that hosting services cannot rely on the safe harbour when they are not "neutral, in the sense that its conduct is merely technical, automatic and passive, pointing to a lack of knowledge or

---

[265] Articles 12 to 14 of Directive 2000/31/EC (*Directive on electronic commerce*).

[266] Communication COM(2017)555 final, p. 4.

[267] Recital 15 of Commission Recommendation (EU) 2018/334 of 1 March 2018 on measures to effectively tackle illegal content online, *OJ L 63*, *6.3.2018* (data.europa.eu/eli/reco/2018/334/oj).

[268] Article 14(1)(a) of Directive 2000/31/EC (*Directive on electronic commerce*).

[269] Article 14(1)(b) of Directive 2000/31/EC (*Directive on electronic commerce*).

[270] Article 14(1) of Directive 2000/31/EC (*Directive on electronic commerce*).

[271] Article 12(1) of Directive 2000/31/EC (*Directive on electronic commerce*).

[272] Article 13(1) of Directive 2000/31/EC (*Directive on electronic commerce*).

control of the data which it stores."[273] The ECJ seems to suggest that hosting services could only rely on the safe harbour of Article 14 as long it has no involvement in the content of user-provided information. Unfeasible for providers that are often involved in user-provided information. This involvement, for example, exists in offering recommendations of user-provided information with content that may interest the user. As Van Eecke observes, Recital 42 of the Directive caused this confusion because its wording suggests that the requirement of a "mere technical, automatic and passive nature" also applies to hosting services. Such an interpretation would obscure between mere conduit, caching, and hosting service. For hosting services, the bar would be raised to rely on the safe harbour. Van Eecke does not view such a criterion as viable now, "hosting providers will almost necessarily have some degree of involvement with their users."[274] Hosting service providers that offer the possibility for users to upload user content or social networking functionalities to encounter and interact with information from other users go beyond such a "mere technical, automatic and passive nature".[275]

The ECJ clarified in *L'Oréal v. eBay* that only providers that, due to their active role, gain "knowledge of, or control over, the data"[276] lose protection under Article 14. While this would strengthen the safe harbour, this does not resolve the so-called "Good Samaritan-paradox".[277] This paradox expresses that providers of internet intermediary services are discouraged from engaging in voluntary moderation of information with illegal content because this may be too active to rely on the safe harbour laid down in Article 14 of the Directive. Providers may fear that they would gain knowledge of or control over the illegal content they failed to moderate.[278]

While the EC emphasized that voluntary monitoring would not cause providers of hosting services to lose their safe harbour, the EC argued that "in such cases the online platform continues to have the possibility to act expeditiously to remove or to disable access to the information in question upon obtaining such knowledge or awareness."[279] The EC tries to assure hosting service providers that they ought not to worry about liability resulting from moderation out of their own initiative. This assurance, according to Kuczerawy, is "somewhat confusing and perhaps even misleading"[280] since "the EC attempts to convince hosting providers that they will not lose the protection – as long as they act according to the expectations of policy makers."[281] The lack of "Good Samaritan"-protections combined with the request of the EC to proactively take down user-provided information with illegal or unlawful content forces providers to choose between passivity or an active approach accompanied by perfect moderation.

---

[273] Judgement of the Court (Grand Chamber) of 23 March 2010 in *C-236/08, C-237/08 and C-238/08, Google France SARL and Google Inc. v. Louis Vuitton Malletier SA*, ECLI:EU:C:2010:159, in particular Rec. 114.

[274] Van Eecke, 'Online service providers and liability: A plea for a balanced approach', *Common Market Law Review*, 2011, pp. 1482-1483.

[275] Van Eecke, 'Online service providers and liability: A plea for a balanced approach', *Common Market Law Review*, 2011, pp. 1482-1483.

[276] Judgement of the Court (Grand Chamber) in *C-324/09* (*L'Oréal v. eBay*), in particular Rec. 116.

[277] Van Eecke, 'Online service providers and liability: A plea for a balanced approach', *Common Market Law Review*, 2011, pp. 1483-1484.

[278] Kuczerawy, 2018, 'The EU Commission on voluntary monitoring: Good Samaritan 2.0 or Good Samaritan 0.5?'; Van Hoboken & Keller, 2019, 'Design Principles for Intermediary Liability Laws', p. 8.

[279] Communication COM(2017)555 final, p. 12.

[280] Kuczerawy, 2018, 'The EU Commission on voluntary monitoring: Good Samaritan 2.0 or Good Samaritan 0.5?'.

[281] Kuczerawy, 2018, 'The EU Commission on voluntary monitoring: Good Samaritan 2.0 or Good Samaritan 0.5?'.

In sum, the EU framing of providers of internet intermediary services in the e-Commerce Directive is highly ambivalent. The Directive suggests that hosting service providers should keep their distance from the content of user-provided information to count on the safe harbour as an internet intermediary. On the other hand, the EC encourages providers to actively moderate illegal, unlawful, and even harmful information. While the EC signals that providers ought not to worry about liability for moderation as long as they remove information with illegal or unlawful content,[282] the EU regime does not offer an exemption for liability from under- and over-removal. Providers may become liable for information with illegal content they accidentally fail to remove. Besides, the e-Commerce Directive does not offer a safe harbour for user claims against the removal of content that is not unlawful.[283]

Internet intermediaries in the US: the general rule of immunity

In contrast to the EU approach, Section 230 does not distinguish between different services – all providers that offer interactive computer services can rely on the immunities provided under this section. The definition of interactive computer service is

> any information service, system, or access software provider that provides or enables computer access by multiple users to a computer server, including specifically a service or system that provides access to the Internet and such systems operated or services offered by libraries or educational institutions.[284]

All kinds of services that make use of the internet are thus considered interactive computer services by the US courts, including hosting services, internet marketplaces, and dating websites.[285] Providers of intermediary services that fall within this definition, according to Section 230(c)(1), can "not be treated as the publisher or speaker of any information provided by another information content provider."[286] Wilman characterises the protection offered by Section 230 as "extreme" but not as absolute since there are exceptions made to the statute and case law.[287] Chapter 3 of this dissertation discusses the exceptions.

Next to Section 230(c)(1) exempting providers from liability for user-provided information, Section 230(c)(2)(A) offers protection to providers that actively intervene in the

---

[282] See, for example, European Commission, 2021, 'Code of Practice on Disinformation'; European Commission, 2016, 'Code of Conduct on Countering Illegal Hate Speech Online'.

[283] Klos, "Wrongful moderation': Aansprakelijkheid van internetplatforms voor het beperken van de vrijheid van meningsuiting van gebruikers', *Nederlands Juristenblad*, 2020/2976; Van Eecke, 'Online service providers and liability: A plea for a balanced approach', *Common Market Law Review*, 2011, pp. 1467-1468; Kuczerawy, 2018, 'The EU Commission on voluntary monitoring: Good Samaritan 2.0 or Good Samaritan 0.5?'.

[284] 47 USCA § 230(f)(2) (West 2018, Westlaw Next through PL 116-91).

[285] Holland, 'In Defense of Online Intermediary Immunity: Facilitating Communities of Modified Exceptionalism', *University of Kansas Law Review*, 2008, pp. 374-375; Gillespie, 2018, *Custodians of the Internet*, p. 34.

[286] 47 USCA § 230(c)(1) (West 2018, Westlaw Next through PL 116-91).

[287] Par. 6.41 of Wilman, 2020, *The Responsibility of Online Intermediaries for Illegal User Content in the EU and the US*, p. 187.

content provided by their users.[288] Section 230(c)(2)(A) reads that "[n]o provider or user of an interactive computer service shall be held liable on account of"[289]

> any action voluntarily taken in good faith to restrict access to or availability of material that the provider or user considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally protected;[290]

As Goldman notes, it is rare for providers of internet intermediary services to rely on Section 230(c)(2)(A) since it does not offer similar far-reaching protections as Section 230(c)(1). Goldman argues that providers also can rely on their terms of service, which allows them to intervene in user-provided information.[291] In addition, as discussed in the third chapter, the 'free speech clause' of the First Amendment also extends to providers that make editorial decisions regarding user-provided information.[292] However, it could be argued that Section 230(c)(2)(A) may offer protection for civil liability when a provider fails to remove content excluded from Section 230 protection, such as sex trafficking advertisements.[293] Goldman, however, questions whether this argument would hold up in court.[294] Section 230(c)(2)(A), however, reveals the legislator's view of providers as providers that make far-reaching decisions in what user-provided information they do and do not permit on their service.

Evaluation: EU versus the US approach

The approaches in the US and the EU know fundamental differences. Internet intermediaries are 'framed' in legal definitions which have legal meaning. Because the safe harbours and immunities provided by legislation link to the legal definitions, providers may ensure they fall within these legal categories. Altering these definitions may cause internet intermediary providers to change their conduct. Ultimately, this may have consequences for what users may and may not do on their services. The same, of course, is valid for the immunities and safe harbours offered to these categories of providers.

The US chose to keep it simple and only defined one category of internet intermediary services in their generic legislation. Providers that offer an "interactive computer service" can rely on the protection of Section 230(c)(1), which offers immunity for liability as a publisher or speaker for user-provided information. As long as a provider is not "responsible, in whole or in part, for the creation or development of information"[295] offered to the service, Section 230(c)(1) offers

---

[288] 47 USCA § 230(c) (West 2018, Westlaw Next through PL 116-91). The goal of Section 230 was to encourage internet intermediaries to set standards themselves, see Kosseff, 2019, *The Twenty-Six Words That Created the Internet*, pp. 64-66. Section 230, thus, does not require internet intermediaries to be 'neutral' towards the content users provided to them, see Gillespie, 2018, *Custodians of the Internet*, pp. 30-31; E. Harmon, 'No, Section 230 Does Not Require Platforms to Be "Neutral"', *Electronic Frontier Foundation,* 12 April 2018, available at eff.org/deeplinks/2018/04/no-section-230-does-not-require-platforms-be-neutral (retrieved on 15 February 2022).

[289] 47 USCA § 230(c)(2) (West 2018, Westlaw Next through PL 116-91).

[290] 47 USCA § 230(c)(2)(A) (West 2018, Westlaw Next through PL 116-91).

[291] Goldman, 2020, 'An Overview of the United States' Section 230 Internet Immunity', p. 160.

[292] In contrast, 'must-carry' rules prohibiting removal of content are considered unconstitutional, see Keller, 2019, 'Who Do You Sue? State and Platform Hybrid Power over Online Speech', pp. 2 and 9-12.

[293] 47 USCA § 230(e)(5) (West 2018, Westlaw Next through PL 116-91).

[294] Goldman, 'The Complicated Story of FOSTA and Section 230', *First Amendment Law Review*, 2019, p. 283.

[295] 47 USCA § 230(f)(3) (West 2018, Westlaw Next through PL 116-91).

protection for liability for user-provided information.[296] Chapter 3 discusses the specific criteria. For now, it is sufficient to conclude that providers that offer an interactive computer service may rely on Section 230, which protects a (very) broad range of internet intermediary activities.

As noted, the EU e-Commerce Directive relies on the broader definition of "Information Society services".[297] While this comes close to an interactive computer service defined in the US, the EU distinguished between three roles these Information Society services could fulfil. Although the Directive does not define what an intermediary is, the Directive places these three roles under the subheading "Liability of intermediary service providers"'[298] Section 230, unlike the e-Commerce Directive, does not distinguish between mere conduit, caching, and hosting services. Section 230(c)(1) and (2) could also apply to non-hosting services (for example, an ISP that offers filtering of harmful websites on behalf of the user but mistakenly over blocks websites). The majority of the issues discussed here, however, concern interactive computer services that involve hosting. While the usability of the immunities provided under Section 230 may differ amongst different services, the general US approach toward internet intermediary liability gives one definition, including all different internet intermediary functions.[299] This difference in scope is not of concern for this dissertation. The following chapters mainly focus on providers that involve hosting user-provided information.

There is uncertainty over how active a hosting service may get under the e-Commerce Directive. The rule laid down in Section 230 perceives interactive computer services not as "mere technical, automatic and passive"[300] providers but as (potentially) actively involved in the content of user-provided information. As noted, Section 230 offers protection for 'Good Samaritan' moderation of user-provided information while the e-Commerce Directive does not.[301] While the Directive stimulates hosting services providers to take down user-provided information after they become aware of its illegal content, it discourages intermediaries from engaging in content moderation themselves. The ECJ requires hosting service providers not to become too involved when they wish to rely on the safe harbour of Article 14.[302] Hosting providers must prevent gaining

---

[296] During my PhD-project following twitter account of EU and US legal scholars was very helpful. In some cases, they would 'retweet' content provided by other users. Which means that they would share content that other users posted on twitter. In a way these twitter accounts functioned as an intermediary between those users and me.

[297] Article 1(1)(b) of Directive (EU) 2015/1535.

[298] See Section 4 of Directive 2000/31/EC (*Directive on electronic commerce*).

[299] Of course, this may be different in statutes that see to a specific category of content, see 17 USCA § 512(a), (b), and (c) (West 2010, Westlaw Next through PL 116-179).

[300] Judgement of the Court (Grand Chamber) in *C-236/08, C-237/08 and C-238/08* (*Google France*), in particular Rec. 114.

[301] Van Eecke, 'Online service providers and liability: A plea for a balanced approach', *Common Market Law Review*, 2011, pp. 1483-1484.

[302] In *Google France* the ECJ argued that:

> in order to establish whether the liability of a referencing service provider may be limited under Article 14 of Directive 2000/31, it is necessary to examine whether the role played by that service provider is neutral, in the sense that its conduct is merely technical, automatic and passive, pointing to a lack of knowledge or control of the data which it stores.

see Judgement of the Court (Grand Chamber) in *C-236/08, C-237/08 and C-238/08* (*Google France*), in particular Rec. 114. A year later in *L'Oréal v. eBay* the ECJ adopted the similar but slightly changed criteria that in that when:

control or knowledge of the content of user-provided information to rely on the safe harbour. In the EU, passive hosting service providers may rely on the safe harbour offered by Article 14. At the same time, providers that are too active may lose safe harbour protection. While this passive/active distinction may be more a dichotomy than a clear distinction, it is not beforehand clear when an internet intermediary becomes too active. Providers, however, may be active on one part of their service and more passive on other parts. Providers may rely on the safe harbour for the passive parts while forfeiting this right for other parts of the service on which they became too active.[303]

Section 230 and the e-Commerce Directive codify different expectations legislators have of providers of internet intermediary services. The EU in the Directive seems to assume that internet intermediaries have minimal involvement with user-provided content. While the Directive does not prohibit active involvement of hosting services in user-provided information, they may lose their safe harbours protection which offers a powerful incentive to not moderate.[304] The argument could be made that forfeiting the safe harbour does not mean that the provider no longer qualifies as a provider.[305] Service providers, however, are often dependent on safe harbours to prevent legal liability for user-provided information. Since the safe harbour of the Directive is tied to some passivity, this suggests that providers are not expected to become active towards user-provided information. As noted, this is different in the context of Section 230, which left open how active moderation of user-provided information could be by offering exemptions for liability arising from moderation but also from not moderating.[306]

However, Section 230 and the Directive are not so different regarding their expectation of providers of internet intermediary services with respect to filtering out illegal or unlawful content before publication. While the expectations were not made explicit,[307] Section 230(c)(1) does not require a provider to review all user-provided information before admission. Without Section 230, the fear exists that providers would overregulate user-provided information or cease to moderate out of fear of liability.[308] Some services even may close services out of fear of liability.[309] Of course, these fears are not without critics arguing that Section 230 immunities may be too overstretched

---

> the operator has provided assistance which entails, in particular, optimising the presentation of the offers for sale in question or promoting those offers, it must be considered not to have taken a neutral position between the customer-seller concerned and potential buyers but to have played an active role of such a kind as to give it knowledge of, or control over, the data relating to those offers for sale.

see Judgement of the Court (Grand Chamber) in *C-324/09* (*L'Oréal v. eBay*), in particular Rec. 116.

303 J. van Hoboken, et al., *Hosting intermediary services and illegal content online: An analysis of the scope of article 14 ECD in light of developments in the online service landscape*, Luxembourg, Publications Office of the EU, 2018, doi:10.2759/284542, pp. 7, 14 and 31-36.

304 Article 14 of Directive 2000/31/EC (*Directive on electronic commerce*).

305 The provider still qualifies as an "Information Society service".

306 Kosseff, 2019, *The Twenty-Six Words That Created the Internet*, pp. 64-66.

307 Gillespie, 2018, *Custodians of the Internet*, pp. 43-44. See also, on the question whether Section 230 (should) protect 'bad actors' or 'Bad Samaritans', see Citron & Wittes, 'The Internet Will Not Break: Denying Bad Samaritans § 230 Immunity', *Fordham Law Review*, 2017, pp. 416-417.

308 Gillespie, 2018, *Custodians of the Internet*, p. 43.

309 Goldman, 'The Complicated Story of FOSTA and Section 230', *First Amendment Law Review*, 2019, pp. 288-289.

to include bad actors, antitrust conduct, and (potential) avoidance of other legislation.[310] Chapter 5 discusses these criticisms. For now, it is sufficient to state that Section 230 assumes that internet intermediaries may engage in content moderation and does not require them to do so.

As Wilman notes, there is no active encouragement for providers offering interactive computer services to engage in content moderation by offering a safe harbour for moderation decisions. As Wilman puts it, "Section 230(c)(2) principally only removes a potential disincentive for intermediaries to do so."[311] However, incentivising providers to remove content could run into constitutional (in the EU) or human rights issues (both the EU and the US) – especially in the US. According to Keller, this is already the case when the legislation would "foreseeably cause platforms to restrict legal speech".[312] While the encouragement does not consist of a carrot or a stick, it is the best the US legislator could do to offer a moderation-friendly environment for providers of internet intermediary services.

Section 230 presupposes that a provider of an internet intermediary service is not in the position to take full responsibility for all content of user-provided information. However, they can moderate unrestrained behaviour by taking measures against user-provided information with illegal or undesirable content. Section 230 protects providers against liability for user-provided information with only a few exemptions. The e-Commerce Directive has a similar view on internet intermediary services. Like Section 230, the Directive protects providers from the requirement to proactively screen for illegal and unlawful content.[313] The EU approach diverges from Section 230 by making the liability of hosting service providers conditional to having "actual knowledge of illegal activity or information" or being "aware of facts or circumstances from which the illegal activity or information is apparent".[314] While not expressed in Article 14, the expectation codified in the Directive is that hosting services do not have knowledge or awareness of user content by default. Knowledge or awareness is the exception. As the case law of the ECJ shows, hosting services providers are expected to uphold some passivity towards user information to rely on safe harbours.[315] A different explanation in which knowledge and awareness would be the default would render the safe harbour provided by Article 14 useless.

As discussed in the following paragraph, providers fulfil a broad range of intermediary roles and functions on the internet. How these different approaches work out for these providers and their users are discussed in Chapters 3 and 4. The definitions and categories used to define internet intermediary roles codify the policy assumptions in the different intermediary roles. The US and EU approaches are similar in the expectation that internet intermediaries were (and are) not able to check all the content of the information provided and requested by users beforehand. The US and the EU consider that internet intermediary activities are different from traditional intermediary activities in volume. However, how service providers should deal with this user

---

310 For example, Gillespie, 2018, *Custodians of the Internet*, pp. 43-44; Citron & Wittes, 'The Internet Will Not Break: Denying Bad Samaritans § 230 Immunity', *Fordham Law Review*, 2017, pp. 419-423; Pasquale, 'Platform Neutrality: Enhancing Freedom of Expression in Spheres of Private Power', *Theoretical Inquiries in Law*, 2016, pp. 494-496.

311 Par. 4.41 of Wilman, 2020, *The Responsibility of Online Intermediaries for Illegal User Content in the EU and the US*, p. 115.

312 Keller, 2021, 'Six Constitutional Hurdles for Platform Speech Regulation'.

313 Article 15 of Directive 2000/31/EC (*Directive on electronic commerce*).

314 Article 14(1)(a) of Directive 2000/31/EC (*Directive on electronic commerce*).

315 See Judgement of the Court (Grand Chamber) in *C-236/08, C-237/08 and C-238/08* (*Google France*), in particular Rec. 114; Judgement of the Court (Grand Chamber) in *C-324/09* (*L'Oréal v. eBay*), in particular Rec. 116.

content is different. While the US approach is to take away the legal hurdles to not disincentivise providers to become active, the EU e-Commerce Directive codifies the assumption that providers, as a rule, do not have knowledge or awareness of the content provided by their users and thus have a passive relationship with user-provided information. Noteworthy is that the legislation and its interpretation by the ECJ are different from the policy articulated by the EC. In the proposal of the DSA in December 2020, the EC seeks to codify its policy that providers are encouraged to moderate.[316] The DSA proposal also includes a legal definition of 'internet intermediary'.[317] However, as discussed in Paragraph 4.3, this does not change what is expected from providers.

1.3.3 Internet intermediaries: the functional dimension

The concept 'internet intermediary' covers many companies, functions, and activities. As mentioned, it is not always easy to distinguish a company from its intermediary roles and its specific functioning and activities. Intertwining different intermediary functions and non-intermediary activities makes it difficult to consider when a provider functions as an internet intermediary service. Besides, different intermediary functions are governed by different (legislative) norms, as mentioned above. Neither the technological nor de legal dimension provides the whole picture of how internet intermediary services providers relate to user-provided information. The technological dimension only sets out the technological possibilities of providers. The legal dimension complements this by setting out what intermediary roles are regulated and what legal obligations providers have regarding the services they offer. Neither of these dimensions sets out what providers of intermediary services functionally do. The technological dimension clarifies what providers *can*, while the legal dimension sets the boundaries for the legal liability of providers for user-provided information. Therefore, this paragraph offers a functional approach by distinguishing different internet intermediary roles and functions and setting out different intermediary activities.[318]

As with the previous paragraph, the focus lies on regulating user-provided information either by providers or by the state through providers. All providers of internet intermediary services have a level of control over user-provided information and thus could intervene in what, how, and when specific content is allowed. In setting out how different services relate to user-provided information, Balkin distinguishes between three intermediary functions. These functions lead to varying degrees of control, involvement, and legal and technological possibilities to regulate what their users may or may not do their services. Balkin distinguishes between basic internet services, payment services and content curators.[319] Balkin's approach complements and forms an alternative to the two technological approaches discussed in paragraph 1.3.1.[320] As noted above, internet content regulation should only take place on the application layer at the edges of the internet. Besides, only providers that offer hosting services fit the legal categories expected to intervene in user-provided information directly. Balkin's classification of internet intermediary services in "basic internet services", "payment services", and "content curators" serve as a

316 Commission Proposal COM(2020) 825 final (*Digital Services Act*).

317 Dinwoodie, 2020, 'Who are Internet Intermediaries?', pp. 38-39.

318 See, for the functional approach of the EU, Par. 2.35 of Wilman, 2020, *The Responsibility of Online Intermediaries for Illegal User Content in the EU and the US*, p. 26.

319 Balkin, 'Free Speech is a Triangle', *Columbia Law Review*, 2018, p. 2038.

320 Note 119 of Balkin, 'Free Speech is a Triangle', *Columbia Law Review*, 2018, p. 2037.

framework to understand how the internet intermediary service that is offered relates to technological control and legal responsibilities.[321]

While sometimes internet content regulation is technically possible and legally allowed, some providers should refrain from regulating user-provided information because of the character of the service they offer. Balkin, for example, argues that e-mail providers should remain neutral and non-discriminatory with respect to the information sent and received by their users.[322] Grouping intermediaries based on their functional involvement with information helps to understand and discuss (new proposals for) regulation for intermediary liability for user-provided information with illegal or unlawful content.[323] Besides, differentiating between the intermediary roles contributes to understanding the regulatory capabilities of intermediaries with respect to internet content irrespective of technological and legal constraints.

Basic internet services

As discussed, internet content regulation normally takes place on the application layer. The application layer offers the possibility to manipulate the content of the information as shown to the users.[324] The application layer is also the most visible layer for users regarding what service is responsible for interventions on user-provided information. When something happens to an account of specific information provided by the user, the application layer service is often the first point of contact. The application layer is the most visible layer for users. Every internet user can name a few platforms (social media, media sharing platforms). Everyone with an internet connection uses gateways (search engines) to find information. When placing an order on their favourite web shop, a transaction network (payment provider) is used to make the payment.[325]

In contrast, the physical and network layers are usually invisible to users. A user does not see how a server in a data centre on the other side of the world transfers information through a patchwork of cables and internet nodes. A user usually lacks awareness of the domain controllers translating readable website addresses into numbers pointing to the correct (physical) computer. The user typically does not notice that a hosting service hosts a website. The only times a user actively thinks about the ISP is when the monthly bill is due or when the internet connectivity malfunctions. Users of internet intermediary services would not think to address these services when confronted with user-provided information that is removed or made inaccessible.

The most invisible group of internet intermediary services, what Balkin calls basic internet services, form the technical infrastructure of the internet. According to Balkin, hosting, telecommunication, domain name, and caching and defence services are basic internet services.[326] Riordan adds cloud services and certificate authorities to this list.[327] Hosting services are services that consist of hosting the data that is required for other services to function. Telecommunication services are services such as ISPs that provide access to the internet. Domain name services offer a service that consists of registering (such as universiteitleiden.nl) and resolving domain names. Domain names allow users to enter user-friendly addresses in their browsers that point to the

---

321 Balkin, 'Free Speech is a Triangle', *Columbia Law Review*, 2018, p. 2038.

322 Note 119 of Balkin, 'Free Speech is a Triangle', *Columbia Law Review*, 2018, p. 2037.

323 Par. 2.40-2.43 of Riordan, 2016, *The Liability of Internet Intermediaries*, pp. 36-37.

324 Par. 2.29-2.30 and 2.40-2.43 of Riordan, 2016, *The Liability of Internet Intermediaries*, pp. 34 and 36-37.

325 Riordan, 2016, *The Liability of Internet Intermediaries*, pp. 40-46.

326 Balkin, 'Free Speech is a Triangle', *Columbia Law Review*, 2018, p. 2038.

327 Par. 2.46-2.56 of Riordan, 2016, *The Liability of Internet Intermediaries*, pp. 38-40.

website's location. Caching and defence services provide a faster connection, availability, and security features.[328] Certificate authorities function as the notaries of the internet: they issue certificates that confirm that the connection between the user's computer and the server is secured.[329] Cloud services partly overlap with hosting but add other computing services such as caching and defence functionalities to their service.[330] Cloud services can optimise the availability of websites or applications hosted elsewhere.[331] One of the most well-known caching and defence services, Cloudflare, has 'cloud' in its name for a reason.[332] Cloud services are examples of services that may provide network and application layer functionalities.[333]

A provider offering hosting services must not engage in content regulation by making decisions about what is not allowed on the hosting service. Especially when it comes to user-provided information stored on the service by another provider, the hosting service provider should refrain from extensive moderation. An ISP, in its turn, must not engage by itself in filtering instances of information or restricting access to services because of the information available on these services.[334] As noted, regulation of user-provided information on the physical and network layer tends to lead to overregulation because of the lack of control of the providers active on these layers.[335] Unplugging a server might mean that hundred or even thousands of websites are unplugged in the process. Completely blocking a service because of the availability of information with illegal or unlawful content also blocks access to the available legal information. Taking down a whole server or website may only be suitable when it dedicates itself exclusively to information with illegal or unlawful content such as sexual child abuse imagery. Interventions on the physical or network layer are not a suitable option when the aim is to take down one post or a few images on a host that generally provides services for user-provided information with legal content.[336]

While basic internet services should remain neutral regarding the information offered to or through them, Balkin makes an exception for one type of service: domain name services. A domain name must be unique for the system to function. For example, universiteitleiden.nl cannot be registered by two parties at the same time because users would never know on which website they would end up. Neutrality resulting in two users registering the same domain name for their website would mean chaos. Domain name controllers, according to Balkin, should, however, remain neutral with respect to the usage of the domain name.[337] Domain name controllers, thus,

---

328 Balkin, 'Free Speech is a Triangle', *Columbia Law Review*, 2018, p. 2038.

329 Par. 2.56 of Riordan, 2016, *The Liability of Internet Intermediaries*, p. 40.

330 Balkin, 'Free Speech is a Triangle', *Columbia Law Review*, 2018, p. 2038.

331 Par. 2.51 of Riordan, 2016, *The Liability of Internet Intermediaries*, p. 39.

332 See, cloudflare.com.

333 Par. 2.51 of Riordan, 2016, *The Liability of Internet Intermediaries*, p. 39.

334 Balkin, 'Free Speech is a Triangle', *Columbia Law Review*, 2018, pp. 2038-2039.

335 Par. 2.45 and 2.48 of Riordan, 2016, *The Liability of Internet Intermediaries*, p. 38.

336 See, again, For example by the ECtHR, see *Ahmet Yıldırım v. Turkey*, no. 3111/10, § 66, ECHR 2012-VI, 18 December 2012; *Cengiz and Others v. Turkey*, no. 48226/10 and 14027/11, § 64, ECHR 2015-VIII, 1 December 2015; *Kablis v. Russia*, no. 48310/16 and 59663/17, § 94, 30 April 2019; *Engels v. Russia*, no. 61919/16, § 33, 23 June 2020; *Vladimir Kharitonov v. Russia*, no. 10795/14, § 38, 23 June 2020; *OOO Flavus and Others v. Russia*, no. 12468/15, 23489/15 and 19074/16, § 36-39, 23 June 2020.

337 Balkin, 'Free Speech is a Triangle', *Columbia Law Review*, 2018, pp. 2038-2039.

should not refrain from offering services because of the content of information that is made available on or through the service that uses the domain name.[338]

In sum, regulation of the physical and network layer can both lead to overregulation. While the providers that offer the physical and network layer services *can,* they *should not* regulate user-provided information. The difference with internet content regulation on the application layer is that physical and network layer services that regulate user-provided information tend to block entire services, not specific information with illegal or unlawful content.

Payment services

Besides basic internet services, Balkin distinguishes between payment services and content curators.[339] In the taxonomy provided by Riordan, payment services and content curators both function on the application layer.[340]

Payment services are exceptional since, as Balkin notes, "payment systems are not, strictly speaking, layers of internet traffic".[341] There is not an internet layer that deals with payment. Monetary transactions, offline and online, are subjected to other types of regulation.[342] Riordan views payment systems as a subclass of marketplaces on the application layer. Riordan subdivides marketplaces into internet marketplaces (including online marketplaces such as eBay, ticket portals, 'retail emporia' such as Amazon, and app stores) and transaction networks.[343] The last category, transaction networks, is what Balkin seems to have in mind when referring to payment providers. Transaction networks, according to Riordan, encompass "the services and software with which value is transferred between internet users."[344] Riordan counts, for example, card issuers and payment networks (such as Mastercard), online payment systems (for such as PayPal) and micropayment providers to transaction networks.[345]

Technically speaking, payment systems function in the application space and form a clear example of an edge service. When a payment system refuses a user to create an account to make or receive payments or rejects a payment, it is often transparent what provider is responsible for this rejection. Due to their role, there are good reasons to require prudence from payment services regarding internet content regulation.

Payment systems may even be (one of the most) potent internet content regulators.[346] With the help of these payment systems, an internet marketplace can verify the domicile of a customer

---

338 In the past, for example, GoDaddy which also offers domain name controller services, refused to offer services to Gab and the Daily Stormer, see S. Byford, 'Gab.com goes down after GoDaddy threatens to pull domain', *The Verge,* 28 October 2018, available at theverge.com/2018/10/28/18036520/gab-down-godaddy-domain-blocked (retrieved on 14 February 2022); T. Ong, 'Neo-nazi site Daily Stormer threatened by hosting providers and possible hackers', *The Verge,* 14 August 2017, available at theverge.com/2017/8/14/16142384/daily-stormer-site-go-daddy-hosting-providers-hackers-anonymous (retrieved on 15 February 2022).

339 Balkin, 'Free Speech is a Triangle', *Columbia Law Review*, 2018, p. 2038.

340 Riordan, 2016, *The Liability of Internet Intermediaries*, pp. 40-42 and 44-46.

341 Note 119 of Balkin, 'Free Speech is a Triangle', *Columbia Law Review*, 2018, p. 2037.

342 Par. 2.85 of Riordan, 2016, *The Liability of Internet Intermediaries*, p. 45.

343 Riordan, 2016, *The Liability of Internet Intermediaries*, pp. 44-46.

344 Par. 2.83 of Riordan, 2016, *The Liability of Internet Intermediaries*, p. 45.

345 Par. 2.84 of Riordan, 2016, *The Liability of Internet Intermediaries*, p. 45.

346 PayPal, for example, banned accounts related to Trump after the US Capitol was invaded by his supporters, see L. Hautala, 'PayPal and Shopify remove Trump-related accounts, citing policies against supporting violence', *cnet,* 7 April 2021, available at cnet.com/news/paypal-and-shopify-remove-trump-related-accounts-citing-policies-against-supporting-violence (retrieved on 15 February 2022).

or prevent the shipment of goods that are legal in one jurisdiction to a jurisdiction that does not allow these goods. On the other end, customers can profit from payment systems that verify the trustworthiness of an internet marketplace, expecting that payment systems cease services to malicious actors. However, the usage of these regulatory capabilities can also combat forbidden sales or other illegal transactions.[347]

Considering the role of payment systems, Balkin argues that they should refrain from content regulation with only a few exceptions.[348] Payment systems have an enormous potential to influence the exercise of freedom of expression rights. What if a payment system bars transactions to perfectly legal websites because the content is deemed indecent or undesirable by the payment service provider? Balkin, therefore, concludes that payment services must not impose regulations on other providers on what is allowed by refusing payments.[349] Balkin argues that payment systems should only be allowed to intervene when the usage of their services facilitates illegal transactions or other conduct that violates criminal law.[350]

Content curators

The third group of service providers, content curators, are different. Content curators, according to Balkin, "act as curators and personalizers, they cannot really avoid making decisions about content."[351] Neutrality, thus, is not a feasible and even a silly standard for providers of such intermediary services. A lack of neutrality is what characterises content curators. Curation, as noted, encompasses activities that see to "[s]elect, organize, and present (online content, merchandise, information, etc.), typically using professional or expert knowledge."[352] Curation, thus, involves decisions about what information is shown to who, where, and when based on its content.

Balkin groups search engines and social media platforms under content curators that make decisions regarding user-provided information. According to Balkin, search engines and social media platforms perform three functions: firstly, they enable the public to participate. Secondly, content curators organise the public debate. Without search engines and social media platforms, it would be harder to participate in the public debate, and it would be much harder to find information. A third function, according to Balkin, is that search engines and social media platforms, as content curators, offer curation of public opinions. For example, both search engines and social media platforms may offer personalisation of information based on its content.[353] Balkin, however, also includes content moderation in this broad definition of content curation by pointing out that community guidelines allow content curators to enforce norms to safeguard a civil discussion.[354]

Both curation and moderation are application layer activities. Search engines (as gateways), for example, may personalise the results of their users. Internet marketplaces could learn what

---

347 Par. 2.86 of Riordan, 2016, *The Liability of Internet Intermediaries*, pp. 45-46.
348 Balkin, 'Free Speech is a Triangle', *Columbia Law Review*, 2018, p. 2039.
349 Balkin, 'Free Speech is a Triangle', *Columbia Law Review*, 2018, p. 2039.
350 Balkin, 'Free Speech is a Triangle', *Columbia Law Review*, 2018, p. 2039.
351 Balkin, 'Free Speech is a Triangle', *Columbia Law Review*, 2018, p. 2041.
352 Lexico, 'Meaning of curate in English'.
353 Balkin, 'Free Speech is a Triangle', *Columbia Law Review*, 2018, p. 2041.
354 Balkin, 'Free Speech is a Triangle', *Columbia Law Review*, 2018, p. 2041.

their users buy and recommend new products or advertisements to their users.[355] Regarding content moderation, online platforms and marketplaces (allowing third-party sellers) seem to stand out: they can best decide what is allowed and what is not and are often legally required to do so.[356] Technically, there is little reason to assume that only application layer providers engage in content moderation. Content-based interventions are possible when there is technological control over the application layer.

A functional approach toward internet intermediary service providers

The expectation is that providers that offer platform and gateway functionalities engage in content curation.[357] Next to these two functionalities, the expectation exists that providers that function as an internet marketplace are involved in the content of user-provided information.[358] In contrast, providers that offer transaction networks (or payment services) carrying out content-related interventions pose a risk similar to content-based interventions by basic internet services. Like basic internet services, payment services are 1) unmissable and 2) more likely to impose restrictions on a service or account level.[359]

Large internet companies usually do not limit themselves to one intermediary function. For example, Amazon, Apple, and Google offer services within multiple groups. Google and Amazon both offer cloud services.[360] Amazon, Apple, and Google all three offer payment services,[361] and all these providers engage in content curation on a multitude of different platforms.[362] Sometimes these functionalities provided by services are related to each other (all online platforms rely on hosting services). These services are not necessarily related to each other, such as payment services. A provider may offer a payment service for users in and outside its ecosystem. The payment service provider may restrict its usage by setting standards on what goods and services can be purchased. The payment service is no longer just an ancillary functionality; it becomes a stand-alone service forming new points of regulation.

A functional approach thus requires reviewing what functions are ancillary to the provider's primary service. For example, in the case of a social media platform, hosting is a necessary but subordinate function to social media networking functions. Hosting is unmissable for social media platforms, but it becomes obsolete when the provider cancels its social networking functions. Restricting providers of social media platforms from curating or moderating because they also fulfil a basic internet service role would be too strict. The opposite is true for payment services used by services offered by the same provider and outside the service. In the latter case, content-based regulation through payment services also leads to the regulation of information on services other services than those offered by the provider.

The functional approach thus requires uncovering how intermediary functions relate to each other. Providers that offer a service fulfilling a subordinate function and have little direct involvement in the content of user-provided information offered on or to other services must

---

355 Riordan, 2016, *The Liability of Internet Intermediaries*, pp. 40 and 43-46.

356 Riordan, 2016, *The Liability of Internet Intermediaries*, pp. 40-44.

357 Balkin, 'Free Speech is a Triangle', *Columbia Law Review*, 2018, p. 2041.

358 Par. 2.78-2.82 of Riordan, 2016, *The Liability of Internet Intermediaries*, pp. 44-45.

359 Par. 2.85-2.86 of Riordan, 2016, *The Liability of Internet Intermediaries*, pp. 45-46.

360 See cloud.google.com and aws.amazon.com.

361 See pay.amazon.com, apple.com/apple-pay, and pay.google.com/about.

362 See amazon.com, apple.com/app-store, play.google.com.

refrain from content-based regulation. They are functionally not the designated service providers to engage in such moderation. However, this does not mean that these providers are always technically incapable or have no legal obligations to engage in such moderation.

Conclusion

As Dinwoodie noted, terminology such as "internet intermediary" offers little precision.[363] As discussed in this chapter, providers differ in terms of their technological capabilities, legal obligations, and functional involvement regarding the content of user-provided information. According to Dinwoodie, alternative terminology such as "online service provider" and "internet service provider" may be more suitable to describe the actors that function as internet intermediaries.[364] In this chapter, I distinguished between the provider of an intermediary service, the internet intermediary service provided, and the specific activities, roles, and functions these providers may fulfil. Unlike catch-all terminology such as "internet intermediary", this distinction in provider, service and activities allows discriminating between the different roles the providers fulfil. As noted, providers may be active in content moderation on one part of their service while remaining passive on other parts. In addition, providers may provide services that are passive per se due to their technological nature or legal responsibilities. A provider offering e-mail and social networking services is potentially subject to different (legal) norms. While providers of internet intermediary services can vastly differ from traditional information intermediaries, they could also be remarkably similar in terms of control. Providers thus could differ extensively from each other in terms of control over user-provided information. Even one provider offering two different services may have various levels of control over the content of the information.

Distinguishing between different providers, intermediary services, and the roles they fulfil is thus essential in the case of integration of services or when roles are interwoven but can be used independently of each other (in the case of usage of a payment service outside of the marketplace). In the case of usage of a payment service outside the providers' ecosystem, content-based regulation through this payment service could also affect the content on other services.

As discussed in the introduction, the main interest of this dissertation is in the providers that function as content curators. As discussed in this chapter, content curators are providers that can exercise direct control over the content of user-provided information. This criterion of direct control is stricter than only allowing application layer or edge service providers to regulate the content of user-provided information. For example, payment service providers – normally – do not have direct access to the content of the information while they do function on the application layer service. Providers that function as content curators (such as online platforms and search engines) know no technological constraints in how they moderate or curate user-provided information. Unlike services that operate on the physical or the network layer of the internet, they have the technological capabilities to regulate user-provided information. Content curators are, unlike other services, not legally required or functionally obligated to remain neutral towards user content.[365] Content curation services providers offer a service directed at the user to

---

[363] Dinwoodie, 2020, 'Who are Internet Intermediaries?', p. 45.

[364] Dinwoodie, 2020, 'Who are Internet Intermediaries?', p. 45.

[365] The Directive only 'protects' intermediaries of a "mere technical, automatic and passive nature", see Recital 42 of Directive 2000/31/EC (*Directive on electronic commerce*). This is – unfortunately – sometimes explained as 'neutrality. For example, in Judgement of the Court (Grand Chamber) in *C-324/09* (*L'Oréal v. eBay*), in particular Rec. 116; Judgement of the Court (Grand Chamber) in *C-236/08, C-237/08 and C-238/08* (*Google France*), in particular Rec. 114. In these judgements an "active role" is confused or at least conflated with "neutrality".

help the user encounter information with relevant content. Providers that offer internet intermediary services encompassing curation have a vital function. Because of this importance, it is necessary to set out how these curators regulate user-provided information.

2 Internet content regulation: between legal harms and illegal remedies

Introduction

Not only the internet intermediary services landscape has grown dense. Since 2000 the regulation targeting providers and the user-provided information they handle has rapidly increased. This regulation increase could partly be explained by providers expanding the services offered. In the 1990s, an online bulletin board with text and low-resolution images formed a large portion of the internet intermediary landscape. Nowadays, providers are active in almost every aspect of the information landscape. Providers invented new services that did not exist before the internet, but they also disrupted old services by offering legal (but sometimes illegal) online alternatives.[366]

The types of services and the functionalities offered by these services were expended. Search engines became 'smarter' by recommending search results tailored to individual users.[367] Social media platforms fostered meaningful contacts by recommending information from others that the user in question holds dear. Providers began to curate user-provided information for their users.[368] According to some, not always for good. The downside is that users may give up their privacy by allowing providers to harvest their data to allow providers to feed personalised recommendations.[369] Next to privacy risks, there are some risks identified for the democratic process as well.[370]

Against this background, policy proposals influence how providers handle user-provided content. As noted in the first chapter, these proposals are tied to the exceptional nature of providers. Internet intermediary regulation has to relate to the exceptionalist statutes that are enacted. In the US and the EU, legal provisions limit the liability of internet intermediaries for third-party content.[371] These provisions introduce some path-dependency in regulating internet content. Internet intermediaries can be made liable by limiting or abolishing exceptionalist statutes offering immunity (US) or 'safe harbours' (EU). While these statutes refer to fostering a freedom of expression-friendly environment,[372] economic growth and internet innovations were also on the

---

366 Spotify, for example, is a legal alternative to the compact disc. However, services that allow streaming music or television shows from illegal sources, are not so legal.

367 J. Hull, 'Google Hummingbird: Where No Search Has Gone Before', *Wired,* 15 October 2013, available at wired.com/insights/2013/10/google-hummingbird-where-no-search-has-gone-before (retrieved on 15 February 2022).

368 Klos, 2021, 'Closed Online Communities and Freedom of Speech', pp. 195-200.

369 S. Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, London, Profile Books, 2019, pp. 93-97.

370 D. Susser, B. Roessler & H. Nissenbaum, 'Technology, autonomy, and manipulation', *Internet Policy Review*, Vol. 8, No. 2, 2019, doi:10.14763/2019.2.1410, p. 11.

371 In the US 47 USCA § 230(c)(1) and (2) (West 2018, Westlaw Next through PL 116-91). In the EU Article 14 of Directive 2000/31/EC (*Directive on electronic commerce*). Goldman refers to Section 230 as 'a flagship example of mid-1990s efforts to preserve Internet utopianism.' see Goldman, 2010, 'The Third Wave of Internet Exceptionalism', p. 165.

372 In the case of the EU see, Van Eecke, 'Online service providers and liability: A plea for a balanced approach', *Common Market Law Review*, 2011, pp. 1464-1465; Recital 9 and 46 of Directive 2000/31/EC (*Directive on electronic commerce*). For the US see Wu, 'Collateral Censorship and the Limits of Intermediary Immunity', *Notre Dame Law Review*, 2011, pp. 315-318.

minds of the legislators.[373] New legislation may depart from these initial goals in favour of new ones. Of course, lawmakers then must clarify why these goals are no longer relevant or why new goals are more important than the old ones.

As set out, the new internet intermediary regulation also reflects exceptionalism in how service providers are made responsible for upholding due process requirements in dealing with user-provided information. Especially in the EU, service providers are made responsible for combating content provided by users with an illegal or unlawful character (for example, terrorist content)[374] and preventing specific harms (manipulation of elections by spreading misleading or wrong information).[375] These responsibilities do not always involve legal liability. Instead, the EC concludes legally non-binding codes as a form of self-regulation while warning that failing to uphold these codes may lead to legislation.[376] When the EU chooses legislative instruments, such legislation often requires member states to enact legislation to back these instruments by an administrative fine.[377] For example, some obligations regarding terrorist content are backed by "financial penalties of up to 4% of the hosting service provider's global turnover of the last business year."[378] Of course, this legislation is meant as an addition to or a harmonisation of legislation by EU member-states.

Choosing what actors are regulated, by what instruments, and the scope of these instruments may have severe effects. Therefore, this chapter explores the (international) scope of the instruments and remedies the targets of internet content regulation can deploy in regulating user-provided information. In this chapter, first, the actors that are made responsible for content regulation are discussed. Then the instruments these actors can deploy are discussed, followed by the remedies that providers can impose. Lastly, the scope of these remedies is discussed. The scope deals with the potential (international) effect of content regulation which raises freedom of expression concerns due to the differences in standards of what does and does not fall within reach of these rights.

2.1 Target: bad actors or good intermediaries (or the other way around)

As discussed in the first chapter, the providers and the services they provide are differently regulated than traditional information intermediaries such as newspapers. Providers are

---

373 For the EU, see Recital 2 and 60 of Directive 2000/31/EC (*Directive on electronic commerce*). For the US, see 47 USCA § 230(b)(1) (West 2018, Westlaw Next through PL 116-91).

374 Some categories of hate speech where already criminalised, the specific responsibilities of internet intermediaries are laid down in a non-binding 'code of conduct', see Council Framework Decision 2008/913/JHA; European Commission, 2016, 'Code of Conduct on Countering Illegal Hate Speech Online'. Online terrorist content, however, is a new category laid down in a new regulation adopted in April 2021, see Regulation (EU) 2021/784.

375 Disinformation is new category of content with obligations for internet intermediaries laid down in a non-binding 'code of practice', see European Commission, 2021, 'Code of Practice on Disinformation'. This non-binding 'code of practice', however, may have indirect legal effect as two courts cases in the Netherlands show, see Rb. Amsterdam (vzr.), 9 September 2020, ECLI:NL:RBAMS:2020:4435, Rec. 4.4-4.5 and 4.11 (*YouTube*); Rb. Amsterdam (vzr.), 13 October 2020, ECLI:NL:RBAMS:2020:4966, Rec. 4.24, *Computerrecht* 2021/66, m.nt. M. Klos (*Facebook*).

376 As Commissioner Věra Jourová stated with respect to transparency requirements laid down in the Code of Practice on Disinformation: 'The time has come to go beyond self-regulatory measures.' see European Commission, 2020, 'Disinformation: EU assesses the Code of Practice and publishes platform reports on coronavirus related disinformation'.

377 See, for example, Article 18 of Regulation (EU) 2021/784.

378 Article 4(1) and (2) and 18(4) of Regulation (EU) 2021/784.

exceptional, legitimising exceptions to and even some immunities for liability. As shown, the exceptionalism of the providers is expressed in equally exceptional regulation. The observation that providers are exceptionally regulated is thus supported by the nature of these providers. As discussed in the previous chapter, providers that offer application-layer services are in the best position to regulate the content of user-provided information because these providers have actual control over this content. The possibility of control raises the question of to what extent service providers could be held liable for user-provided information that is illegal or otherwise unlawful now control suggests (legal) responsibility.[379] Such legal responsibility comes next to or in the place of the user's responsibility. The question, thus, is who can target who with internet content regulation?

2.1.1 Internet intermediary liability regimes

In assuming legal responsibility for providers, the user's role must not be forgotten. As discussed in this paragraph, it is possible to distribute liability for user-provided content between the provider(s) and the services' user(s). However, how providers are regulated causes regulators to neglect the role of the users – which is exceptional with respect to offline intermediaries. In this respect, Balkin distinguishes between "old-school" and "new-school" speech regulation.[380] Balkin defines "old-school speech regulation" as government regulation directly aimed at individuals or legal entities through "threats of fines, penalties, imprisonment, or other forms of punishment or retribution".[381] In contrast, "new-school speech regulation" targets an intermediary "to get the infrastructure to surveil, police, and control speakers."[382] While old-school regulation targets the offender, new-school regulation explicitly targets the intermediary to regulate the offender. In other words: the provider is targeted by the internet content regulation to regulate user-provided content. The provider is made liable for the content of user-provided information besides or in the place of the responsible user.[383]

Providers could be made liable for user-provided information in numerous ways. Gillespie distinguishes between "strict liability", "conditional liability", and "broad immunity".[384] Providers subjected to strict liability are directly liable for the illegal or unlawful content of user-provided information. According to Gillespie, an example of strict liability forms the internet intermediary liability regime in China. In China, providers must take a proactive role in preventing user-provided information with illegal or unlawful content from being published on their service. When they fail to do so, they instantly become liable for the content of user-provided information. As the opposite of strict liability, broad immunity lies on the other side of the continuum. Broad immunity means that providers cannot be held liable for the content of user-provided information.[385] An example of such a broad immunity approach is US Section 230, which prevents civil liability of

---

[379] See, for example, *Delfi AS v. Estonia* [GC], no. 64569/09, § 157, ECHR 2015-II, 16 June 2015.

[380] Balkin, 'Free Speech is a Triangle', *Columbia Law Review*, 2018, p. 2015.

[381] Balkin, 'Free Speech is a Triangle', *Columbia Law Review*, 2018, p. 2015.

[382] Balkin, 'Free Speech is a Triangle', *Columbia Law Review*, 2018, pp. 2015-2016.

[383] Which may be a failure of the regulators to regulate the repsonsible party, see M.R. Leiser, 'Regulating computational propaganda: lessons from international law', *Cambridge International Law Journal*, Vol. 8, No. 2, 2019, doi:10.4337/cilj.2019.02.03, p. 221.

[384] Gillespie, 2018, *Custodians of the Internet*, p. 33.

[385] Gillespie, 2018, *Custodians of the Internet*, p. 33.

providers for user-provided information with only a few exceptions.[386] The third option distinguished by Gillespie, conditional liability (or conditional immunity), takes a middle position between strict liability and broad immunity.[387] As discussed in Chapter 4, conditional immunity forms the EU approach toward internet intermediary regulation.[388] Conditional liability (or immunity) regimes have in common that a provider cannot be held legally liable for the content of user-provided information as long as they do (or do not) fulfil a set of conditions.[389] Conditional liability can also be understood as conditional immunity: a provider can count on the safe harbour as long as the provider maintains some distance from the content of the user-provided information.[390] These liability regimes imply an allocation of (legal) responsibility between the provider and its users, which will be discussed in the next paragraph.

2.1.2 Allocating liability: between responsibility and effectiveness

Who is responsible for the content of user-provided information? While this may seem a principal discussion, the allocation of legal liability between the provider and the user of the service is fuelled by practical concerns. Providers (usually) do not materially contribute to the illegal or unlawful content of the information provided by users. Most providers do not have knowledge or awareness of the illegal or unlawful content of user-provided information. At the same time, the provider may be in the best position to remedy the harmful effects of such content. The internet as a global network makes it hard for affected individuals and nation-states to hold the responsible parties accountable. The relative anonymity the internet provides to users makes it hard to reveal the identity of the person that provided the information. The legal procedures are lengthy when successful, while the content may cause harm every minute it remains up.[391]

In the case of defamatory content (content that is, for example, slanderous or libellous aiming to hurt the good reputation of an individual), the distribution of liability between the user and the providers may have far-reaching consequences for the possibility for the affected party to pursue effective enforcement of their rights. Perry and Zarsky distinguish five liability models for civil claims based on defamation law.[392]

In the first model distinguished by Perry and Zarsky, neither the provider nor the user responsible for the user-provided information could be held liable for the illegal or unlawful content of the information. Perry and Zarsky quickly dismiss this option since they did not find any examples of such a liability regime in the real world.[393] Such a liability regime would (of course) be highly undesirable and potentially incompatible with international human rights standards that

---

386 At least in the context of 47 USCA § 230(c) (West 2018, Westlaw Next through PL 116-91); The DMCA follows a different approach in 17 USCA § 512 (West 2010, Westlaw Next through PL 116-179).

387 Gillespie, 2018, *Custodians of the Internet*, p. 33.

388 Articles 12, 13 and 14 of Directive 2000/31/EC (*Directive on electronic commerce*).

389 Gillespie, 2018, *Custodians of the Internet*, p. 33.

390 See, for example, Judgement of the Court (Grand Chamber) in *C-324/09* (*L'Oréal v. eBay*), in particular Rec. 116.

391 Regulating providers as gatekeepers for illegal and unlawful user-provided content may reduce costs of enforcement while potentially increasing the incentive to prevent social harms, see J. Riordan, 'A Theoretical Taxonomy of Intermediary Liability', *in* G. Frosio (Ed.) *Oxford Handbook of Online Intermediary Liability*, Oxford, Oxford University Press, 2020, doi:10.1093/oxfordhb/9780198837138.013.3, pp. 75-76.

392 R. Perry & T.Z. Zarsky, 'Who Should Be Liable for Online Anonymous Defamation?', *University of Chicago Law Review Dialogue*, Vol. 82, 2015, p. 163.

393 Perry & Zarsky, 'Who Should Be Liable for Online Anonymous Defamation?', *University of Chicago Law Review Dialogue*, 2015, p. 163.

put a high premium on the protection of individual rights that may be impacted by the expressions of others on the internet.[394] A second model that can be easily dismissed is "exclusive indirect liability", which is also not used at large. This liability regime only imposes liability on the provider while the user that provided the information with illegal or unlawful content is exempted from liability.[395] The moral argument can be made that it is odd that the person to blame cannot be held legally accountable while the provider is.

The third and fourth models described by Perry and Zarsky have a more significant impact on internet intermediary liability regimes due to their usage by the US and the EU. "Exclusive direct liability" only imposes liability on the user responsible for the content of the information while exempting the provider from liability (which forms the US approach towards internet intermediary liability).[396] The fourth model, "concurrent liability", imposes liability on both the user that provided the information and the service provider. This fourth model forms the EU approach towards internet intermediary liability.[397]

While these two models are popular, all four liability regimes know significant pitfalls. Perry and Zarsky propose a fifth model, "residual indirect liability", as an alternative to the first four models. In this model, Perry and Zarsky argue that "the speaker is exclusively liable, but if he or she is not reasonably reachable, the content provider becomes liable."[398] In other words, the responsibility and thus the liability for user-provided information is placed where it belongs: the user as the responsible party for the existence of the illegal or unlawful content in the first place. When the user is not "reasonably reachable", the provider that offers the service to the user becomes liable instead.[399]

While Perry and Zarsky concern themselves with civil liability for defamatory content, the Dutch Criminal Code knows a similar regime for the criminal liability of printers and publishers. When the publication is not accompanied by identifying information of the author, the printer or publisher may be prosecuted for criminal participation. However, the publisher or printer could prevent prosecution by revealing the author after being requested by the examining magistrate.[400] The plus side of this approach is that enforcement becomes less costly for providers while users' freedom of expression rights is better protected than under concurrent liability. Service providers are only required to check or remove user-provided content when the user in question fails or

---

394 For example, the ECtHR, "acknowledges that important benefits can be derived from the Internet in the exercise of freedom of expression," but, the ECtHR "is also mindful that the possibility of imposing liability for defamatory or other types of unlawful speech must, in principle, be retained, constituting an effective remedy for violations of personality rights.", see *Delfi AS v. Estonia* [GC], no. 64569/09, § 110, ECHR 2015-II, 16 June 2015; M. Husovec, 'General monitoring of third-party content: compatible with freedom of expression?', *Journal of Intellectual Property Law & Practice*, Vol. 11, No. 1, 2016, doi:10.1093/jiplp/jpv200, p. 20.

395 Perry & Zarsky, 'Who Should Be Liable for Online Anonymous Defamation?', *University of Chicago Law Review Dialogue*, 2015, pp. 167-168.

396 Perry & Zarsky, 'Who Should Be Liable for Online Anonymous Defamation?', *University of Chicago Law Review Dialogue*, 2015, p. 163.

397 Perry & Zarsky, 'Who Should Be Liable for Online Anonymous Defamation?', *University of Chicago Law Review Dialogue*, 2015, p. 170.

398 Perry & Zarsky, 'Who Should Be Liable for Online Anonymous Defamation?', *University of Chicago Law Review Dialogue*, 2015, p. 172.

399 Perry & Zarsky, 'Who Should Be Liable for Online Anonymous Defamation?', *University of Chicago Law Review Dialogue*, 2015, p. 172.

400 Article 53 and 54 of Wetboek van Strafrecht (Dutch Criminal Code).

refuses to provide identifying information. The fifth model, "residual indirect liability", also remedies a severe pitfall of "exclusive direct liability", which leaves those harmed by illegal or unlawful content emptyhanded because the provider is not liable for the content of user-provided information. The user that provided the information may hide in a veil of anonymity. Of course, "residual indirect liability" also has some downsides. Perry and Zarsky warn that the approach laid down in this model may require balancing with other rights such as privacy rights because the provider may ask for identifying information from all users to avoid liability.[401] Such a balance may not be easy in jurisdictions that put a high premium on privacy rights.

2.1.3 Size, function, or the content of information

The previous two paragraphs discussed regulating the content of user-provided information by exposing providers of internet intermediary services to legal liability for illegal or unlawful content. While exposing providers to legal liability for the content of user-provided information is a popular regulatory instrument, making providers liable often impacts a broad range of different providers. These providers may be very different in terms of userbase, revenue, or the services they offer to their users. Because providers are pretty different, imposing regulation on all providers may have severe unintended side effects, which may work counterproductive. For example, providers that are new or have a small crew are unlikely to adhere to the same level of compliance as very large providers.[402]

Exposing all providers to legal liability is not the only way state actors can regulate user-provided information on services. Providers can also be regulated by imposing obligations directly on their capacity as an intermediary upon fulfilling a predefined set of criteria. An advantage of such regulation is that it allows more differentiation between providers. Some legislation may impose norms on all providers, all services, and all activities, while other regulations may differ between types of services or specific activities. Some regulation only targets specific services such as social media platforms or video platforms. Other regulations may consider the size of the provider in terms of active users or revenue.[403] In addition, internet intermediary regulation may target specific types of infringements or illegal or unlawful content.[404]

As noted, it does matter *how* providers are regulated. Imposing legal liability to providers may lead to unintended and (perhaps) unwanted removal of user-provided information that contains content that is not illegal or unlawful. Such interventions on user-provided information

---

[401] Perry & Zarsky, 'Who Should Be Liable for Online Anonymous Defamation?', *University of Chicago Law Review Dialogue*, 2015, pp. 173-174.

[402] For example, a large online platform such as Facebook required 30 000 moderators in 2020, see C. Jee, 'Facebook needs 30,000 of its own content moderators, says a new report', *Technology Review,* 8 June 2020, available at technologyreview.com/2020/06/08/1002894/facebook-needs-30000-of-its-own-content-moderators-says-a-new-report (retrieved on 15 February 2022).

[403] Such regulation, however, may provoke measures undertaken by providers that are not unintended nor desired by regulation, see E. Goldman & J. Miers, 'Regulating Internet Services by Size', *CPI Antitrust Chronicle*, 2021 (available at ssrn.com/abstract=3863015), p. 7.

[404] For example, Regulation (EU) 2021/784.

are referred to as "over-removal",[405] "over-censorship",[406] and "over-blocking".[407] These phenomena may be caused by how internet intermediary services are regulated by governmental actors or by how content moderation is shaped internally by the service provider.[408] The common denominator of this "collateral censorship"[409] is, according to Felix Wu, that "a (private) intermediary suppresses the speech of others in order to avoid liability".[410]

Collateral censorship thus also impacts user-provided information with legal content. Content that may be protected under freedom of expression rights.[411] Such over-removal may be out of fear of liability, but according to Keller, also to "spare […] the operational expense of paying lawyers to assess content."[412] For a provider, the (legal) costs are lower when they overregulate borderline content than risking legal liability. This risk may, of course, be higher when small providers are targeted by such regulation. Some proposals for new legislation recognise that smaller providers may be less able to bear such legal responsibilities. Very large service providers are targeted with new obligations in these proposals,[413] while smaller services are even excluded.[414] Other proposals for legislation do not differentiate between the size of different providers.[415]

When it is hard for the provider to assess whether the content of user-provided information is illegal, there is a significant risk of overregulation. Citron, for example, notes that hate speech, terrorist content, and extremist speech are highly ambiguous and context-dependent. Because of ambiguous concepts and this context-dependency, there is a clear risk of overremoval – mainly when providers are nudged or forced to deploy automatic means to detect such content.[416] The content of user-provided information may seem illegal (infringement of intellectual property rights).[417] However, facts or circumstances may derogate from its illegality (the right to cite).[418]

Legislators do not only distinguish between services and the content of user-provided information but also between platform functionalities. As noted, providers may be regulated as mere conduit, caching, or hosting service providers. Such regulation differentiates the level of involvement of the provider in user-provided information. Regulation, however, can also target

---

405 Keller, 2020, 'Empirical Evidence of "Over-Removal" by Internet Companies Under Intermediary Liability Laws'.

406 T. McGonagle, 'Free Expression and Internet Intermediaries: The Changing Geometry of European Regulation', *in* G. Frosio (Ed.) *Oxford Handbook of Online Intermediary Liability*, Oxford, Oxford University Press, 2020, doi:10.1093/oxfordhb/9780198837138.013.24, p. 483.

407 Benedek & Kettemann, 2020, *Freedom of Expression and the Internet*, pp. 127-128.

408 For example, because the policy is not available in the language of the content that is considered by a moderator, see 10. Policy recommendation of Oversight Board, 'Case decision 2021-007-FB-UA', *Oversight Board,* 11 August 2021, available at oversightboard.com/decision/FB-ZWQUPZLZ (retrieved on 15 February 2022).

409 Balkin, 'Free Speech is a Triangle', *Columbia Law Review*, 2018, pp. 2016-2017.

410 Wu, 'Collateral Censorship and the Limits of Intermediary Immunity', *Notre Dame Law Review*, 2011, p. 295.

411 Balkin, 'Free Speech is a Triangle', *Columbia Law Review*, 2018, pp. 2016-2017; Council of Europe, 2021, 'Content moderation: best practices towards effective legal and procedural frameworks for self-regulatory and co-regulatory mechanisms of content moderation', p. 25.

412 Keller, 2019, 'Who Do You Sue? State and Platform Hybrid Power over Online Speech', p. 3.

413 Article 25 of Commission Proposal COM(2020) 825 final (*Digital Services Act*), p. 59.

414 Article 16 of Commission Proposal COM(2020) 825 final (*Digital Services Act*), p. 53.

415 Health Misinformation Act of 2021, S. 2448, 117th Cong. (2021).

416 Citron, 'Extremist Speech, Compelled Conformity, and Censorship Creep', *Notre Dame Law Review*, 2018, pp. 1052-1055.

417 Article 17(4) of Directive (EU) 2019/790.

418 Article 17(7) of Directive (EU) 2019/790.

what categories of user-provided information the providers allow users to provide. For example, the EU Audiovisual Media Services Directive obligations only apply to video-sharing platform providers.[419] These providers must be "devoted to providing programmes, user-generated videos, or both, to the general public, for which the video-sharing platform provider does not have editorial responsibility" when the provider organises these videos by deploying algorithms.[420]

Thus, providers can be regulated based on their size (monthly active users, number of employees, annual turnover), the intermediary services they offer (for example, a video-sharing platform service), and the content of user-provided information. The first two categories (size and platform functionalities) form an example of direct regulation of the provider. The last category (the content of user-provided information) may be either direct (imposing obligations on providers because of their capacity as providers) or indirect (imposing liability to everyone who may deal with such content). All types of regulation may have the unintended consequence that providers may adjust their conduct so that they no longer fall within these categories. Would regulating very large service providers hamper their growth? Would it cause new providers to refrain from offering specific services because the cost of legal compliance is too high? Or would providers ban specific content altogether out of fear of liability?

2.1.4 Soft regulation of providers

Providers do not only engage in regulating user-provided information because of state legislation. Service providers may regulate the content of user-provided information out of various motives.[421] As Balkin notes, services providers exercise a form of "private governance" over "online speakers, communities, and populations".[422] While governments may be a potent regulators of user-provided information, they are nowhere without their governors.[423] These governors do not only moderate user-provided information for illegal or unlawful content because they are legally required to do so. Service providers also voluntary regulate user-provided information for content that is not illegal or unlawful but deemed undesirable.

In numerous examples, some state pressure can be identified when providers prohibit content of user-provided information that they were not legally required to do so. One of the examples is disinformation policies that followed concerns over election interference[424] and Covid-19-disinformation.[425] Service providers were not legally required to enact these policies. The government, however, did request providers to enact policies prohibiting these categories of

[419] Article 28(a) of Directive (EU) 2018/1808.

[420] Article 1(1)(aa) of Directive (EU) 2018/1808.

[421] For example, economic reasons, see Gillespie, 2018, *Custodians of the Internet*, p. 35.

[422] Balkin, 'Free Speech is a Triangle', *Columbia Law Review*, 2018, p. 2021.

[423] As Klonick calls them, see Klonick, 'The New Governors: The People, Rules, and Processes Governing Online Speech', *Harvard Law Review*, 2018.

[424] Communication COM(2018)236 final, pp. 11-12; Ž. Švedkauskas, C. Sirikupt & M. Salzer, 'Russia's disinformation campaigns are targeting African Americans', *The Washington Post,* 24 July 2020, available at washingtonpost.com/politics/2020/07/24/russias-disinformation-campaigns-are-targeting-african-americans/ (retrieved on 15 February 2022).

[425] Joint Communication JOIN(2020) 8 final; Judd, Vazquez & O'Sullivan, 2021, 'Biden says platforms like Facebook are 'killing people' with Covid misinformation'.

disinformation. While this regulation was not backed by legislation or any legal liability, it may have influenced what providers allow on their service.[426]

These effects, however, are partly caused by the possibility of enacting legislation when non-legislative regulation does not have the desired effect. According to Citron, service providers adopted terms and conditions on hate speech and terrorist content after the EC asked them to.[427] According to Citron, service providers "accommodated these demands because regulation of extremist speech was a real possibility."[428] Adapting new regulations addressing online terrorist content shows Citron and the providers were not wrong.[429]

As shown in the previous paragraphs, content moderation policies enacted by providers thus can be either 1) completely voluntary, 2) requested by another (either private or state) actor but voluntary enacted, 3) or legally required by state actors backed by fines or other state sanctions.

2.2 Instruments: overregulation and underregulation by moderation and curation

Content regulation can be directed at the service user or the provider that offers the service. As noted, providers can enact moderation policies for many reasons, including their own. Providers may also influence what is shown to (individual) users without removing user-provided information from the service altogether.

There is thus a difference between moderation (remedy a rule violation) and curation (indexing, organising, and recommending) of user-provided information. In both cases, a provider makes decisions concerning the visibility of the content of user-provided information. Moderation leads to a remedy following a rule violation which usually results in the inaccessibility of user-provided information. In the case of curation, the provider seeks to offer relevant user-provided information to the user, resulting in higher or lower visibility of specific information based on its content. When a service provider curates, other facts and circumstances than the content of the user-provided information may be considered. For example, curation may also occur based on previous interactions with other user-provided information. The user of the internet intermediary service could be offered information similar to the content of earlier clicked information. Curation for individual users is often referred to as personalisation. Content curation, however, can also apply to all users of a service. During the COVID-19 pandemic, providers promoted authoritative information from governments and health officials while ranking user-provided information with (possible) misinformation or disinformation lower.[430]

---

[426] In the case of COVID-19 disinformation, see Twitter, 'COVID-19 misleading information policy'; Facebook, 'COVID-19 policy updates and protections', *Facebook Help Center*, available at facebook.com/help/230764881494641 (retrieved on 14 February 2022); Google, 'COVID-19 medical misinformation policy', *YouTube Help,* 20 May 2020, available at support.google.com/youtube/answer/9891785 (retrieved on 15 February 2022).

[427] Citron, 'Extremist Speech, Compelled Conformity, and Censorship Creep', *Notre Dame Law Review*, 2018, pp. 1037-1038.

[428] Citron, 'Extremist Speech, Compelled Conformity, and Censorship Creep', *Notre Dame Law Review*, 2018, p. 1038.

[429] Regulation (EU) 2021/784.

[430] Communication COM(2021) 262 final of the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions of 26 May 2021 European Commission Guidance on Strengthening the Code of Practice on Disinformation, p. 14.

As argued, vague legal definitions may lead to the over-removal of user-provided information with severe consequences for users' freedom of expression rights.[431] The question is whether the curation of user-provided information may reach similar concerns. Therefore, in this chapter, the moderation and curation efforts based on the content of user-provided information are discussed.

2.2.1 Moderation

Often moderation is reviewed in discussing overregulation or underregulation by providers. Moderation encompasses providers' interventions on user-provided information and/or on the user accounts because of an (alleged) violation of a rule. Moderation results typically in interventions that encompass remedies affecting the availability of user information or the possibility for a user to access the service.[432] Moderation, however, can involve other remedies that are discussed in paragraph 2.3. Interventions on the content of user-provided information that does not involve a remedy following a rule violation fall outside the scope of this concept of moderation. For example, providers that affect the visibility of user-provided information to other users based on personalisation are not moderation but curation, which is discussed in paragraph 2.2.2.

The concept of content moderation, like the concept of internet intermediary, is not defined in early legislation that deals with provider liability. Recognising that content moderation by providers may impact users' freedom of expression rights,[433] the EC seeks to change this with the DSA. In the proposal for the DSA, the following definition is proposed:

> 'content moderation' means the activities undertaken by providers of intermediary services aimed at detecting, identifying and addressing illegal content or information incompatible with their terms and conditions, provided by recipients of the service, including measures taken that affect the availability, visibility and accessibility of that illegal content or that information, such as demotion, disabling of access to, or removal thereof, or the recipients' ability to provide that information, such as the termination or suspension of a recipient's account;[434]

While content moderation, following this definition, seems a clear-cut concept, the opposite is true. Content moderation, as a concept, is highly contested. To compare, the Steering Committee on Media and Information Society (hereafter: CDMSI) of the Council of Europe defines content moderation in a Guidance Note as:

> The process whereby a company hosting online content assesses the [il]legality or compatibility with terms of service of third-party content, in order to decide whether certain content posted, or attempted to be posted, online should be demoted […] tagged as being potentially inappropriate or incorrect, demonetised, not sanctioned or removed, for some or all audiences, by the service on which it was posted.[435]

---

[431] M. Masnick, 'Protocols, Not Platforms: A Technological Approach to Free Speech', *Knight Columbia,* 21 August 2019, available at knightcolumbia.org/content/protocols-not-platforms-a-technological-approach-to-free-speech (retrieved on 15 February 2022), p. 12.

[432] Goldman, 'Content Moderation Remedies', *Michigan Technology Law Review*, 2021, pp. 5-6.

[433] Commission Proposal COM(2020) 825 final (*Digital Services Act*), p. 2.

[434] Article 2(p) of Commission Proposal COM(2020) 825 final (*Digital Services Act*), p. 45.

[435] Council of Europe, 2021, 'Content moderation: best practices towards effective legal and procedural frameworks for self-regulatory and co-regulatory mechanisms of content moderation', p. 11.

Both definitions clarify that content moderation sees to 1) detecting and identifying user-provided information that contains content that may violate the rules and then 2) assessing whether this instance of content indeed violates the rules. Rule violation, in both definitions, sees to information with illegal or unlawful content and content that violates the terms and conditions set by the providers. Both definitions, thus, do not limit content moderation to either public or private rules. Besides, both definitions provide examples of sanctions and remedies that could follow a rule violation. The rules may be private (exclusively laid down in the terms and conditions) or public (laid down in legislation but often translated in the terms and conditions of the provider).

Regarding the remedies that may follow a rule violation, the DSA offer a more comprehensive definition. The DSA views all remedies that "affect the availability, visibility and accessibility" of user-provided information next to and the ability of the user to provide new information to the intermediary service as potential remedies. The CDMSI only considers action undertaken against a specific instance of information as a remedy following moderation. In the case of moderation, the detection of potential rule violating content, the interpretation and enforcement of the rules followed by an appropriate remedy are all equally important. This paragraph focuses on the first two stages (detection and assessment), while paragraph 2.3 discusses how an appropriate remedy should address the rule violation.

Moderation efforts are increasingly put under scrutiny by academics, civil society organisations, and governmental actors.[436] Especially governments may force providers to change how they moderate. While there are legitimate interests in reviewing moderation efforts by providers of intermediary services, state actors must be cautious in imposing regulation on moderation because they are unhappy with how providers perform moderation tasks. Moderation, after all, is not an easy task. As Gillespie argues:

> Moderation is hard because it is resource intensive and relentless; because it requires making difficult and often untenable distinctions; because it is wholly unclear what the standards should be; and because one failure can incur enough public outrage to overshadow a million quiet successes.[437]

As noted, providers must moderate user-provided information because they are legally required to do so. They, however, may also moderate for various other reasons – including reasons of their own. Moderation, whether state-sanctioned or out of the initiative of the intermediary itself, may lead to conflicts between users and providers. The provider may argue that it is legally required or at least legally justified to moderate, while the users may believe that the provider limits their freedom of expression rights. Users could accuse providers of moderating user-provided information whose content does not violate state legislation and even may be considered protected

[436] For example, various initiatives have attempted to subject content moderation to certain standards, see Manila Principles on Intermediary Liability, 'Manila Principles on Intermediary Liability: Best Practices Guidelines for Limiting Intermediary Liability for Content to Promote Freedom of Expression and Innovation', *Manila Principles on Intermediary Liability,* 24 March 2015, available at eff.org/files/2015/10/31/manila_principles_1.0.pdf (retrieved on 15 February 2022); The Santa Clara Principles, 'Santa Clara Principles 1.0', *The Santa Clara Principles on Transparency and Accountability in Content Moderation,* 7 May 2018, available at santaclaraprinciples.org/scp1/ (retrieved on 15 February 2022). Besides, the European Commission seeks to influence moderation practices with a proposal for the Digital Services Act, see Commission Proposal COM(2020) 825 final (*Digital Services Act*).

[437] Gillespie, 2018, *Custodians of the Internet*, p. 9.

speech under (international, regional, or constitutional) freedom of expression rights.[438] In European jurisdictions, this may result in the user suing the provider. A provider arguing that the directions of the state were followed could be exonerated from blame in such procedures.[439]

Such an outcome is, however, not a given.[440] Unlike the US legislation, the EU e-Commerce Directive does not know an exemption for the legal liability of providers when they moderate user-provided information they genuinely believe to violate state legislation.[441] As Van Eecke observes, the Directive even emphasises[442] that hosting services in moderating user information must take into account the freedom of expression rights of the user.[443] Moderation is hard for service providers when providers are required to take down illegal and unlawful content because they could mistakenly pass illegal content as legal. Content moderation becomes almost impossible when legislation sets boundaries on what providers can moderate at their initiative.[444] Expecting providers to be exactly right in terms of content moderation is expecting providers to wield supernatural powers.

Providers operate in legal limbo. There is little to no certainty on the boundaries of moderating user-provided information. Users of internet intermediary services have good reasons to complain over a lack of legal protections against wrongful removal of information they provided to their service or termination of user accounts. For users, it is hard to win a case against a provider that wrongfully moderates – if it is possible to sue in the first place. In the EU, there are no clear legal limitations on what providers can and cannot do when it comes to moderation – it depends on the facts and circumstances in each case. In the US, providers are offered broad discretion in moderating the content of user-provided information: both for moderating and not moderating.[445]

Overregulation caused by legal liability regimes may be foreseen or unforeseen and intentional or accidental.[446] While providers are not open about how they carry out content regulation, some empirical evidence exists that over-removal occurs on a large scale.[447] Overregulation caused by how states impose legal liability on providers is troublesome because of the state-intermediary dynamic. Balkin warns that states may (ab)use the providers' capabilities of

---

438 Klos, "Wrongful moderation': Aansprakelijkheid van internetplatforms voor het beperken van de vrijheid van meningsuiting van gebruikers', *Nederlands Juristenblad*, 2020/2976.

439 Rb. Amsterdam (vzr.), 9 September 2020, ECLI:NL:RBAMS:2020:4435, Rec. 4.11 (*YouTube*).

440 C. Goujard, 'German Facebook ruling boosts EU push for stricter content moderation', *Politico,* 29 July 2021, available at politico.eu/article/german-court-tells-facebook-to-reinstate-removed-posts (retrieved on 15 February 2022); Rb. Noord-Holland (vzr.), 6 October 2021, ECLI:NL:RBNHO:2021:8539, Rec. 4.24 (*Kamerlid/LinkedIn*).

441 See 47 USCA § 230(c)(2) (West 2018, Westlaw Next through PL 116-91); 17 USCA § 512(g)(1) (West 2010, Westlaw Next through PL 116-179).

442 Recital 46 of Directive 2000/31/EC (*Directive on electronic commerce*).

443 Van Eecke, 'Online service providers and liability: A plea for a balanced approach', *Common Market Law Review*, 2011, p. 1468.

444 Klos, "Wrongful moderation': Aansprakelijkheid van internetplatforms voor het beperken van de vrijheid van meningsuiting van gebruikers', *Nederlands Juristenblad*, 2020/2976.

445 47 USCA § 230(c) (West 2018, Westlaw Next through PL 116-91).

446 For example, the risks that books that are expected to contain hate speech are 'deshelved' out of fear of criminal liability, see Paragraph 13 of B.P. Vermeulen, 'Artikel 7 - Vrijheid van meningsuiting', *NederlandRechtsstaat*, available at nederlandrechtsstaat.nl/grondwet/inleiding-bij-hoofdstuk-1-grondrechten/artikel-7-grondwet-vrijheid-van-meningsuiting (retrieved on 15 February 2022).

447 Keller, 2020, 'Empirical Evidence of "Over-Removal" by Internet Companies Under Intermediary Liability Laws'.

internet intermediaries to carry out state regulation.[448] Making providers responsible for enforcing state law requires providers to interpret the law and decide whether the content of user-provided information violates (their interpretation) of the law.[449] Because providers may become liable for failing to (correctly) apply state regulation, they may decide to also remove user-provided information with content that may violate the law without being sure.[450] The CDMSI argues that state regulation should offer predictability regarding liability to remedy such harmful effects.[451] The CDMSI even notes that making internet intermediaries liable for illegal or unlawful content of user-provided information "may not be the most effective, proportionate and targeted way towards achieving a balanced outcome."[452]

Because of these risks, NGOs and academics cooperated in drafting *The Manila Principles on Intermediary Liability* (2015), setting out seven principles for an intermediary liability framework. These seven principles sought to prevent the over-removal of user-provided information. The first principle, which deals with the liability of providers for user-provided information, is the most important for this paragraph. The first principle rejects strict liability: providers should not be held liable for user-provided information by merely offering a service. To clarify the boundaries of the liability regime, legislation dealing with internet intermediary liability should be "precise, clear, and accessible". The first principle of *The Manila Principles* sets out that providers should be immunised from liability for user-provided information. The only exception is that providers should not be immunised when they modify the content of user-provided information. Providers should not be burdened with monitoring user-provided information for illegal content.[453] As discussed in chapters 3 and 4, this first principle (partly) comes back in the liability regimes in the EU and the US. The most crucial difference with the EU regime is that the e-Commerce Directive does not offer complete immunity for liability for user-provided information to providers but makes liability dependent on knowledge or awareness of illegal or unlawful content.[454]

The relationship between user and provider in the EU is governed by contract law without a legal shield similar to Section 230. The absence of such a provision enables users to bring complaints about removing user-provided information or account termination before a judge. However, users are likely to lose the case because of the terms of services of the internet intermediary service provider.[455] Judges setting aside this contract to safeguard user freedom of expression rights seem to form an exception.[456] At the same time, interventions by a provider may

---

[448] Balkin, 'Free Speech is a Triangle', *Columbia Law Review*, 2018, p. 2029.

[449] Land, 'Against Privatized Censorship: Proposals for Responsible Delegation', *Virginia Journal of International Law*, 2020, p. 408.

[450] Keller, 2019, 'Who Do You Sue? State and Platform Hybrid Power over Online Speech', p. 3.

[451] Council of Europe, 2021, 'Content moderation: best practices towards effective legal and procedural frameworks for self-regulatory and co-regulatory mechanisms of content moderation', p. 20.

[452] Council of Europe, 2021, 'Content moderation: best practices towards effective legal and procedural frameworks for self-regulatory and co-regulatory mechanisms of content moderation', p. 28.

[453] Principle 1 of Manila Principles on Intermediary Liability, 2015, 'Manila Principles on Intermediary Liability: Best Practices Guidelines for Limiting Intermediary Liability for Content to Promote Freedom of Expression and Innovation'.

[454] Article 14(1) of Directive 2000/31/EC (*Directive on electronic commerce*).

[455] Rb. Midden-Nederland (vzr.), 8 October 2020, ECLI:NL:RBMNE:2020:4348, *Computerrecht* 2021/65, m.nt. M.G.A. Berk.

[456] For example, when the terms of service or its application is not sufficiently clear, see Rb. Noord-Holland (vzr.), 6 October 2021, ECLI:NL:RBNHO:2021:8539, Rec. 4.20-4.24 (*Kamerlid/LinkedIn*).

have severe consequences for the user in question. Service providers have a clear legal interest to moderate because of the (potential) liability from illegal or unlawful content. Besides, users that are repeatedly violating the rules may be harmful to the business interests of the provider or other users of the service.

While there may be legitimate interests in engaging in moderation, this does not mean that providers should be granted unlimited discretion to decide on the rules on a case-to-case basis. Various civil society initiatives seek to bind providers to principles designed to safeguard user rights. An example of such an initiative is *The Santa Clara Principles on Transparency and Accountability in Content Moderation* (2018) which articulate norms for providers. *The Santa Clara Principles* require providers to publish how many content removals and interventions on accounts they undertook. These numbers include how many removals and suspensions (for example, following 'flagging') the provider has imposed for different formats (for example, text or video) of user-provided information. In addition, the provider has to report what type of rule violations it encounters and how the provider was notified of the violation. For example, the provider may receive notifications from governmental actors. The reports must also reflect where the notification came from and which groups of users were impacted (for example, by hiding posts based on the geographical location). Next to a breakdown in numbers, the provider should notify users of the rule violation. This notification requirement holds that providers point out what user-provided information is affected, the specific rule violated, and how the provider became aware of the rule violation. Besides, the provider must set out how the user can appeal the decision in the notification. The requirements for appeal are laid down in the third principle of *The Santa Clara Principles,* in which minimum requirements for providers are set out, which includes due process requirements such as independent review by a human, the possibility for users to submit supplementary information taken into account by the human reviewer in the appeal process, and a reasoned decision by the provider after review.[457]

The Santa Clara Principles set out principles on transparency and accountability of providers that engage in content moderation.[458] Of course, providers do not operate in a (legal) vacuum but are restricted by the legal landscape in which they operate. Therefore, it is necessary to complement *the Santa Clara Principles* with the already mentioned *Manila Principles*. As already noted, the *Manila Principles* are primarily aimed at the state. The state must restrict the liability of providers for the content of user-provided information to prevent over-removal.[459] The *Manila Principles* also include principles directed at providers. For example, the fifth principle sets out that providers should offer users "mechanisms to review decisions to restrict content in violation of the intermediary's content restriction policies" and "should reinstate the content" when no rule violation is found after review.[460] Besides, *The Manila Principles* articulate that providers should adhere to human rights

[457] The Santa Clara Principles, 2018, 'Santa Clara Principles 1.0'.

[458] The Santa Clara Principles, 2018, 'Santa Clara Principles 1.0'.

[459] The Manila Principles, however, have some overlap with the Santa Clara Principles with respect to transparency and notification requirements, see Principle VI(c) and, to some extent, (g) of Manila Principles on Intermediary Liability, 2015, 'Manila Principles on Intermediary Liability: Best Practices Guidelines for Limiting Intermediary Liability for Content to Promote Freedom of Expression and Innovation', p. 5.

[460] Principle V(c) and (d) of Manila Principles on Intermediary Liability, 2015, 'Manila Principles on Intermediary Liability: Best Practices Guidelines for Limiting Intermediary Liability for Content to Promote Freedom of Expression and Innovation', p. 4.

requirements in setting out their community guidelines and enforcing these policies.[461] These policies must be in "clear language and accessible formats" online available. These policies must be kept up to date. In case of an update, users must be notified of changes.[462] In the case a provider restricts access to or removes information, the provider must place "a clear notice that explains what content has been restricted and the reason for doing so."[463]

Imposing regulation that increases the liability of providers without safeguards does not help users but may lead to new restrictions. Service providers may moderate more extensively out of fear of liability for user-provided information and – in the most extreme circumstances – may even change how or what services they offer.[464] Therefore, *The Manilla Principles* prohibit "extra-judicial measures to restrict content" such as "collateral pressures to force changes in terms of service, to promote or enforce so-called 'voluntary' practices and to secure agreements in restraint of trade or restraint of public dissemination of content."[465] When a government wishes to impose restrictions on what users of providers can and cannot provide to their services, they have to enact legislation. Legislation, however, is not enough. *The Manila Principles* add that providers should only engage in government-sanctioned moderation after "an order has been issued by an independent and impartial judicial authority that has determined that the material at issue is unlawful."[466] According to the Manila Principles, delegating moderation of user-provided information to providers by declaring content illegal in legislation is not an option.

The Manila Principles were drawn in 2015, and the *Santa Clara Principles* in 2018.[467] A few years after these principles were drafted, the accountability of providers of internet intermediary services is sharp on the minds of scholars and policymakers. However, proposals for new regulations do not necessarily reflect the principles laid down in *The Manila* and *Santa Clara Principles*. Some of the principles find their way into proposals for legislation. For example, requiring providers to lay down precise rules in their terms of services ultimately overseen by out-of-court dispute settlement[468] reflects these principles. Besides, there are a lot of new transparency requirements proposed.[469] Especially online platforms that offer social networking functionalities

---

461 Principle V(f) of Manila Principles on Intermediary Liability, 2015, 'Manila Principles on Intermediary Liability: Best Practices Guidelines for Limiting Intermediary Liability for Content to Promote Freedom of Expression and Innovation', p. 4.

462 Principle VI(c) of Manila Principles on Intermediary Liability, 2015, 'Manila Principles on Intermediary Liability: Best Practices Guidelines for Limiting Intermediary Liability for Content to Promote Freedom of Expression and Innovation', p. 5.

463 Principle VI(f) of Manila Principles on Intermediary Liability, 2015, 'Manila Principles on Intermediary Liability: Best Practices Guidelines for Limiting Intermediary Liability for Content to Promote Freedom of Expression and Innovation', p. 5.

464 Goldman, 'The Complicated Story of FOSTA and Section 230', *First Amendment Law Review*, 2019, pp. 288-289.

465 Principle VI(b) of Manila Principles on Intermediary Liability, 2015, 'Manila Principles on Intermediary Liability: Best Practices Guidelines for Limiting Intermediary Liability for Content to Promote Freedom of Expression and Innovation', p. 5.

466 Principle II(a) of Manila Principles on Intermediary Liability, 2015, 'Manila Principles on Intermediary Liability: Best Practices Guidelines for Limiting Intermediary Liability for Content to Promote Freedom of Expression and Innovation', p. 2.

467 Manila Principles on Intermediary Liability, 2015, 'Manila Principles on Intermediary Liability: Best Practices Guidelines for Limiting Intermediary Liability for Content to Promote Freedom of Expression and Innovation', p. 1. The Santa Clara Principles, 2018, 'Santa Clara Principles 1.0'.

468 Article 18 of Commission Proposal COM(2020) 825 final (*Digital Services Act*), pp. 53-55.

469 Article 23 of Commission Proposal COM(2020) 825 final (*Digital Services Act*), p. 58.

with a large userbase or with a minimum annual global turnover are targeted by new legislation.[470] Most far-reaching, however, are proposals that codify that not the provider of the intermediary service but the state or the user community should set the standards of moderation.[471]

In regulating providers, intermediary accountability and transparency get much attention. In contrast, government transparency and accountability seemed moved to the background, while some government instruments regulating providers are highly questionable in light of the *Manila Principles*. Not only are internet intermediary services confronted with regulation targeting illegal or unlawful content, but also with regulation that targets harmful but not necessarily illegal user-provided information. User-provided information that may qualify as harmful may even be protected under (international) freedom of expression rights.[472] For example, the DSA empowers the EC to conclude codes of conduct that are made part of a co-regulatory regime, meaning that upholding the code of conduct is effectively part of the audits of "very large online platforms".[473] Oversight over the behaviour of the EC with respect to these codes of conduct is less codified, while this behaviour can easily lead to government coercion.[474]

Even if the DSA is adopted, how this regulation works out should be subjected to constant review. As understood by *The Manila Principles*, content moderation is a collective effort that is never finished. How internet intermediary liability regimes work out should therefore be critically followed.[475] Therefore, the *Manilla Principles* recommend that governments, civil society, and the provider of internet intermediary services should collaborate in "independent, transparent, and impartial oversight mechanisms to ensure the accountability of the content restriction policies and practices."[476] Accountability of both providers and the government regarding content regulation is necessary to safeguard users' freedom of expression rights.

As noted, the attention shifted from holding governments accountable and increasing government transparency to provider accountability and transparency. This shift in attention may be risky. The impact of such governmental regulation through providers may be hidden because content moderation is attributed to the service provider. However, this shift in attention can be easily explained by the fact that providers are placed not under the auspices of the state but besides the state. Service providers are framed as state-like actors regarding their capabilities, possibilities, and financial and political power.[477] At the same time, states increasingly rely on providers for their

---

[470] See, for example, Article 16 and 25 of Commission Proposal COM(2020) 825 final (*Digital Services Act*).

[471] See, for example, the recommendation of the Dutch government councils Adviesraad Internationale Vraagstukken, 2020, 'Regulering van online content: Naar een herijking van het Nederlandse internetbeleid (AIV-advies 113)', pp. 11-13; Van Huijstee, et al., 2021, 'Online ontspoord: Een verkenning van schadelijk en immoreel gedrag op het internet in Nederland', pp. 139-141.

[472] For example, Article 17 and 18 of Commission Proposal COM(2020) 825 final (*Digital Services Act*).

[473] Recital 67-70 of Commission Proposal COM(2020) 825 final (*Digital Services Act*), pp. 34-35; Article 27 and 28 of Commission Proposal COM(2020) 825 final (*Digital Services Act*), pp. 60-61.

[474] Citron, 'Extremist Speech, Compelled Conformity, and Censorship Creep', *Notre Dame Law Review*, 2018, p. 1070.

[475] Principle VI(h) of Manila Principles on Intermediary Liability, 2015, 'Manila Principles on Intermediary Liability: Best Practices Guidelines for Limiting Intermediary Liability for Content to Promote Freedom of Expression and Innovation', p. 5.

[476] Principle VI(g) of Manila Principles on Intermediary Liability, 2015, 'Manila Principles on Intermediary Liability: Best Practices Guidelines for Limiting Intermediary Liability for Content to Promote Freedom of Expression and Innovation', p. 5.

[477] For example, in terms of lobbying power with respect to standard-setting, see Mak, 2020, *Legal Pluralism in European Contract Law*, pp. 209-210 and 221-222. But also in terms of distorting competition and impact on the

technological and bureaucratic possibilities to moderate user-provided information where the state cannot.[478] This dependency on providers makes it rather strange to frame these providers as state-like actors that must be brought back under governmental control.[479]

The US and the European context, however, are very different. In the US, regulating providers requiring them to moderate user-provided information may conflict with the 'free speech clause' of the First Amendment. In contrast, under the ECHR, states may even have a positive obligation to regulate providers.[480] While the First Amendment severely restricts state involvement in what content is allowed in the US, the ECHR (as interpreted by the ECtHR) may include a positive state obligation to require providers to have transparent and predictable rules for what user-provided information is allowed.[481]

Government actors that seek to influence what providers are required to moderate and what they cannot moderate must relate to these freedom of expression safeguards. Elsewhere I argued that it would be unwise for state legislators and the judiciary to severely limit the possibility for providers of internet intermediary services to enact content moderation policies of their own.[482] As already noted, moderation also deals with what remedy can, must, and should be imposed after a rule violation. These rules may have two sources. The rules can be a direct consequence of state regulation (both legislative and non-legislative) and the result of providers imposing rules on their own. Proposals to regulate moderation by providers seek to restrict the latter while expanding the first.[483] Providers are not subjected to the same human rights obligations and legal restrictions as state actors. Providers have more room to regulate the information provided by their users than the state. Of course, this discretionary room to set and enforce standards can potentially be abused while leaving the user of the services empty-handed.[484] Providers may, in the worst case, set standards that align with their viewpoints while prohibiting information with content that opposes this view. When the dependency of users on internet intermediary services for their media

---

public by excluding others from their infrastructure, see Wu, 2011, *The Master Switch*, pp. 57-59; G. Lakier, 'The Non-First Amendment Law of Freedom of Speech', *Harvard Law Review*, Vol. 134, No. 7, 2021 (available at harvardlawreview.org/2021/05/the-non-first-amendment-law-of-freedom-of-speech), pp. 2319-2320.

[478] Balkin, 'Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation', *U.C. Davis Law Review*, 2018, p. 1175; H. Bloch-Wehba, 'Global Platform Governance: Private Power in the Shadow of the State', *SMU law review*, Vol. 72, No. 1, 2019, p. 39.

[479] For example, Adviesraad Internationale Vraagstukken, 2020, 'Regulering van online content: Naar een herijking van het Nederlandse internetbeleid (AIV-advies 113)', p. 41.

[480] As Judges Raimondi, Karakaş, De Gaetano and Kjølbro argued a joint concurring opinion, this may come down to balancing the right to respect for private and family life (Article 8) and freedom of expression rights (Article 10) in regulating providers, see their separate opinion of *Delfi AS v. Estonia* [GC], no. 64569/09, § 10, ECHR 2015-II, 16 June 2015.

[481] Compare Keller, 2021, 'Six Constitutional Hurdles for Platform Speech Regulation'; Council of Europe, 2021, 'Content moderation: best practices towards effective legal and procedural frameworks for self-regulatory and co-regulatory mechanisms of content moderation', p. 19 and 24.

[482] Klos, "Wrongful moderation': Aansprakelijkheid van internetplatforms voor het beperken van de vrijheid van meningsuiting van gebruikers', *Nederlands Juristenblad*, 2020/2976.

[483] For example, before the DSA was proposed the EC concluded two 'voluntary' codes and proposed legislation that sees to terrorist content, see European Commission, 2016, 'Code of Conduct on Countering Illegal Hate Speech Online'; European Commission, 2021, 'Code of Practice on Disinformation'; Regulation (EU) 2021/784.

[484] Yemini, 'The New Irony of Free Speech', *Columbia Science and Technology Law Review*, 2018, p. 193.

consumption increases,[485] such standard-setting may significantly distort the possibility for users to express and receive viewpoints as they wish. In such cases, boundary setting for providers may be deemed required by imposing limitations on what providers can and cannot moderate.[486]

Such a requirement may be unwise and counterproductive for three reasons. The first reason is that providers are simply not able to engage in perfect state-sanctioned moderation. Service providers will almost certainly miss illegal or unlawful content that they should moderate while passing content as legal. Sometimes providers will moderate content that is not illegal. Any requirement for providers to only moderate illegal content presupposes perfect moderation that providers simply cannot uphold.[487] Moderation based on legislative standards is especially hard. While a court quickly takes a few months to a few years before rendering a decision about whether the content of speech violates the law, internet intermediary providers are expected to decide in an hour to a few days whether the rules are violated. Legislation dealing with freedom of expression rights often requires a careful contextual assessment, raising multiple interpretation issues. In other words, it is hard to decide whether an expression is indeed defamatory. However, even when it is easy to establish its defamatory character, numerous factors are taken into account to establish its unlawful character. When does the personal interest or the public interest exonerate the speaker from liability? When is an expression offensive to a group, and what groups are protected? These are challenging questions that are not easy to answer for providers. Ambiguous legislation could easily lead to overregulation.[488] For the provider and the users of the intermediary service, it may be preferable to set clear, (perhaps) broader standards that are easily understandable for the enormous userbase of the intermediary services.[489]

A third reason it would be unwise to restrict content moderation by service providers is that it may be desirable that providers moderate content that is not prohibited by legislation. For example, in the US, the First Amendment, as interpreted by the Supreme Court of the United States (hereafter: SCOTUS), limits content-based restrictions by the legislator.[490] Such a restriction does not bind providers. In the European context, legislative restrictions on sharing content should be considered an *ultimum remedium*.[491] When the state has a legitimate interest in enacting content-

---

485 For example, A.W. Geiger, 'Key findings about the online news landscape in America', *Pew Research Center,* 11 September 2019, available at pewresearch.org/fact-tank/2019/09/11/key-findings-about-the-online-news-landscape-in-america (retrieved on 14 February 2022).

486 New or proposed legislation often encompass such limitations, see 2021 Fla. Sess. Law Serv. Ch. 2021-32 (SB 7072) (West); 'Draft Online Safety Bill', *Department for Digital, Culture, Media & Sport,* 12 May 2021, available at gov.uk/government/publications/draft-online-safety-bill (retrieved on 15 February 2022); Commission Proposal COM(2020) 825 final (*Digital Services Act*).

487 Gillespie, 2018, *Custodians of the Internet*, p. 9.

488 Citron, 'Extremist Speech, Compelled Conformity, and Censorship Creep', *Notre Dame Law Review*, 2018, pp. 1052-1055.

489 See, for example, the examples in Facebook's 'Hate Speech' policy, Meta, 2021, 'Hate Speech'.

490 A. Guiora & E. Park, 'Hate Speech on Social Media', *Philosophia*, Vol. 45, No. 3, 2017, doi:10.1007/s11406-017-9858-4, pp. 964-965.

491 See in case of disinformation and political expressions, for example, Van Hoboken, et al., 2019, 'Het juridisch kader voor de verspreiding van desinformatie via internetdiensten en de regulering van politieke advertenties', p. 128.

based restrictions, it usually takes a while before legislation is passed. Providers may pioneeringly enact policies before state regulation makes it to the law books.[492]

Should this mean that providers should get a blank check concerning content moderation? While it is necessary to prevent providers from arbitrarily restricting content because providers may skew the public debate towards their ends, this does not mean that providers should be limited to moderating strictly illegal content. Instead of limiting what providers can include in their moderation policies, it may be wiser to oversee how providers apply their policies. To prevent providers from skewing the public debate, they could be required to enact policies that can be enforced in an indiscriminate matter. For example, a provider can enact a policy prohibiting promoting medical products, which should not be enforced arbitrarily. Therefore, some public oversight of moderation practices is desirable and necessary.[493]

2.2.2 Curation and customisation

Providers may also influence what user-provided information is offered to other users by curating and offering customisation tools. As I understand it, curation and customisation differ from each other. For example, curation is carried out by the provider without any direct influence of the user that consumes the curated information. Customisation means that a provider offers tools to users to customise for themselves how and what information is shown. I first discuss how I view curation, and then I turn to customisation as an alternative for curation.

Curation encompasses all interventions of providers on what information is shown to whom, when, where, and how. Curation may take the shape of personalisation. In the case of personalisation, the provider curates the information provided to one specific user based on the characteristics of the user in question. Curation, however, does not always take the form of personalisation. Providers may also curate user-provided information for all users, for example, by leaving out (potential) harmful (but lawful) user-provided content out of the search results or the suggestions that are shown when a search term is entered. In its Guidance Note, the CDMSI defines content curation as:

> The process of deciding which content should be presented to users (in terms of frequency, order, priority, and so on), based on the business model and design of the platform.[494]

Curation, thus, encompasses interventions on how user-provided information is presented. Curation differs from moderation in two ways: curation does not (necessarily) occur after a rule violation is established, nor does curation deal with removing user-provided information or other restrictions on the availability of its content. Curation, however, may affect actual availability and thus the reach of user-provided information. For example, information may be shown less or placed in a position that is hard to find. In other words, curation does not see to the availability of the content user-provided information in a strict sense. In contrast, the actual availability in terms of visibility may be affected positively or negatively. Curation by providers thus may contribute to

---

[492] Van Huijstee, et al., 2021, 'Online ontspoord: Een verkenning van schadelijk en immoreel gedrag op het internet in Nederland', pp. 40-41.

[493] Klos, "Wrongful moderation': Aansprakelijkheid van internetplatforms voor het beperken van de vrijheid van meningsuiting van gebruikers', *Nederlands Juristenblad*, 2020/2976, pp. 3321-3322.

[494] Council of Europe, 2021, 'Content moderation: best practices towards effective legal and procedural frameworks for self-regulatory and co-regulatory mechanisms of content moderation', p. 11.

the virality of user-provided information and the fact that some information may be near impossible to find.

As the CDMSI definition shows, curation is based on the "business model" or the "design of the platform",[495] which suggests that the interests of providers may put a fair amount of weight in the balance.[496] The direct involvement of providers in curating user-provided content caused Keller, Fukuyama and Masnick to diagnose the bundling of the internet intermediary functions as a risk for users' freedom of expression rights. Keller argues that providers give "a common point of control".[497] According to Masnick, such a point of control offers centralised control over user-provided information. This control is grouped in the hands of a few companies.[498] According to Fukuyama et al., these companies gained an "economic, social, and political influence"[499] that is unprecedented. Providers may not always serve the user's interest in curating user-provided information.[500]

Because of curation's (possible) intrusive character, proposals are made to decouple curation from other intermediary functions. One of the possible alternatives is discussed by Keller: the so-called 'Magic API'. An application programming interface (API) is a computer code that allows different computer programs to communicate. For example, Twitter allows developers to their API to view, analyse, and interact with user-provided content (called Tweets) on their service, allowing developers to build their software around Twitter.[501] APIs, however, are limited to what the provider of the API allows. Besides, there may be limitations on the API usage or functionalities that require premium or enterprise licenses for which the provider may charge extra. Keller explores the 'Magic API' as an alternative for platform-centric curation. The provider would provide the user-provided information through the API before curation. This API allows others to develop curation services for the intermediary service. Users of internet intermediaries can decide themselves what content curation service they choose.[502] In other words, users are not dependent on the curation service offered by the provider – they can use other curation providers as well.

The "Magic API" can be viewed as a less far-reaching alternative to Masnick's proposal to open the protocols of platforms.[503] As Masnick notes, an online platform is a bundle of different protocols that add to platform functionalities concentrated on private services. Allowing others to

---

[495] Council of Europe, 2021, 'Content moderation: best practices towards effective legal and procedural frameworks for self-regulatory and co-regulatory mechanisms of content moderation', p. 11.

[496] Providers that place their own goals in the place of user goals may be problematic from the user perspective, see J. Grimmelmann, 'Speech Engines', *Minnesota Law Review*, Vol. 98, No. 3, 2014 (available at scholarship.law.umn.edu/mlr/299), p. 874.

[497] Keller, 2019, 'Who Do You Sue? State and Platform Hybrid Power over Online Speech', p. 27.

[498] Masnick, 2019, 'Protocols, Not Platforms: A Technological Approach to Free Speech', p. 6.

[499] F. Fukuyama, et al., 'Middleware for Dominant Digital Platforms: A Technological Solution to a Threat to Democracy', *Stanford Cyber Policy Center,* 2021, available at cyber.fsi.stanford.edu/content/biden-recommendations-cyber-policy-center (retrieved on 14 February 2022), p. 1.

[500] Balkin, 'Free Speech is a Triangle', *Columbia Law Review*, 2018, pp. 2040-2041. In the DSA, the EC proposes due process requirements for providers that qualify as 'online platform', see Recital 34 and 35 of Commission Proposal COM(2020) 825 final (*Digital Services Act*), p. 25.

[501] Twitter, 'Twitter API', *Twitter Developer Platform*, available at developer.twitter.com/en/docs/twitter-api (retrieved on 15 February 2022).

[502] Keller, 2019, 'Who Do You Sue? State and Platform Hybrid Power over Online Speech', pp. 26-27.

[503] Keller, 2019, 'Who Do You Sue? State and Platform Hybrid Power over Online Speech', p. 26.

use these protocols by opening these platforms up would remedy the situation that internet intermediary functionalities are concentrated in the hands of a few providers. Removing barriers to the usage of these protocols would allow others to develop, for example, content filters, curation services, or interfaces built upon the information offered to these providers. The protocols-not-platforms approach allows for a dichotomy between concentration and complete decentralisation. It is possible to open up parts of the platform by allowing access to a restricted number of protocols by offering a Magic API.[504] According to Masnick, opening up the protocols for using others to develop services around user-provided content could remedy (alleged) bias of providers and harmful effects of market concentration. Besides, opening up the platforms would form a way to answer calls for social responsibility regarding content moderation. The provider would be no longer exclusively responsible for content moderation because others could take up the glove and develop filters and curating services.[505]

Platformisation, as Masnick notes, has given providers exclusive control over what happens on their platform – not only in terms of content moderation and curation. Due to centralisation, providers can harvest user data. This user data can target users with (personalised) advertisements.[506] Advertising and related data services became the principal revenue stream for providers. Protocolisation means that this control is given away by opening up to other providers. Protocolisation, thus, may have beneficial effects on user rights and competition between providers.[507]As Masnick notes, it is not necessary "to build an entirely new Facebook if you already have access to everyone making use of the 'social network protocol'".[508]

The Magic API and protocolisation are both examples of what Fukuyama et al. call "middleware", defined as "software, provided by a third party and integrated into the dominant platforms, that would curate and order the content that users see."[509] Middleware would limit the control of providers over political content by allowing users to choose between different curation services.[510]According to Fukuyama et al., middleware solutions would be preferable to breaking up providers that would be technologically hard and might even be counterproductive to reaching other goals, such as preventing the amplification of harmful user-provided information.[511] Keller, however, is not convinced that these proposals (including the Magic API) would be beneficial below the line but views it preferable to consider these alternatives than imposing a must-carry obligation for providers.[512]

Where moderation, according to Gillespie, is "the commodity" platforms offer,[513] some providers use curation to keep users' attention to their platforms by recommending relevant

---

504 Keller, 2019, 'Who Do You Sue? State and Platform Hybrid Power over Online Speech', pp. 26-27.

505 Masnick, 2019, 'Protocols, Not Platforms: A Technological Approach to Free Speech', pp. 5-7 and 14.

506 Masnick, 2019, 'Protocols, Not Platforms: A Technological Approach to Free Speech', p. 11.

507 Masnick, 2019, 'Protocols, Not Platforms: A Technological Approach to Free Speech', p. 15.

508 Masnick, 2019, 'Protocols, Not Platforms: A Technological Approach to Free Speech', p. 15.

509 Fukuyama, et al., 2021, 'Middleware for Dominant Digital Platforms: A Technological Solution to a Threat to Democracy', p. 3.

510 Fukuyama, et al., 2021, 'Middleware for Dominant Digital Platforms: A Technological Solution to a Threat to Democracy', p. 3.

511 Fukuyama, et al., 2021, 'Middleware for Dominant Digital Platforms: A Technological Solution to a Threat to Democracy', p. 4.

512 Keller, 2019, 'Who Do You Sue? State and Platform Hybrid Power over Online Speech', pp. 26-27.

513 Gillespie, 2018, *Custodians of the Internet*, p. 207.

content and monetising their services by offering relevant ads to their users.[514] As noted, there are beneficial effects to be expected for users' freedom of expression and privacy rights from middleware solutions. However, it would shake up the business model of many providers, making it hard to predict what the effects of middleware solutions would become to mean for the availability of providers. Besides, it deserves attention to how middleware providers could monetise their services.[515]

Some alternatives do not remedy the central position of providers but may offer user control over curation. Providers may offer tools for the user of these services to customise their experience on the service by choosing what categories of content they wish to see. Some of these possibilities still qualify as curation by the provider; other possibilities are entirely controlled by users and thus customisation. According to Goldman, user-controlled interventions have significant benefits over service-level interventions. User-controlled interventions do not affect all users, while service-level interventions do.[516] Service-level interventions, however, are the default. As Masnick notes, providers are, due to centralisation and concentration, able to make such decisions for a large user base.[517] In some jurisdictions, such as the EU, internet content regulation is tied to the possibility of service-level interventions.[518] In other words, internet intermediary regulation's success depends on large internet platforms that can moderate user-provided information that contains illegal or unlawful content.[519] Therefore, leaving content-based interventions over to users is limited to curating and moderating content that is not illegal or unlawful.

Goldman points out that user-controlled interventions know risks as well. User-controlled curation, for example, may lead to reinforcement of beliefs users already hold because they choose content that fits their convictions. Such "filter bubbles" are, according to Goldman, however, preferable over service-level interventions.[520]

2.3 Remedies: a sanction regime that fits the violation

Service providers can affect user-provided content in numerous ways. Providers, for example, can remove content or make content inaccessible for groups of users. Such interventions occur after the violation of a rule laid down in the terms and conditions of the intermediary. As noted, these terms of conditions also encompass requirements by state legislation. Following moderation, an individual video, photograph, or post may be removed or made inaccessible by a service provider. Besides, providers may impose remedies on a group, page, or whole accounts. As already mentioned, the whole process of rule-setting, interpretation, detecting violations and choosing

---

[514] See, for this mechanism, Zuboff, 2019, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, pp. 93-97.

[515] Fukuyama, et al., 2021, 'Middleware for Dominant Digital Platforms: A Technological Solution to a Threat to Democracy', pp. 8-9.

[516] Goldman, 'Content Moderation Remedies', *Michigan Technology Law Review*, 2021, p. 54.

[517] Masnick, 2019, 'Protocols, Not Platforms: A Technological Approach to Free Speech', p. 17. See also, Goldman, 'Content Moderation Remedies', *Michigan Technology Law Review*, 2021, p. 55.

[518] For example, in the EU the obligations of providers are related to the amount of users, see Article 25(1) of Commission Proposal COM(2020) 825 final (*Digital Services Act*), p. 59.

[519] A provider may lose protection under the safe harbour in the EU when it fails to remove of disable access to illegal or unlawful content when it gains knowledge of such content, see Article 14 of Directive 2000/31/EC (*Directive on electronic commerce*).

[520] Goldman, 'Content Moderation Remedies', *Michigan Technology Law Review*, 2021, p. 55.

remedies is referred to as content moderation. The last step, deciding what remedy fits the violation, is the field of content moderation remedies. According to Goldman, what happens in the last step of content moderation is easily dominated by the other steps. The blind spot for the remedy toolbox is not without risks. A lack of definitional clarity and a more refined toolset of moderation options may cause service providers and states to go to the fall-back default option: removal.[521] Removal of all remedies, of course, is one of the most impactful on users' freedom of expression rights.

2.3.1 Content moderation remedies: definition

As noted above, there is a lack of clarity on what should be considered content moderation remedies. The DSA, for example, offers a comprehensive definition viewing all remedies that "affect the availability, visibility and accessibility" and "the recipients ability to provide that information" as content moderation remedies.[522] In contrast, the CDMSI only views action undertaken against a specific instance of user-provided information as a content moderation remedy.[523] Direct interventions on the visibility of user-provided information by removing or blocking access to (specific instances) of content are generally understood as content moderation remedies when they occur after a rule violation.[524] Goldman dubbed this the "binary approach". User-provided information is either left up or taken down after assessing whether its content violates the rules.[525] According to Goldman, this is the default approach guiding governmental and non-governmental thinking about choosing a remedy.[526] State actors such as the EC seek to address content that is not illegal but still potentially harmful with a more diverse set of tools such as labelling, prioritising, warning and counter-speech.[527] According to Goldman, all remedies imposed after rule violation could be considered content moderation remedies – irrespective of the nature of the rule. Goldman:

> the responses are intended to remediate the rule violation, in the same way that a court grants remedies to successful litigants who are entitled to legal relief.[528]

While the remedy is deployed after the rule violation is established, this does not exclude the possibility of ex-ante moderation by screening user-provided information for rule violating content. "Post-production moderation", meaning that the providers moderate content after publication, is the norm. However, this norm does not exclude other moderation efforts. For example, "pre-production moderation" (reviewing content before admission) or other moderation systems such as "peer-based moderation" (leaving moderation to the users themselves) are not considered moderation.[529] How the rule violation is uncovered is not decisive to speak of content moderation remedies.

---

[521] Goldman, 'Content Moderation Remedies', *Michigan Technology Law Review*, 2021, pp. 5-9.

[522] Article 2(p) of Commission Proposal COM(2020) 825 final (*Digital Services Act*).

[523] Council of Europe, 2021, 'Content moderation: best practices towards effective legal and procedural frameworks for self-regulatory and co-regulatory mechanisms of content moderation', p. 11.

[524] While removal by the provider for a different reason than a rule violation is hard to imagine, hiding content in a specific region to prevent violation of intellectual property rights may be not considered content moderation.

[525] Goldman, 'Content Moderation Remedies', *Michigan Technology Law Review*, 2021, pp. 4-6.

[526] Goldman, 'Content Moderation Remedies', *Michigan Technology Law Review*, 2021, p. 12.

[527] See, for example, European Commission, 2021, 'Code of Practice on Disinformation'.

[528] Goldman, 'Content Moderation Remedies', *Michigan Technology Law Review*, 2021, p. 9.

[529] OECD, 2007, 'Participative Web and User-Created Content', p. 92.

Besides impacting user-provided information, moderation can also impact a whole platform offered by other providers. For example, when a provider ceases to offer services to another provider because of the content of user-provided information shared on the platform. Van Dijck, De Winkel & Schäfer call this "deplatformization", which "applies to tech companies' efforts to *reduce toxic content by pushing back controversial platforms and their communities to the edge of the ecosystem, denying them access to basic infrastructural services needed to function online*."[530] It is not hard to see how such moderation can have more severe restrictions on the possibility for users to express themselves.

In addition, content moderation remedies can be public or private. Goldman distinguishes remedies in private content moderation from remedies that the state can deploy. Goldman "focuses on editorial decisions implemented by private entities, not decisions made by government state actors",[531] arguing that "[p]rivate actors, with their structurally different attributes, raise different considerations".[532] Of course, private actors are different from state actors, especially in terms of accountability and the possible remedies that can be used.[533] Providers may require the help of the state to make use of some remedies.[534]

Because of this demarcation, the discussion is whether state efforts should be seen as content moderation remedies. Is the involvement of the providers necessary to speak of content moderation remedies? Can, for example, a court sentence for posting hate speech be regarded as content moderation? Although the service provider takes no action, the violation of the rules is addressed with a remedy. In some cases, such an approach may be preferable to direct intervention by the service provider. For example, a rule violation may be better to be left to the courts when the rule violation is hard or impossible to establish by a service provider. For example, in the case of libel or slander, a remedy imposed by the provider may do more harm than good. Sometimes a remedy chosen by the provider is not enough, for example, in the case of a severe violation of legal rights. An extreme example is online child sexual abuse material. Some delineation, however, is necessary. Not all state interventions related to the content of user-provided information on internet intermediary services should be understood as content moderation remedies. To be called such, a content moderation remedy should relate directly to the posted content and not all events related to this content.

Providers can also affect the visibility of user-provided content more subtle. Such an intervention affecting the visibility of user-provided information may not necessarily follow a rule violation and is not always considered a content moderation remedy. For example, providers can stop recommending user-provided information with specific content categories to (specific) groups of users, delisting from the search, or altering content prioritisation. Interventions on the

---

[530] J. van Dijck, T. de Winkel & M.T. Schäfer, 'Deplatformization and the governance of the platform ecosystem', *New Media & Society*, 2021 (available at journals.sagepub.com/doi/full/10.1177/14614448211045662), doi:10.1177/14614448211045662, p. 4.

[531] Goldman, 'Content Moderation Remedies', *Michigan Technology Law Review*, 2021, p. 10.

[532] Goldman, 'Content Moderation Remedies', *Michigan Technology Law Review*, 2021, p. 12.

[533] Goldman, 'Content Moderation Remedies', *Michigan Technology Law Review*, 2021, pp. 9-13.

[534] For example, to prevent users circumventing remedies imposed by the provider, see C. D'Anastasio, 'Twitch Sues Users Over Alleged 'Hate Raids' Against Streamers', *Wired,* 10 September 2021, available at wired.com/story/twitch-sues-users-over-alleged-hate-raids (retrieved on 14 February 2022).

visibility of content without removing or making the content inaccessible for other reasons than to remedy a rule violation can be called content curation.[535]

While there is a difference between content moderation and content curation, these terms are often used interchangeably or are conflated within content moderation.[536] For good reasons: from the user's viewpoint, both moderation and curation may affect the possibility of reaching an audience. Not being recommended to other users may be equally impactful as removing or hiding an individual instance of user-provided information because of its content.

2.3.2 Limitations on content moderation remedies?

Goldman points out that private actors conduct content moderation and thus impose the remedies to user accounts or the information provided by the user. Private actors, of course, differ from government actors in terms of accountability and constitutional limits.[537] However, there are voices to subject providers to the same norms as the state when engaging in content moderation in academic and governmental debates.[538] The CDMSI, for example, argues that "removal of an online post is a limitation of a user's freedom of expression, so this also needs to be done in a way which is predictable, legitimate, necessary and proportionate."[539] The CDMSI emphasises that moderation decisions may result from private or/and state decision-making.[540] Service providers enforce terms and conditions that may leave something to wish for when it comes to clarity. Such unclarity may also be caused by equally unclear terminology in legislation.[541]

Besides, content moderation is not tied to, for example, the physical presence of a user as the state is: if it is possible to program it, it is possible to use it as a remedy.[542] However, some remedies that states can use are not available to providers. For example, a provider cannot seize the users' physical possessions for not fulfilling their end of a transaction without the help of the state.[543]

Goldman argues that removal is considered the default remedy in internet content regulation. Removal, however, has a considerable disadvantage.[544] The CDMSI notices that content moderation sees to different problems. Not only the subject of the content that is

---

535 Council of Europe, 2021, 'Content moderation: best practices towards effective legal and procedural frameworks for self-regulatory and co-regulatory mechanisms of content moderation', p. 11.

536 See, for example, Council of Europe, 2021, 'Content moderation: best practices towards effective legal and procedural frameworks for self-regulatory and co-regulatory mechanisms of content moderation', p. 3.

537 Goldman, 'Content Moderation Remedies', *Michigan Technology Law Review*, 2021, pp. 9-13.

538 Adviesraad Internationale Vraagstukken, 2020, 'Regulering van online content: Naar een herijking van het Nederlandse internetbeleid (AIV-advies 113)', p. 13; Office of the United Nations High Commissioner for Human Rights, *Guiding Principles on Business and Human Rights: Implementing the United Nations "Protect, Respect and Remedy" Framework*, UN Doc. HR/PUB/11/04, pp. 13-16 (2011).

539 Council of Europe, 2021, 'Content moderation: best practices towards effective legal and procedural frameworks for self-regulatory and co-regulatory mechanisms of content moderation', p. 14.

540 Council of Europe, 2021, 'Content moderation: best practices towards effective legal and procedural frameworks for self-regulatory and co-regulatory mechanisms of content moderation', p. 18.

541 Council of Europe, 2021, 'Content moderation: best practices towards effective legal and procedural frameworks for self-regulatory and co-regulatory mechanisms of content moderation', p. 18.

542 Goldman, 'Content Moderation Remedies', *Michigan Technology Law Review*, 2021, p. 11.

543 Goldman, 'Content Moderation Remedies', *Michigan Technology Law Review*, 2021, pp. 10-11.

544 Goldman, 'Content Moderation Remedies', *Michigan Technology Law Review*, 2021, pp. 21-22.

moderated may differ, but also the problems tied to such content.[545] Goldman argues that removal causes 'collateral damage' such as 1) the removal of evidence, 2) leaving other posts interacting with the removed content without context, 3) causing linkrot, 4) removing content that does not violate the rules (comments on or interactions with removed content or in the case of account removal all posts of the user in question).[546] Goldman first distinguishes between moderation remedies directed at 1) the content of individual instances of user-provided information and 2) individual accounts from which the user-provided information with the violating content is provided. Besides, both accounts as user-provided information 3) can be subjected to regulation by reducing the visibility. Of course, the visibility of one instance of user-provided information is less far-reaching than affecting the visibility of all user-provided information posted from an account. In addition to remedies seeing to user accounts or user-provided information, it is possible to impose 4) monetary remedies on the ability to monetise the usage of a service or even contractual fines, and 5) a category with other remedies that do not fit the previous categories.[547] As Goldman noted, the regulatory toolkit of providers is only limited by imagination and technological possibilities.[548] A broad range of instruments is beneficial, considering that "removal by default" may undoubtedly result in overregulation.[549]

How do content moderation and content curation differ now that curation can also be used to remedy rule violations similar to moderation? Not the intervention, but the reason should be decisive: curation to remedy a rule violation should be considered moderation and be subjected to enhanced oversight. Content curation is deployed not to remedy rule violations but to ensure quality control should be left to the providers. Therefore, there may be good reasons to leave categories of content unregulated – especially when a provider has to decide on the quality.

2.4 Scope: international, state, and intermediary regulation

Many providers, one way or another, have an international presence. These providers may offer services to users across jurisdictions, have a physical (for example, servers) or legal (daughter companies) presence in multiple jurisdictions, or even facilitate the cross-border exchange of goods and services as part of their service. Because providers operate globally, this raises questions over the applicability of regulation from the territorial state to providers and their users.

Internet content regulation may be carried out at multiple points on the network. Firstly, it is possible to impose regulations on users who posted or received information with illegal or unlawful content. Hence, the state where the user is physically present can claim jurisdiction over the user.[550] Besides users, providers rely on physical locations. Territorial states can impose regulations on the physical location or computers where the user-provided information is hosted. Next to the location of the user and the physical location of the servers of providers, the third possibility for regulation is the legal entity that exploits the different internet intermediary services.

---

[545] Council of Europe, 2021, 'Content moderation: best practices towards effective legal and procedural frameworks for self-regulatory and co-regulatory mechanisms of content moderation', p. 16.

[546] Goldman, 'Content Moderation Remedies', *Michigan Technology Law Review*, 2021, pp. 21-22.

[547] Goldman, 'Content Moderation Remedies', *Michigan Technology Law Review*, 2021, pp. 23-24.

[548] Goldman, 'Content Moderation Remedies', *Michigan Technology Law Review*, 2021, p. 11.

[549] Council of Europe, 2021, 'Content moderation: best practices towards effective legal and procedural frameworks for self-regulatory and co-regulatory mechanisms of content moderation', p. 19.

[550] In criminal law, states may also claim jurisdiction over criminal offenses committed outside the territory of the state see, for example, Article 7 of Wetboek van Strafrecht (Dutch Criminal Code).

The legal entity and the physical location of the servers do not necessarily correspond. The jurisdiction in which the internet intermediary has its legal establishment (or its subsidiary) or in which its legal owner has a presence may impose and enforce regulations on these legal entities. Fourthly, (non-profit and for-profit) organisations maintain (parts) of the network layer of the internet infrastructure, such as domain registries. As noted in Chapter 1, these entities could also be subjected to state regulation, while this may be undesirable.[551]

Of course, it is hard to successfully prosecute, convict and execute penalties when the defendant is not within the state's territory. It is not always possible to successfully execute a court order, for example, when the defendant's state does not respect the foreign court ruling in question.[552] Some jurisdictions require providers to establish an office or appoint a legal representative within their territory, which may increase compliance with state regulations. An example is the EC proposing such an obligation in new draft legislation for the DSA.[553] The fear exists that some countries, such as Turkey and India, may use such a representative as a target for pressuring a provider to censor content for the government.[554]

Because it may be hard to regulate the hosting service provider offering the user-provided information with illegal or unlawful content, sometimes the state chooses network layer interventions. Network layer interventions can be imposed by regulating the ISPs that offer services within the jurisdiction that seeks to block specific instances of information. A clear example is a legal requirement for some Dutch ISPs to block connections to an illegal online file-sharing platform called *The Pirate Bay*, which led to lengthy legal proceedings before the Dutch court and the ECJ.[555] Network layer interventions are critically reviewed. One downside of network layer interventions is that it is not easy to discriminate between legal and illegal content. Blocking *The Pirate Bay* also blocks access to content that could be considered legal or even protected under

---

[551] B. de La Chapelle & P. Fehlinger, 'From Legal Arms Race to Transnational Cooperation', *in* G. Frosio (Ed.) *The Oxford Handbook of Online Intermediary Liability*, Oxford, Oxford University Press, 2020, doi:10.1093/oxfordhb/9780198837138.013.38, p. 729.

[552] One of the first cases discussing the enforcement of foreign judgements against internet intermediaries in the US was *Yahoo! Inc. v. La Ligue Contre Le Racisme et l'antisémitisme (LICRA)*, 433 F.3d 1199, 1218-1221 (9th Cir. 2006). Since 2010, the US bars the enforcement of foreign judgments concerning defamation, unless the defamation law 'provided at least as much protection for freedom of speech and press in that case as would be provided by the first amendment', see §4102. Recognition of foreign defamation judgments, 28 USCA § 4102(a)(1)(A) (West 2010, Westlaw Next through PL 116-150). Interactive computer services – the legal category also encompassing internet intermediaries – are protected for the enforcement of such judgements 'unless the domestic court determines that the judgment would be consistent with section 230 if the information that is the subject of such judgment had been provided in the United States.', see 28 USCA § 4102(c)(1) (West 2010, Westlaw Next through PL 116-150). See, for a definition of 'interactive computer service' 47 USCA § 230(f)(2) (West 2018, Westlaw Next through PL 116-91). See also, Goldman, 2020, 'An Overview of the United States' Section 230 Internet Immunity', p. 160.

[553] 'Providers of intermediary services which do not have an establishment in the Union but which offer services in the Union shall designate, in writing, a legal or natural person as their legal representative in one of the Member States where the provider offers its services.' see Article 11(1) of Commission Proposal COM(2020) 825 final (*Digital Services Act*).

[554] M. Santora, 'Turkey Passes Law Extending Sweeping Powers Over Social Media', *The New York Times,* 29 July 2020, available at nytimes.com/2020/07/29/world/europe/turkey-social-media-control.html (retrieved on 15 February 2022); BBC, 'Twitter fears for freedom of expression in India', *BBC,* 27 May 2021, available at bbc.com/news/world-asia-india-57265331 (retrieved on 14 February 2022).

[555] Which led to lengthy procedures, see, for example, Judgment of the Court (Second Chamber) of 14 June 2017 in *C-610/15, Stichting Brein v Ziggo BV and XS4All Internet BV*, ECLI:EU:C:2017:456; HR, 13 November 2015, ECLI:NL:HR:2015:3307, *Nederlandse Jurisprudentie* 2018/110, m.nt. P.B. Hugenholtz.

freedom of expression regulation. Such collateral damage normally renders network layer interventions unsuitable for content regulation. However, it may be an option to raise the barriers to accessing a service known for facilitating user-provided information with illegal content. It is necessary to emphasise that it is only barrier raising because it is impossible to prevent users from accessing the service.[556]

Because content regulation aims to delete specific instances of illegal content, states imposing such regulations wish to target hosting service providers. Because of the regulatory capabilities of these providers, providers are a potential target for state regulation. When hosting service providers impose application layer restrictions, these restrictions may even only limit access from jurisdictions where specific content is illegal. Such efforts by providers offer a new mode of regulation for states.[557] However, this does not remedy clashing state norms and thus questions the applicability of these norms. Such conflicts could especially arise if states impose different norms on providers. Besides, states seeking to increase their regulatory capability by regulating the infrastructure instead of the application layer service may directly impact the sovereignty of other states.

For example, De La Chapelle and Fehlinger warn that increasing state ambitions to impose regulation over internet content may lead to "either extending sovereignty beyond national frontiers or strictly reimposing national borders."[558] Extending sovereignty may lead to adverse effects on the global nature of the internet. Already in 2014, multiple scholars warned in the Financial Times that the internet might become "balkanised" because democracies enforce new policies to protect their citizens while Turkey and Russia enforce similar policies to get a tighter grasp on the internet for security reasons.[559] The EC has deployed many policy instruments to tackle hate speech,[560] online terrorist content[561] and disinformation.[562] Citron warns that such policies may also impact jurisdictions that consider the regulated categories of speech protected under freedom of expression rights.[563]

In 2020, the Dutch Advisory Council on International Affairs (hereafter: AIV) warned in a policy advisory report for the Dutch government that national or regional policies

---

[556] Gerechtshof Amsterdam, 2 June 2020, ECLI:NL:GHAMS:2020:1421, Rec. 3.8.9; Judgment of the Court (Fourth Chamber) of 27 March 2014 in *C-314/12, UPC Telekabel Wien GmbH v Constantin Film Verleih GmbH and Wega Filmproduktionsgesellschaft mbH*, ECLI:EU:C:2014:192, in particular Rec. 62.

[557] In the *Yahoo!*-case, the US Court argued that the first amendment does not necessarily offer protection for internet intermediaries against governmental interference in a domestic context in which no US citizens are involved: "Yahoo! is necessarily arguing that it has a First Amendment right to violate French criminal law and to facilitate the violation of French criminal law by others. As we indicated above, the extent – indeed the very existence – of such an extraterritorial right under the First Amendment is uncertain.", see *Yahoo! Inc. v. La Ligue Contre Le Racisme et l'antisémitisme (LICRA)*, 433 F.3d 1199, 1221 (9th Cir. 2006).

[558] De La Chapelle & Fehlinger, 2020, 'From Legal Arms Race to Transnational Cooperation', p. 732.

[559] FT reporters, 'Tying up the internet', *Financial Times,* 16 September 2014, available at ft.com/content/2f2f7274-3a5e-11e4-bd08-00144feabdc0 (retrieved on 18 June 2021).

[560] European Commission, 'European Commission and IT Companies announce Code of Conduct on illegal online hate speech', *European Commission,* 31 May 2016, available at ec.europa.eu/commission/presscorner/detail/en/IP_16_1937 (retrieved on 14 February 2022); European Commission, 2016, 'Code of Conduct on Countering Illegal Hate Speech Online'.

[561] Regulation (EU) 2021/784.

[562] Communication COM(2018)236 final; European Commission, 2021, 'Code of Practice on Disinformation'.

[563] Citron, 'Extremist Speech, Compelled Conformity, and Censorship Creep', *Notre Dame Law Review*, 2018.

> creates the risk of a disintegrated and fragmented 'splinter net'. Such cyber-balkanisation will inevitably undermine the internet as a cross-border medium for free expression and access to information.[564]

Fragmentation of the internet as a global network along jurisdictional (legal) or political borders is high on the policy agenda, especially since the global nature of the internet is considered vital for exercising freedom of expression rights.[565] Freedom of expression also includes the right to receive information – "regardless of frontiers".[566] Regulatory interventions in the EU and the US are increasingly diverging, which may cause new fragmentation.[567] Next to governmental policy, courts are more involved in internet content regulation leading to court decisions that are either given a global reach or are merely enforced locally.[568] Fragmentation along legal lines is not surprising since regulation of providers (such as protection of human rights of users of the services) takes place along the lines of the territorial state.[569] Therefore, the Dutch AIV argues that the protection of human rights on the internet should be prioritised higher than maintaining an open and global internet. Some fragmentation should be taken for granted when this protects human rights values.[570]

One example of such an effort is that states seek to expand their regulatory capabilities by requiring providers to keep user data as much as possible within their territory and thus jurisdiction.[571] Russia's RuNet aims to function autonomously from the global internet – an effort that, according to Musiani, can be labelled as "internet sovereignty" – is a clear example of such an attempt.[572] De La Chapelle and Fehlinger argue that asserting sovereignty over the internet leads to paradoxes. The first paradox is that extraterritorial regulation enacted to assert sovereignty often impacts the sovereignty of other states. Extraterritorial regulation tends to violate the principle of non-intervention which underpins sovereignty. The second paradox is that asserting sovereignty by territorialising parts of the internet does not safeguard the sovereignty of states that cannot maintain large data centres. The second paradox thus also decreases the sovereignty of some states

---

564 In Dutch "Hierdoor ontstaat het risico op een uiteen gevallen en gefragmenteerd 'splinternet'. Een dergelijke cyberbalkanisering' zorgt voor een onvermijdelijke aantasting van het internet als grensoverschrijdend medium voor vrije expressie en toegang tot informatie.", see Adviesraad Internationale Vraagstukken, 2020, 'Regulering van online content: Naar een herijking van het Nederlandse internetbeleid (AIV-advies 113)', p. 46.

565 Benedek & Kettemann, 2020, *Freedom of Expression and the Internet*, p. 18.

566 Article 11(1) of Charter of Fundamental Rights of the European Union, *OJ C 326*, *26.10.2012* (data.europa.eu/eli/treaty/char_2012/oj); Article 19(2) of International Covenant on Civil and Political Rights, 16 December 1966, 999 U.N.T.S. 171; Article 10(1) of the European Convention on Human Rights.

567 Transatlantic High Level Working Group on Content Moderation Online and Freedom of Expression, 'Freedom and Accountability: A Transatlantic Framework for Moderating Speech Online', *The Annenberg Public Policy Center of the University of Pennsylvania,* 2020, available at annenbergpublicpolicycenter.org/feature/transatlantic-working-group-freedom-and-accountability, p. 12.

568 A. Callamard, 'Are courts re-inventing Internet regulation?', *International Review of Law, Computers & Technology*, Vol. 31, No. 3, 2017, doi:10.1080/13600869.2017.1304603, pp. 333-334.

569 G. De Gregorio, 'Democratising online content moderation: A constitutional framework', *Computer Law & Security Review*, Vol. 36, No. 105374, 2020, doi:10.1016/j.clsr.2019.105374, p. 9.

570 Adviesraad Internationale Vraagstukken, 2020, 'Regulering van online content: Naar een herijking van het Nederlandse internetbeleid (AIV-advies 113)', p. 11.

571 De La Chapelle & Fehlinger, 2020, 'From Legal Arms Race to Transnational Cooperation', p. 734.

572 F. Musiani, '*Infrastructuring* digital sovereignty: a research agenda for an infrastructure-based sociology of digital self-determination practices', *Information, Communication & Society*, Vol. 25, No. 6, 2022, doi:10.1080/1369118X.2022.2049850, pp. 793-796.

now that data transfers between states are not limited.[573] In other words, the state with the means to require local storage of user data sees its regulatory capabilities increase while states that do not have such means lose those capabilities. For example, when Germany (hypothetically) would require local storage of their users' data, this may mean that providers no longer desire to maintain data centres in the Netherlands. Instead, providers would prefer to move those centres to Germany to comply with German law while these data centres still could maintain their regional function.

While it is hard to notice borders on the internet, states may willingly or accidentally establish such borders. Svantesson, therefore, argues that "there is a fundamental clash between the global, largely borderless, internet on the one hand, and the practice of lawmaking and jurisdiction anchored in territorial thinking."[574] This difficulty, however, does not render the territorial state obsolete. The question, however, is what should happen when a local court seeks to apply local or regional standards globally.[575]

Is it not possible to make state borders on the internet irrelevant? A solution would be to harmonise internet content regulation globally. However, as Svantesson points out, the debate dances around two conflicting views on what values the internet should uphold the possibility of almost unlimited (absolute) freedom of expression rights and, on the other hand, the possibility of regulating content that is considered harmful.[576] While there is no consensus on which of these two options should guide such internet content regulations, it becomes even harder to gain consensus on the material aspect of internet content regulation. What should be considered illegal? What should be considered harmful? Even among the EU Member States, there are considerable differences between states.[577] From a state perspective, harmonising internet content regulation seems next to impossible.

What would the users of the internet choose? Users would benefit from an internet guided by human rights standards and providers that would refuse regulation that does not comply with these human rights standards. Svantesson, therefore, considers the option of an "international law doctrine of selective legal compliance".[578] Service providers should follow legislation and court orders that respect human rights law while ignoring those that violate human rights. However, as already argued, there are conflicting views on how these human rights must be interpreted. Besides, sometimes newly rights are explicitly acknowledged as human rights in one or more jurisdictions. An example forms the right to data protection in the EU.[579] Leaving providers to decide how and

---

[573] De La Chapelle & Fehlinger, 2020, 'From Legal Arms Race to Transnational Cooperation', p. 735.

[574] D.J.B. Svantesson, 'Internet Jurisdiction and Intermediary Liability', *in* G. Frosio (Ed.) *The Oxford Handbook of Online Intermediary Liability*, Oxford, Oxford University Press, 2020, doi:10.1093/oxfordhb/9780198837138.013.3, pp. 691-692.

[575] De La Chapelle & Fehlinger, 2020, 'From Legal Arms Race to Transnational Cooperation', pp. 733-734.

[576] Svantesson, 2020, 'Internet Jurisdiction and Intermediary Liability', pp. 692-693.

[577] For example, the German NetzDG, is one of its kind, see the Network Enforcement Act 2017 (*Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken*). The EU does not offer any harmonisation for criminalisation of hate speech, see Council Framework Decision 2008/913/JHA.

[578] Svantesson, 2020, 'Internet Jurisdiction and Intermediary Liability', p. 693.

[579] For example, the 'Protection of personal data' laid down in Article 8 of Charter. Paragraph 1 reads: "Everyone has the right to the protection of personal data concerning him or her." Paragraph 2 sets out what kind of protection: "Such data must be processed fairly for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis laid down by law. Everyone has the right of access to data which has been collected concerning him or her, and the right to have it rectified." Between EU member-states and the parties of the ECHR differences between how privacy and data protection rights are enforced exist, see Council of Europe,

to what extent human rights should affect their services may lead to undesirable consequences for one or more jurisdictions. Besides, there are concerns about legitimacy as well. Not private providers but democratically legitimised states should decide on human rights to online services. Geist warns that allowing providers to cherry-pick would lay too much power in their hands.[580]

The complexity of the international dimension of regulating providers leaves us with two uncomfortable (possible) outcomes. Geist points out that the first possible outcome is that providers are made the ultimate arbiters regulating the content of user-provided information because they decide how court orders from national states are given effect.[581] On the other hand, laying internet content regulation in the hands of large providers raises questions about due process requirements.[582] The CDMI, therefore, recommends state involvement as a positive obligation under the ECHR, requiring the state to set out the legal framework in which content moderation takes place. For example, the state could enact legislation that imposes requirements on the terms of services of providers.[583]

A second possible outcome, according to Geist, is that content regulation is left over to the local courts. These rulings, however, could lead to new problems when they are given global effect.[584] Svantesson argues against the position that substantive laws seeing to internet content regulation of one jurisdiction, should automatically apply globally. Giving local laws global effects would ultimately raise conflicts with the laws in other jurisdictions.[585] Geist, therefore, argues that a global takedown should only be issued "where it is clear that the underlying right and remedy are also available in affected foreign countries."[586] Svantesson argues that courts should take notice of the "scope of jurisdiction". While the court may have personal and subject matter jurisdiction, this does not mean that the court should not consider the geographical scope of a court order.[587] According to Svantesson, states should only claim jurisdiction when there is a "substantial connection" and "legitimate interest" and when exercising jurisdiction "is reasonable given the balance between the state's legitimate interest and other interests."[588]

'Comparative Study on Blocking, Filtering and Take Down of Illegal Internet Content', *Council of Europe,* 2017, available at edoc.coe.int/en/internet/7289-pdf-comparative-study-on-blocking-filtering-and-take-down-of-illegal-internet-content-.html, p. 16. Data protection rights such as 'the right to be forgotten' may impact other jurisdictions where such privacy rights are not recognised there to the same extent, see Judgment of the Court (Grand Chamber) of 13 May 2014 in *C-131/12, Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González*, ECLI:EU:C:2014:317, in particular Rec. 96-98; Balkin, 'Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation', *U.C. Davis Law Review*, 2018, pp. 1204-1206. The 'right to be forgotten' may even come in conflict with First amendment protections in the US, see Bloch-Wehba, 'Global Platform Governance: Private Power in the Shadow of the State', *SMU law review*, 2019, pp. 58-59.

[580] M. Geist, 'The Equustek Effect: A Canadian Perspective', *in* G. Frosio (Ed.) *The Oxford Handbook of Online Intermediary Liability*, Oxford, Oxford University Press, 2020, doi:10.1093/oxfordhb/9780198837138.013.37, p. 714 and 724.

[581] Geist, 2020, 'The Equustek Effect: A Canadian Perspective', p. 724.

[582] Goldman, 'Content Moderation Remedies', *Michigan Technology Law Review*, 2021, p. 53.

[583] Council of Europe, 2021, 'Content moderation: best practices towards effective legal and procedural frameworks for self-regulatory and co-regulatory mechanisms of content moderation', p. 24.

[584] Geist, 2020, 'The Equustek Effect: A Canadian Perspective', p. 725.

[585] Svantesson, 2020, 'Internet Jurisdiction and Intermediary Liability', pp. 702-703.

[586] Geist, 2020, 'The Equustek Effect: A Canadian Perspective', p. 726.

[587] Svantesson, 2020, 'Internet Jurisdiction and Intermediary Liability', pp. 699-700.

[588] Svantesson, 2020, 'Internet Jurisdiction and Intermediary Liability', p. 699.

The scope of the norms is not the only concern. Also, what remedies may follow on violation of these norms must be considered. Goldman argues that providers should localise remedies. Instead of global measures, remedies should only be implemented in countries where the content of user-provided information violates the local law.[589] Goldman, however, prefers private remedies over judicial remedies. Courts may be slow, costly, and constrained by rules about jurisdiction. Remedies imposed by service providers, of course, raise different questions. Faster decisions may pose a risk to the quality of the procedure. Besides, Goldman points out that court procedures may be counterproductive since this may increase the attention to user-provided information with illegal content.[590]

Conclusion

This chapter discusses the scope and limitations of internet content regulation by regulating providers. Distinguished in this chapter are four dimensions that influence regulating internet content. The target (the first dimension), instruments (the second dimension), and remedies deployed (the third dimension) by providers influence potential overregulation and underregulation. The territorial scope of the application (the fourth dimension) is, in its turn, decisive for potential (extraterritorial) effects of the regulation.

As discussed, the target of content regulation is dependent on the liability regime that is enacted. Some liability regimes completely exonerate providers from any liability that may arise from the content of user-provided information, making the user who provided the information with violating content a target for regulation. However, as shown in Part 2 of this dissertation, the liability regimes discussed here have a more refined approach to distributing liability between the provider and the user that submitted the content. Service providers are generally exonerated from legal liability when they fulfil a set of conditions. However, under such regimes, providers may become liable for the content of user-provided information when they (for example) gain knowledge or awareness of illegal or unlawful content. Regulation may also be differentiated between the providers based on their size and the roles these providers may fulfil. As noted, such differentiation is not without risk. As argued in Chapter 1, such regulation may cause providers to alter their services so that they are not included under the definitions of such regulation.

In addition to the service provider, the regulation targets categories of conduct or even content. Regulation may aim to regulate how providers moderate the content of user-provided information. Other regulations may be imposed to influence what user-provided information is recommended to other users and what is not. As shown in this chapter, both moderation and curation efforts may severely influence users' freedom of expression rights. The distinction between moderation and curation is thus mainly one of responsibility for the service provider. As noted, moderation is reactive (after a rule violation is established). Curation, in contrast, is an ongoing process: providers often curate user-provided information on an ongoing basis. Both efforts, however, can be exposed to regulation.

The third dimension is the remedies that follow a rule violation. As noted, the default option is removal after a rule violation is established. Such a rule violation is often complemented with a strike. A user that accumulates enough strikes may be confronted with account-level sanctions (in the most extreme case, even termination of the user account). A more diverse system

[589] Goldman, 'Content Moderation Remedies', *Michigan Technology Law Review*, 2021, pp. 53-54.
[590] Goldman, 'Content Moderation Remedies', *Michigan Technology Law Review*, 2021, pp. 52-53.

of potential remedies may prevent the harmful effects of this default option. Some content of user-provided information, for example, is not strictly illegal and thus does not justify removal as a moderation remedy. While providers should not restrict their options following moderation to removal, a significant cause of such an approach comes from state regulation that only considers removal enough remedy to prevent providers from becoming liable. State regulation, thus, also should allow for a multitude of remedies and not just removal – especially in the case of borderline illegal or unlawful content.

Differentiation and diversification of remedies are essential in light of the international dimension of internet content regulation. States should not impose remedies such as removal internationally when these norms and remedies are not explicitly recognised in foreign jurisdictions. Primarily when these norms are enforced with remedies with far-reaching consequences such as removal or account terminations, such a overstretch may severely impact the freedom of expression rights of users in foreign jurisdictions.

Part 2: Regimes of internet intermediary liability: the European and the US approach

3 The US Approach

Introduction

As discussed in Chapter 1, providers are defined by the service they offer. Providers provide access to information by hosting, transmitting, and indexing this information.[591] Without these providers, spreading and accessing information would become more challenging. These providers allow users to encounter new information.[592] Because providers enabled users to share whatever they like to whomever they like – without any gatekeeping – they were seen as a democratising force.[593]

As discussed in Chapter 1, these providers were even granted an exception for liability from user-provided information. In the US, this exception was a response to the *Stratton Oakmont* ruling,[594] which exposed providers to publisher liability whenever they conducted any editorial control over the content of user-provided information.[595] The hearth of the US liability regime forms Section 230 codified in the Communications Decency Act of 1996, which offers immunity for liability from the content of user-provided information while also providing a safe harbour for liability from moderation decisions regarding this information.[596]

This chapter thus discusses the liability regime established by Section 230. The central question in this chapter is how the liability regime provided by Section 230 in the US protects providers from liability for (moderating) the content of user-provided information.

The first part of this chapter sees the liability regime offered by Section 230. What providers can rely on Section 230? Furthermore, for what categories of content or conduct can Section 230 shield these providers? Moreover, what does Section 230 encourages when it comes to the moderation of user-provided information with illegal content? The answer to these questions contributes to understanding how the US liability regime may incentivise over- or underregulation of user-provided information, which will be discussed in Chapter 5. However, it must be clarified what can be attributed to Section 230 and what must be attributed to other legislation. In the Section 230 debate in the US, legal scholars weigh whether providers rely on Section 230 or (ultimately) on the First Amendment to the United States Constitution, which prohibits a wide range of governmental interference in exercising freedom of expression rights.[597] Therefore, the relationship between Section 230 and the First Amendment is discussed.

---

591 Perset, 2010, 'The Economic and Social Role of Internet Intermediaries', p. 9.

592 Van Eecke, 'Online service providers and liability: A plea for a balanced approach', *Common Market Law Review*, 2011, p. 1455.

593 E. Morozov, *The Net Delusion: How Not to Liberate The World*, London, Penguin Books, 2012, p. 4; Gillespie, 2018, *Custodians of the Internet*, p. 25.

594 *Stratton Oakmont, Inc. v. Prodigy Services Co.*, 23 Media L Rep 1794 (N.Y. Sup. Ct. 1995).

595 Kosseff, 2019, *The Twenty-Six Words That Created the Internet*, p. 64; G. Fishback, 'How the Wolf of Wall Street Shaped the Internet: A Review of Section 230 of the Communications Decency Act', *Texas Intellectual Property Law Journal*, Vol. 28, No. 2, 2020, pp. 283-284; Citron & Wittes, 'The Internet Will Not Break: Denying Bad Samaritans § 230 Immunity', *Fordham Law Review*, 2017, pp. 404-406; Klonick, 'The New Governors: The People, Rules, and Processes Governing Online Speech', *Harvard Law Review*, 2018, p. 1605; Par. 4.09 of Wilman, 2020, *The Responsibility of Online Intermediaries for Illegal User Content in the EU and the US*, pp. 100-101.

596 47 USCA § 230(c)(1) and (2) (West 2018, Westlaw Next through PL 116-91).

597 For example, Goldman, 'Why Section 230 Is Better Than the First Amendment', *Notre Dame Law Review*, 2019, pp. 40-42.

As noted in Chapter 1, these providers are increasingly held (morally) responsible for the harmful content of user-provided information they allow to be spread.[598] Besides, there are allegations of political bias.[599] As discussed in this chapter, these concerns relate to numerous proposals to amend Section 230 to hold providers responsible for not interfering in illegal or unlawful content of user-provided information or alleged interference in the political debate by favouring some political speech over others.

3.1 Section 230

Goldman refers to Section 230 codified in the CDA in 1996[600] as an "exceptionalist statute" that "treats the internet differently than other media."[601] As discussed here, it is not hard to see why "exceptionalist" is an accurate terminology. Both how providers are handled and the (broad) immunities offered to them are exceptional. Section 230 even shields providers from civil liability for the content of user-provided information originating from others when the provider is aware of the illegal (nature of the) content and even when the provider is notified that the content in question may lead to (serious) harm.[602]

Section 230 has two features in terms of provided immunity that will be discussed here. At first, Section 230(c)(1) states that

> [n]o provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider[603]

Section 230(c)(1) functions as a "legal shield" for providers for (potential) liability from the content of user-provided information by offering comprehensive protection.[604]

The second feature of Section 230 discussed here is that Section 230(c)(2) offers some protection to providers for liability arising from moderation. Section 230(c)(2) reads that providers are not to be held liable for:

> (A) any action voluntarily taken in good faith to restrict access to or availability of material that the provider or user considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally protected; or

---

598 For example, Anti-Defamation League, 'Stop Hate for Profit', *Anti-Defamation League,* 16 June 2021, available at stophateforprofit.org (retrieved on 14 February 2022).

599 For example, US Department of Justice, 2020, 'Department of Justice's Review of Section 230 of the Communications Decency Act of 1996'.

600 47 USCA § 230 (West 2018, Westlaw Next through PL 116-91).

601 Goldman, 2020, 'An Overview of the United States' Section 230 Internet Immunity', p. 162.

602 Kosseff, 2019, *The Twenty-Six Words That Created the Internet*, p. 5.

603 47 USCA § 230(c)(1) (West 2018, Westlaw Next through PL 116-91).

604 See, for example, D. Wakabayashi, 'Legal Shield for Social Media Is Targeted by Lawmakers', *The New York Times,* 15 December 2020, available at nytimes.com/2020/05/28/business/section-230-internet-speech.html (retrieved on 2 February 2022); L. Nylen, B.W. Swan & C. Lima, 'DOJ proposes crackdown on tech industry's legal shield', *Politico,* 17 June 2020, available at politico.com/news/2020/06/17/doj-crackdown-tech-industry-legal-shield-325594 (retrieved on 15 February 2022).

> (B) any action taken to enable or make available to information content providers or others the technical means to restrict access to material described in paragraph (1) [subparagraph (A), MK].[605]

Section 230(c)(2)(A) thus sets out the categories of the content of user-provided information that the provider can restrict in "good faith". While these categories are broadly defined ("or otherwise objectionable"), providers have to demonstrate that restrictions are taken in "good faith", which conditional nature, as shown in the following paragraphs, makes Section 23(c)(2)(A) less attractive than (c)(1). Section 230(c)(2)(B) sees to offering services to filter the categories of content that are defined in Section 230(c)(2)(A).[606]

Section 230(c) is captioned "Protection for 'Good Samaritan' blocking and screening of offensive material",[607] which forms one of the points for discussion about what Section 230 aims to protect. Is Section 230 only meant to shield providers that act responsibly regarding the content of the information shared by their users?[608] Does Section 230 require (political) neutrality with respect to content moderation?[609] Or is Section 230(c) meant to provide immunity regardless of whether or how the provider moderates?[610] As discussed in Paragraph 3.1.2, the US courts adopted the latter perspective, which sparked political and academic debates over the desirability of such extensive Section 230 protections.

In US debates on Section 230, there is much confusion in which fact and fiction are difficult to separate.[611] As Goldman observes, the approach by Section 230 "is simple and elegant, but is hardly intuitive, and it has had extraordinary consequences for the internet and our society."[612] The outcome of Section 230, in particular, is incomprehensible to many users and contrary to their intuition. Section 230, in an international context, is also exceptional. Especially in comparison with the approach provided by the EU and ECHR,[613] which are discussed in Chapter 4, Section 230 stands out as exceptional in both its protections and broad applicability.

---

605 Paragraph (1) should read subparagraph (A), see 47 USCA § 230(c)(2) (West 2018, Westlaw Next through PL 116-91).

606 Goldman, 2020, 'An Overview of the United States' Section 230 Internet Immunity', p. 160.

607 47 USCA § 230(c) (West 2018, Westlaw Next through PL 116-91). This caption may play a role in the interpretation of (c)(1) and (c)(2). This is for example the case when the interpretation of (c)(1) and (c)(2) produce conflicting results, see *Doe v. GTE Corp.*, 347 F.3d 655, 660 (7th Cir. 2003); *Chicago Lawyers' Committee for Civil Rights Under Law, Inc. v. Craigslist, Inc.*, 519 F.3d 666, 669-670 (7th Cir. 2008); *Fair Housing Council of San Fernando Valley v. Roommates.com, LLC*, 521 F.3d 1157, 1163-1164 (9th Cir. 2008).

608 Citron & Wittes, 'The Internet Will Not Break: Denying Bad Samaritans § 230 Immunity', *Fordham Law Review*, 2017, pp. 407-408.

609 Harmon, 2018, 'No, Section 230 Does Not Require Platforms to Be "Neutral"'.

610 Kosseff, 2019, *The Twenty-Six Words That Created the Internet*, p. 2; Gillespie, 2018, *Custodians of the Internet*, pp. 30-31.

611 See, for an overview, M. Masnick, 'Hello! You've Been Referred Here Because You're Wrong About Section 230 Of The Communications Decency Act', *techdirt,* 23 June 2020, available at techdirt.com/2020/06/23/hello-youve-been-referred-here-because-youre-wrong-about-section-230-communications-decency-act (retrieved on 11 July 2022).

612 Goldman, 2020, 'An Overview of the United States' Section 230 Internet Immunity', p. 155.

613 As observed by, for example, Kosseff, 2019, *The Twenty-Six Words That Created the Internet*, pp. 152-153 and 159; F. Stjernfelt & A.M. Lauritzen, *Your Post has been Removed: Tech Giants and Freedom of Speech*, Cham, Springer, 2019, doi:10.1007/978-3-030-25968-6, p. 169; Jones, 'Silencing Bad Bots: Global, Legal and Political Questions for Mean Machine Communication', *Communication Law and Policy*, 2018, p. 180.

3.1.1 What providers are protected under Section 230?

As noted, Section 230 was a direct reaction to the possibility that courts could hold providers liable for the content of user-provided information as if the provider was a publisher or a speaker of this content itself.[614] In *Stratton Oakmont, Inc. v. Prodigy Services Co.* (1995),[615] the New York Supreme Court[616] argued that Prodigy, an online messaging board (comparable to an internet forum), should be considered the publisher of the defamatory comments posted by users. The judge argued that Prodigy presented itself as an administrator exercising editorial control. Besides, Prodigy removed and filtered some information based on its content. Since Prodigy exercised some editorial control over the content of some information, Prodigy was made responsible for all the content of user-provided information on its service.[617] The New York Supreme Court argued:

> That such control is not complete and is enforced both as early as the notes arrive and as late as a complaint is made, does not minimize or eviscerate the simple fact that PRODIGY has uniquely arrogated to itself the role of determining what is proper for its members to post and read on its bulletin boards.[618]

The New York Supreme Court also discussed the distinction between publishers and distributors. The judge concluded that Prodigy must be regarded as a "publisher rather than a distributor."[619] As Wilman notes, a publisher is exposed to the same legal liability as the author of the information because it can exercise editorial control over its content.[620] By concluding that Prodigy is a *publisher* for the content of information posted by its users, the door was opened to civil liability for all user-provided information. Under the Prodigy ruling, providers could be held liable as a publisher for user-provided information. Service providers that refrain from moderation could escape the stricter publisher liability and could only be held liable as a distributor. Distributor liability for the content of user-provided information, as the United States District Court for the Southern District of New York ruled in *Cubby v. CompuServe* (1991), can only be established when the plaintiff demonstrates that the provider knows or should know of this content on its service.[621] The difference with *Prodigy* was that in *Cubby*, CompuServe did not exercise editorial control over what was uploaded to its services by others. As the District Court argued:

> CompuServe has no more editorial control over such a publication than does a public library, book store, or newsstand, and it would be no more feasible for CompuServe to examine every

---

[614] Kosseff, 2019, *The Twenty-Six Words That Created the Internet*, p. 64; Fishback, 'How the Wolf of Wall Street Shaped the Internet: A Review of Section 230 of the Communications Decency Act', *Texas Intellectual Property Law Journal*, 2020, pp. 283-284; Citron & Wittes, 'The Internet Will Not Break: Denying Bad Samaritans § 230 Immunity', *Fordham Law Review*, 2017, pp. 404-406; Klonick, 'The New Governors: The People, Rules, and Processes Governing Online Speech', *Harvard Law Review*, 2018, p. 1605; Par. 4.08-4.09 of Wilman, 2020, *The Responsibility of Online Intermediaries for Illegal User Content in the EU and the US*, pp. 100-101.

[615] *Stratton Oakmont, Inc. v. Prodigy Services Co.*, 23 Media L Rep 1794 (N.Y. Sup. Ct. 1995).

[616] Unlike the name might suggest, The New York Supreme Court is a trial court and not the highest appellate court in the State of New York.

[617] Fishback, 'How the Wolf of Wall Street Shaped the Internet: A Review of Section 230 of the Communications Decency Act', *Texas Intellectual Property Law Journal*, 2020, p. 282.

[618] *Stratton Oakmont, Inc. v. Prodigy Services Co.*, 23 Media L Rep 1794 (N.Y. Sup. Ct. 1995).

[619] *Stratton Oakmont, Inc. v. Prodigy Services Co.*, 23 Media L Rep 1794 (N.Y. Sup. Ct. 1995).

[620] Par. 4.03 of Wilman, 2020, *The Responsibility of Online Intermediaries for Illegal User Content in the EU and the US*, p. 98.

[621] *Cubby, Inc. v. CompuServe, Inc.*, 776 F.Supp. 135, 139-141 (S.D. New York 1991).

> publication it carries for potentially defamatory statements than it would be for any other distributor to do so.[622]

Publisher liability thus exposes the provider to liability for all the content on its service (independent of knowledge) because the provider exercises editorial control. In contrast, distributor liability is contingent on some scienter of the provider of the existence of illegal or unlawful content.

Of course, the ruling not merely affected Prodigy but a far more extensive range of providers. According to Kosseff, any involvement with user content introduced the risk to the provider of becoming liable for user-provided information as a publisher under the *Stratton Oakmont* ruling.[623] As Klonick argues, *Stratton Oakmont* and *Cubby* "seemed to expose intermediaries to a wide and unpredictable range of tort liability if they exercised any editorial discretion over content posted on their sites."[624] All providers that moderate user-provided information could be exposed to publisher liability. The *Stratton Oakmont* ruling thus meant that providers that engage in editorial control are exposed to liability for the content of all user-provided information. In and outside the US, legal scholars are critical about what an internet under the *Stratton Oakmont* standard would mean. Holding providers liable as a publisher would cause providers to refrain from (voluntary) moderation, may cause providers to withdraw their services, and – ultimately – less freedom of expression.[625] Service providers that did engage in moderation could be held liable as a publisher because they exercised some editorial control over the content of user-provided information.

As discussed under Paragraph 3.1.3, the drafters of Section 230 wished to enable providers to self-regulate by moderating the content of user-provided information on their services.[626] Their proposal for Section 230 thus directly aimed to overturn the *Stratton Oakmont* ruling.[627] Without exempting providers from liability arising from moderation, the barrier for providers to engage in self-regulation is too high.[628] Section 230(c)(1) took down this barrier by shielding providers from liability as a speaker or publisher of the content of user-provided information.[629] Section 230 applies to every "interactive computer service"[630] dealing with third-party information.[631] Note that

---

[622] *Cubby, Inc. v. CompuServe, Inc.*, 776 F.Supp. 135, 140 (S.D. New York 1991).

[623] Kosseff, 2019, *The Twenty-Six Words That Created the Internet*, p. 56.

[624] Klonick, 'The New Governors: The People, Rules, and Processes Governing Online Speech', *Harvard Law Review*, 2018, p. 1605.

[625] Klonick, 'The New Governors: The People, Rules, and Processes Governing Online Speech', *Harvard Law Review*, 2018, p. 1605; Par. 4.10 of Wilman, 2020, *The Responsibility of Online Intermediaries for Illegal User Content in the EU and the US*, p. 101.

[626] Citron & Wittes, 'The Internet Will Not Break: Denying Bad Samaritans § 230 Immunity', *Fordham Law Review*, 2017, p. 405.

[627] Par. 4.10 of Wilman, 2020, *The Responsibility of Online Intermediaries for Illegal User Content in the EU and the US*, p. 101.

[628] Perry & Zarsky, 'Who Should Be Liable for Online Anonymous Defamation?', *University of Chicago Law Review Dialogue*, 2015.

[629] 47 USCA § 230(c)(1) (West 2018, Westlaw Next through PL 116-91).

[630] Section 230 defines "interactive computer service" as "any information service, system, or access software provider that provides or enables computer access by multiple users to a computer server, including specifically a service or system that provides access to the Internet and such systems operated or services offered by libraries or educational institutions.", see 47 USCA § 230(f)(2) (West 2018, Westlaw Next through PL 116-91).

[631] Gillespie, 2018, *Custodians of the Internet*, p. 33.

both the provider and the user are exempted from liability for the content of third-party information.[632] According to Goldman, the courts extended this exemption from immunity to diverse content categories with only a few exceptions codified in Section 230.[633] I discuss these exceptions in Paragraph 3.1.3. However, the general rule is that providers in the US under Section 230(c)(1) cannot be held legally liable – as publisher or speaker – for the content of user-provided information.[634]

Section 230 aimed to protect active providers who exercise editorial discretion while not forcing passive providers to engage in moderation. Therefore, Section 230 provides protections to a broad range of providers. Section 230 does not only apply to providers that offer a service without any knowledge or involvement with the content of user-provided information. In one of the first Section 230 cases in which plaintiffs sought to hold AOL (a service provider) liable for the defamatory content of user-provided information on its service, the United States District Court of the District of Columbia argued that

> it would seem only fair to hold AOL to the liability standards applied to a publisher or, at least, like a book store owner or library, to the liability standards applied to a distributor. But Congress has made a different policy choice by providing immunity even where the interactive service provider has an active, even aggressive role in making available content prepared by others.[635]

While it may be counterintuitive for plaintiffs, Section 230 not merely shields passive providers but also providers that are active in varying degrees. As the wording of Section 230 suggests, its immunities do not apply to providers that are active in such a manner that they contribute to the development of the content of the provided information themselves. Such conduct would cause providers to forfeit Section 230 protections since this would cause the provider to become a content provider itself.[636] As discussed in the following paragraphs, the line between these is sometimes difficult to draw. After all, assuming too quickly that providers are responsible as a developer for the content of user-provided information would render Section 230 protections useless.

3.1.2 For what does Section 230 protect providers?

As noted in the previous paragraph, Section 230 offers a broad and (almost) unconditional immunity for providers for liability arising from the content of user-provided information. However, Section 230, regarding the question of what is protected, is hardly as clear and elegant as some legal scholars wish it to be.[637] Courts especially offer little clarity between the two components of Section 230. Section 230(c)(1), as Wilman puts it, protects providers from under-filtering (a hands-off approach), while Section 230(c)(2) protects providers from over-filtering (a hands-on approach).[638] I first discuss Section 230(c)(1) and then turn to Section 230(c)(2). The focus, to recall, lies on liability for user-provided information or interventions on its content.

---

[632] *Barrett v. Rosenthal*, 146 P.3d 510, 526-529 (Cal. S.C. 2006); Par. 4.18 of Wilman, 2020, *The Responsibility of Online Intermediaries for Illegal User Content in the EU and the US*, pp. 104-105.

[633] Goldman, 2020, 'An Overview of the United States' Section 230 Internet Immunity', p. 158.

[634] Goldman, 2020, 'An Overview of the United States' Section 230 Internet Immunity', p. 159.

[635] *Blumenthal v. Drudge*, 992 F.Supp. 44, 51-52 (D.D.C. 1998).

[636] 47 USCA § 230(f)(3) (West 2018, Westlaw Next through PL 116-91).

[637] For example, Goldman, 2020, 'An Overview of the United States' Section 230 Internet Immunity', p. 155.

[638] Par. 4.16-4.17 of Wilman, 2020, *The Responsibility of Online Intermediaries for Illegal User Content in the EU and the US*, p. 104.

Section 230(c)(1)

As noted, Section 230(c)(1) reads:

> (c) Protection for "Good Samaritan" blocking and screening of offensive material
>
> (1) Treatment of publisher or speaker
>
> No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider.[639]

According to Goldman, three distinct criteria must be fulfilled before providers[640] can rely on the immunities provided by Section 230(c)(1). First, the provider must qualify as a user or operator of an interactive computer service.[641] As Goldman notes, "in practice […] everyone online should satisfy this first element."[642] As set out in the previous paragraph, Section 230 shields users and providers from liability for third-party content that they did not develop themselves. The second criterion is that the plaintiff seeks to hold the provider responsible as a "publisher or speaker". As Goldman notes, this requirement is interpreted broadly by the courts. The courts include all causes of action that rely on holding the provider responsible for the content of user-provided information. When the user seeks to hold the provider responsible for user-provided information while using terminology such as "editor", "publisher", "distributor", or "speaker", the chances are very high that the judge will side with the provider and dismiss the claim.[643] Of course, as long as the content of the information originates from someone other than the service provider. The third criterion is this: the provider cannot rely on Section 230 when it qualifies as the information content provider. Service providers cannot rely on Section 230 when they are "responsible, in whole or in part, for the creation or development of information provided"[644] since that would render the service provider an information content provider itself. Section 230 protections only apply to a "provider or user of an interactive computer service" for liability as a publisher or speaker for "information provided by another information content provider". Providers that develop the content of the information itself thus cannot rely on the immunity provided by Section 230(c)(1). As a rule of thumb, Goldman distinguishes between information containing content authored by employees of the provider and the content of user-provided information. Section 230(c)(1) does not apply to the first, while Section 230(c)(1) shields from liability for the latter.[645]

Section 230(c)(1) renders knowledge of information with illegal or unlawful content meaningless.[646] The immunities provided by Section 230 also extend to providers aware of user-provided information's illegal or unlawful content – even when they are notified of its existence.[647] In *Zeran v. America Online, Inc.* (1997), the Third Circuit argued that a notice-based liability system

---

639 47 USCA § 230(c)(1) (West 2018, Westlaw Next through PL 116-91).

640 Or users, but the focus in this dissertation lies on the service providers.

641 47 USCA § 230(f)(2) (West 2018, Westlaw Next through PL 116-91).

642 Goldman, 2020, 'An Overview of the United States' Section 230 Internet Immunity', p. 158.

643 *Barnes v. Yahoo!, Inc.*, 570 F.3d 1096, 1103-1104 (9th Cir. 2009); *Batzel v. Smith*, 333 F.3d 1018, 1027 (9th Cir. 2003); *Zeran v. America Online, Inc.*, 129 F.3d 327, 331-334 (3rd Cir. 1997); Kosseff, 2019, *The Twenty-Six Words That Created the Internet*, p. 93.

644 47 USCA § 230(f)(3) (West 2018, Westlaw Next through PL 116-91).

645 Goldman, 2020, 'An Overview of the United States' Section 230 Internet Immunity', pp. 158-159.

646 Goldman, 'Why Section 230 Is Better Than the First Amendment', *Notre Dame Law Review*, 2019, p. 38.

647 Kosseff, 2019, *The Twenty-Six Words That Created the Internet*, pp. 93-95.

requiring providers to review notices would "defeat the dual purposes advanced by § 230 of the CDA".[648] The court argued that "liability upon notice reinforces service providers' incentives to restrict speech and abstain from self-regulation".[649] The Third Circuit also mentioned that providers receiving a notification would be incentivised to remove the user-provided information in question by default to prevent liability. According to the Third Circuit, such a system would have "a chilling effect on the freedom of Internet speech."[650] Notice-based liability, thus, would not be compatible with Section 230(c)(1). Such a liability regime imposes responsibilities on a provider as publisher for the content of user-provided information, which is incompatible with the wording and the purpose of Section 230.

Even the most far-reaching real-world harms may not break through the wall of immunity provided by Section 230. For example, "negligence and gross negligence claims are barred by the CDA, which prohibits claims against Web-based interactive computer services based on their publication of third-party content." [651] A provider that does not verify the age of its users while social media functionalities cannot be held liable for negligence or gross negligence when the underage user meets with another user of the service and becomes the victim of sexual assault as a result of this meeting.[652] The other way around, an adult who uses a service to arrange a sex date with someone thought to be an adult cannot sue the provider when the sex date turns out to be underage.[653]

In *Doe v. GTE Corp.* (2009), the Seventh Circuit argued that "Congress is free to oblige web hosts to withhold services from criminals (to the extent legally required screening for content may be consistent with the first amendment)".[654] Under current law, however, "a web host cannot be classified as an aider and abettor of criminal activities conducted through access to the Internet."[655] Plaintiffs sued GTE, the hosting service provider, because it hosted videos showing undressed college athletes. The videos in which the plaintiffs appeared were secretly taped without their consent.[656] Section 230(c)(1) bars such a claim because the hosting service provider is held liable as a publisher.[657] Ultimately, the plaintiffs seek to hold GTE liable for negligence, but they "do not cite any case in any jurisdiction holding that a service provider must take reasonable care to prevent injury to third parties."[658]

In a different case, Section 230 immunities did not apply. In *Doe v. Internet Brands,* Doe sued Internet Brands because the company, according to Doe, knew that there were rapists active on one of its model-scouting websites but failed to warn Doe and other users of their modus operandi.

---

[648] *Zeran v. America Online, Inc.*, 129 F.3d 327, 333 (3rd Cir. 1997).

[649] *Zeran v. America Online, Inc.*, 129 F.3d 327, 333 (3rd Cir. 1997).

[650] *Zeran v. America Online, Inc.*, 129 F.3d 327, 333 (3rd Cir. 1997); This is a recurring argument against the introduction of all kinds of liability of Internet intermediaries for third-party content. An overview of empirical studies regarding the overremoval of content by internet intermediaries can be found at Keller, 2020, 'Empirical Evidence of "Over-Removal" by Internet Companies Under Intermediary Liability Laws'.

[651] *Doe v. MySpace, Inc.*, 528 F.3d 413, 422 (5th Cir. 2008).

[652] *Doe v. MySpace, Inc.*, 528 F.3d 413, 420-422 (5th Cir. 2008).

[653] *Doe v. SexSearch.com*, 502 F.Supp.2d 719, 727-728 (N.D.Ohio 2007); *Doe v. SexSearch.com*, 551 F.3d 412 (6th Cir. 2008).

[654] *Doe v. GTE Corp.*, 347 F.3d 655, 659 (7th Cir. 2003).

[655] *Doe v. GTE Corp.*, 347 F.3d 655, 659 (7th Cir. 2003).

[656] *Doe v. GTE Corp.*, 347 F.3d 655, 656 (7th Cir. 2003).

[657] *Doe v. GTE Corp.*, 347 F.3d 655, 660 (7th Cir. 2003).

[658] *Doe v. GTE Corp.*, 347 F.3d 655, 661 (7th Cir. 2003).

Doe argued before the court that because of this negligent failure to warn, Doe was drugged and raped.[659] The Ninth Circuit concluded that Section 230 does not immunises Internet Brands for negligent failure to warn because "Jane Doe's negligent failure to warn claim does not seek to hold Internet Brands liable as the 'publisher or speaker of any information provided by another information content provider.'"[660]

As noted, Section 230(c)(1) excludes the development of the content of information by the provider from the immunity provided by this section. The question then is when a provider becomes a developer of the content of user-provided information. In case law, Section 230 is interpreted to protect a wide range of activities. Even editing the content of user-provided information does not cause the internet intermediary provider to forfeit its immunity under Section 230. The Ninth Circuit clarified in *Batzel v. Smith* (2003) when a provider could be held liable as an internet content provider. According to the Ninth Circuit, the provider its involvement, before it can be considered to be involved in the "development of information",[661] must be "more substantial than merely editing portions of an email and selecting material for publication."[662] As the Ninth Circuit puts it

> the exclusion of "publisher" liability necessarily precludes liability for exercising the usual prerogative of publishers to choose among proffered material and to edit the material published while retaining its basic form and message.[663]

Section 230 offers providers that engage in publisher activities regarding user-provided information a shield from liability as a publisher. A dating website, thus, is not liable for the content of the information provided by its users – even when a user chooses to submit the personalia of someone else without this person's consent.[664] When providers engage in publisher activities, this does not mean that they directly forfeit Section 230 immunities. Plaintiffs, thus, are required to base a claim on something different than holding the provider liable as a publisher or speaker. Another option is to demonstrate that a provider is "materially contributing"[665] to the illegal or unlawful content of user-provided information, which renders the provider a developer of this content itself.

In *Fair Housing Council of San Fernando Valley v. Roommates.com* (2008), the Ninth Circuit denied Section 230(c)(1) immunity because the provider was regarded to have developed the

---

[659] *Doe v. Internet Brands, Inc.*, 824 F.3d 846, 848-851 (9th Cir. 2016).

[660] *Doe v. Internet Brands, Inc.*, 824 F.3d 846, 851 (9th Cir. 2016).

[661] 47 USCA § 230(f)(3) (West 2018, Westlaw Next through PL 116-91).

[662] *Batzel v. Smith*, 333 F.3d 1018, 1031 (9th Cir. 2003). Merely deleting some symbols or other portions of content does not render the provider a developer of the content, see *Ben Ezra, Weinstein, and Company, Inc. v. America Online Inc.*, 206 F.3d 980, 985-986 (10th Cir. 2000).

[663] *Batzel v. Smith*, 333 F.3d 1018, 1031 (9th Cir. 2003).

[664] In *Carafano v. Metrosplash.com* the 9th Cir. concluded that "despite the serious and utterly deplorable consequences that occurred in this case" that the provider "did not play a significant role in creating, developing or 'transforming' the relevant information." see *Carafano v. Metrosplash.com, Inc.*, 339 F.3d 1119, 1125 (9th Cir. 2003). The difference with *Roommates* is that in *Carafano* "the selection of the content was left exclusively to the user. The actual profile 'information' consisted of the particular options chosen and the additional essay answers provided." see *Carafano v. Metrosplash.com, Inc.*, 339 F.3d 1119, 1124 (9th Cir. 2003).

[665] *Fair Housing Council of San Fernando Valley v. Roommates.com, LLC*, 521 F.3d 1157, 1168 (9th Cir. 2008).

content of the user-provided information.[666] Roommates.com allowed users to – as the name of the website says – find roommates. The service asked a few questions about the user's preferences to help the user find a suitable roommate. These questions were mandatory. The user only could answer these questions by choosing from predefined options. Questions were about sex, sexual orientation, whether the subscriber has children and the sexual orientation of those living in the house. Housing seekers had to specify whether they were willing to live with children or people with a given sexual orientation. They could also exclude living with males or females from the results.[667] Some of these questions were argued to violate federal or state laws against housing discrimination. As the Ninth Circuit notes, "asking questions certainly can violate the Fair Housing Act and analogous laws in the physical world".[668] The Ninth Circuit did not evaluate whether the questions were indeed illegal but only whether Section 230(c)(1) immunities would shield Roommates.com from liability for the content on its websites generated from these questions.[669] According to the Ninth Circuit, Section 230(c)(1) did not apply to Roommates.com conduct now

> By any reasonable use of the English language, Roommate is "responsible" at least "in part" for each subscriber's profile page, because every such page is a collaborative effort between Roommate and the subscriber.[670]

Because Roommates.com required users to submit their preferences by answering questions from a predefined list of options, Roommates.com is at least in part responsible for the profile pages that are generated based on these questions.[671] The Ninth Circuit clarifies that the usage of dropdown menus is not the problem:

> A dating website that requires users to enter their sex, race, religion and marital status through drop-down menus, and that provides means for users to search along the same lines, retains its CDA immunity insofar as it does not contribute to any alleged illegality[672]

Roommates can be compared with *Chicago Lawyers' Committee for Civil Rights Under Law v. Craigslist* (2008), in which the Seventh Circuit granted Craigslist Section 230 immunity. According to the Seventh Circuit, Craigslist did not "induces anyone to post any particular listing or express a preference for discrimination".[673] As Roommates.com was immunised for the content users entered in a free-text field captioned with "Additional Comments",[674] the Seventh Circuit concluded that Section 230(c)(1) shielded Craigslist from liability for (potential) discriminatory housing advertisements.[675] While Roommates.com, in part, was considered a developer of the content of the information provided by its users. Craigslist that did not use mandatory dropdown menus with predefined options was not.

---

[666] *Fair Housing Council of San Fernando Valley v. Roommates.com, LLC*, 521 F.3d 1157, 1166 (9th Cir. 2008); 47 USCA § 230(c)(1) and (f)(3) (West 2018, Westlaw Next through PL 116-91).
[667] *Fair Housing Council of San Fernando Valley v. Roommates.com, LLC*, 521 F.3d 1157, 1165 (9th Cir. 2008).
[668] *Fair Housing Council of San Fernando Valley v. Roommates.com, LLC*, 521 F.3d 1157, 1164 (9th Cir. 2008).
[669] *Fair Housing Council of San Fernando Valley v. Roommates.com, LLC*, 521 F.3d 1157, 1164 (9th Cir. 2008).
[670] *Fair Housing Council of San Fernando Valley v. Roommates.com, LLC*, 521 F.3d 1157, 1167 (9th Cir. 2008).
[671] *Fair Housing Council of San Fernando Valley v. Roommates.com, LLC*, 521 F.3d 1157, 1165-1167 (9th Cir. 2008).
[672] *Fair Housing Council of San Fernando Valley v. Roommates.com, LLC*, 521 F.3d 1157, 1169 (9th Cir. 2008).
[673] *Chicago Lawyers' Committee for Civil Rights Under Law, Inc. v. Craigslist, Inc.*, 519 F.3d 666, 671 (7th Cir. 2008).
[674] *Fair Housing Council of San Fernando Valley v. Roommates.com, LLC*, 521 F.3d 1157, 1172-1173 (9th Cir. 2008).
[675] *Chicago Lawyers' Committee for Civil Rights Under Law, Inc. v. Craigslist, Inc.*, 519 F.3d 666, 671-672 (7th Cir. 2008).

Even when providers require users to provide information that can discriminate between others or for illegal activities, this does not necessarily render the provider a developer. For example, users can discriminate when searching for a date based on their preferences. These preferences and the users' personalia could be used to discriminate or for other illegal purposes. However, the dating website provider cannot be said to contribute to such illegality by merely asking these questions.[676]

The Ninth Circuit in *Roommates* provided some insight into when services may cross the line in contributing to illegality that causes them to lose Section 230(c)(1) immunity. For example, providers steering users to use discriminatory criteria in search or notification systems may not count on Section 230(c)(1) immunity.[677] Of course, this does not expose all search engines that can be (potentially) used to find information with illegal content to liability. A provider is not responsible for the content offered by others when it "comes entirely from subscribers and is passively displayed".[678] The United States District Court for the Northern District of California in *Goddard v. Google* (2009) revisited *Roommates*. The question was whether Google could be held liable for the content of fraudulent advertisements. These advertisements (so-called AdWords) are placed beside search results. According to the plaintiff, Google actively suggests the advertisement's content. For example, when an advertiser creates an advertisement concerning ringtones, the word "free" is suggested. This suggestion, according to the plaintiff, could contribute to the fraudulent nature of the advertisement. According to the plaintiff, users are often charged for so-called "free" ringtones. According to the plaintiff, this renders Google a developer of the content of the advertisement.[679] According to the District Court, this argument does not hold up. According to the District Court, Google "does nothing more than provide options that advertisers may adopt or reject at their discretion."[680] The District Court thus suggested a narrow reading of *Roommates*. Service providers are not quickly rendered a developer of the content of user-provided information. While denying Section 230 immunities does not automatically mean that the provider is indeed held liable, the mere fact that the court case continues may be costly for providers.[681]

Besides the fact that providers are not liable for the content of user-provided information that is offered to them, a provider is also not responsible for how users use search and filter options. However, as the *Roommates* case clarified, as long as the provider does "not use unlawful criteria to limit the scope of the searches conducted on them, nor are […] designed to achieve illegal ends".[682] For example, Section 230(c)(1) bars the enforcement of state law seeking to hold Craigslist liable for offering a "word search" functionality on its service. This search functions as a neutral tool for users to search for content on the service but can also be used to find (illegal) advertisements for prostitution.[683] In other words, the fact that it is possible to search for illegal content does not render the search engine provider a developer of the list with results – even when the results are built from an index of the search provider. However, when the provider steers users

---

676 *Fair Housing Council of San Fernando Valley v. Roommates.com, LLC*, 521 F.3d 1157, 1169 (9th Cir. 2008).
677 *Fair Housing Council of San Fernando Valley v. Roommates.com, LLC*, 521 F.3d 1157, 1167 (9th Cir. 2008).
678 *Fair Housing Council of San Fernando Valley v. Roommates.com, LLC*, 521 F.3d 1157, 1174 (9th Cir. 2008).
679 *Goddard v. Google, Inc.*, 640 F.Supp.2d 1193, 1197 (N.D.Cal. 2009).
680 *Goddard v. Google, Inc.*, 640 F.Supp.2d 1193 (N.D.Cal. 2009).
681 *Goddard v. Google, Inc.*, 640 F.Supp.2d 1193, 1202 (N.D.Cal. 2009).
682 *Fair Housing Council of San Fernando Valley v. Roommates.com, LLC*, 521 F.3d 1157, 1167 (9th Cir. 2008).
683 *Dart v. Craigslist, Inc.*, 655 F.Supp.2d 961, 969 (N.D.Ill. 2009).

to illegal search queries or filtering options predefined by the search provider, based on illegal content the provider actively requested, the provider loses its Section 230 immunity. A search provider, at least, is "responsible" for the "development" of the illegal content of user-provided information when it actively solicits for this content, for example, by paying for illegally obtained documents.[684] The Ninth Circuit concluded: "The message to website operators is clear: If you don't encourage illegal content, or design your website to require users to input illegal content, you will be immune."[685]

Section 230(c)(1) protections, however, do apply to providers that do not undertake action against user-provided information that qualifies as terrorist content. In *Fields v. Twitter* (2016), the United States District Court for the Northern District of California argued that claims seeking to hold Twitter liable for terrorist content on its service are based on Twitter its status as a publisher of the content of user-provided information. Section 230(c)(1) bars such claims based on the actual content that is spread through its service but also claims based on the fact that the provider does not prevent terrorists from using its service. Both claims are based on Twitter being as a publisher liable for the content of user-provided information (that may be) offered through its service.[686]

As noted, Section 230 provides immunity to providers for failing to remove user-provided information with illegal or unlawful content after a notification. However, there is an exception. As the Ninth Circuit first argued in *Barnes v. Yahoo!, Inc.* (2009), "removing content is something publishers do".[687] Therefore, a provider cannot be held liable for failing to remove user-provided information. In *Barnes*, however, Section 230(c)(1) did not immunise Yahoo! as a publisher for the content of user-provided information. The plaintiff notified Yahoo! of defamatory content, which Yahoo! promised to take down, and Yahoo! failed to live up to this promise. The Ninth Circuit argued that holding a provider liable for failing to follow up on a "promissory estoppel" is not based on a provider being "a publisher or speaker of third-party content".[688] Instead, the provider is held liable "as the counter-party to a contract, as a promisor who has breached."[689] After *Barnes,* liability based on failing to live up to a promise is a rarity. Service providers will think twice before making such promises.

However, Section 230(c)(1) does not apply to claims based on the service provider's liability for misrepresentation because of the content of information created by the provider itself. Seeking to hold a provider liable for the misrepresentation of dating profiles as active – while they no longer are – does not seek to hold the provider liable as a publisher or speaker of the content of user-provided information. Instead, the provider is held liable as an "information content provider" itself.[690] In *Anthony v. Yahoo* (2006), the United States District Court for the Northern District of California denied Section 230 immunity for misrepresentation and fraud. While another "information content provider" may have provided the content of the dating profiles, the misrepresentation of profiles as active (while they no longer are) does not fall under Section

---

[684] *F.T.C. v. Accusearch Inc.*, 570 F.3d 1187, 1197-1201 (10th Cir. 2009).

[685] *Fair Housing Council of San Fernando Valley v. Roommates.com, LLC*, 521 F.3d 1157, 1175 (9th Cir. 2008).

[686] *Fields v. Twitter, Inc.*, 200 F.Supp.3d 964, 970-974 (N.D.Cal. 2016). Materially, it is also unlikely that service providers would be held liable for "aiding and abetting" terrorists, see *Colon v. Twitter, Inc.*, 14 F.4th 1213 (11th Cir. 2021).

[687] *Barnes v. Yahoo!, Inc.*, 570 F.3d 1096, 1103 (9th Cir. 2009).

[688] *Barnes v. Yahoo!, Inc.*, 570 F.3d 1096, 1107 (9th Cir. 2009).

[689] *Barnes v. Yahoo!, Inc.*, 570 F.3d 1096, 1107 (9th Cir. 2009).

[690] *Anthony v. Yahoo Inc.*, 421 F.Supp.2d 1257, 1262-1263 (N.D.Cal. 2006).

230(c)(1) protections. Section 230(c)(1), as the District Court puts it, "does not absolve Yahoo! from liability for any accompanying misrepresentations."[691]

As noted, the *Roommates* case was about whether Section 230 immunities *would* apply to Roommates.com when the content of user-provided information which Roommates.com was argued to materially contribute to *was* illegal. While all the preceding would suggest otherwise: Roommates.com did ultimately not violate the law.[692] When a court denies Section 230 immunity, this does not automatically render the provider liable. While Roommates.com did not win a Section 230 motion to dismiss, on the merits of the case, Roommates.com did. In the process, *Roommates*, according to Goldman, left Section 230 with a common law exception. The *Roommates*-exception, however, is not the only hallow exception the Ninth Circuit added. According to Goldman, the Ninth Circuit in *Doe v. Internet Brands*[693] added an exclusion for a failure to warn. However, ultimately there was no such duty to warn for Internet Brands.[694] These exceptions may allow court cases to continue instead of stranding on Section 230 immunities but leave plaintiffs ultimately empty-handed after the case is discussed on its merits. As the saying goes: the plaintiffs did win the battle but ultimately lost the war.

In sum, Section 230(c)(1) offers broad protection to a diverse range of services. Section 230(c)(1) applies to providers that do moderate but also to providers that do not moderate. For Section 230(c)(1), the level of editorial control does not matter.[695] Service providers may, for example, select user-provided information for publication without becoming liable for its content as long as they do not materially contribute to its illegal or unlawful content.[696] Section 230(c)(1), however, only applies to a "provider or user of an interactive computer service"[697] and not to traditional media, which makes Section 230(c)(1) an exceptionalist statute.[698] Most importantly, Section 230(c)(1) immunises providers for action, holding them responsible as a publisher or speaker of the content of user-provider information.

Section 230(c)(1) makes it easy for providers to let a judge grant a motion to dismiss because claims based on publisher or speaker liability have no chance of success. Lengthy trials usually are avoided.[699]

Section 230(c)(2)

As Goldman notes, Section 230(c)(1) is far more critical for providers than Section 230(c)(2) because the wording of (c)(2) suggests that it applies to providers that act in good faith. Service providers relying on (c)(1) may quickly get the case dismissed, while relying on (c)(2) may lead to

---

691 *Anthony v. Yahoo Inc.*, 421 F.Supp.2d 1257, 1263 (N.D.Cal. 2006).

692 *Fair Housing Council of San Fernando Valley v. Roommate.com, LLC*, 666 F.3d 1216 (9th Cir. 2012).

693 *Doe v. Internet Brands, Inc.*, 824 F.3d 846 (9th Cir. 2016); E. Goldman, 'Failure-to-Warn Claim Against Match.com Fails–Beckman v. Match.com', *Technology & Marketing Law Blog,* 27 November 2018, available at blog.ericgoldman.org/archives/2018/11/failure-to-warn-claim-against-match-com-fails-beckman-v-match-com.htm (retrieved on 14 February 2022).

694 E. Goldman, 'The Ninth Circuit's Confusing Ruling Over Snapchat's Speed Filter–Lemmon v. Snap', *Technology & Marketing Law Blog,* 21 May 2021, available at blog.ericgoldman.org/archives/2021/05/the-ninth-circuits-confusing-ruling-over-snapchats-speed-filter-lemmon-v-snap.htm (retrieved on 14 February 2022).

695 Goldman, 2020, 'An Overview of the United States' Section 230 Internet Immunity', p. 159.

696 *Jones v. Dirty World Entertainment Recordings LLC*, 755 F.3d 398, 406-417 (6th Cir. 2014).

697 47 USCA § 230(c)(1) (West 2018, Westlaw Next through PL 116-91).

698 Goldman, 2020, 'An Overview of the United States' Section 230 Internet Immunity', p. 162.

699 Goldman, 2020, 'An Overview of the United States' Section 230 Internet Immunity', p. 159.

a lengthy procedure in which the provider has to demonstrate that it indeed was acting in good faith. The provider still may win, but the cost of litigation may be much higher than under Section 230(c)(1).[700] There are, however, some examples in which a motion to dismiss was granted, which causes Section 230(c)(2) to immunise the provider for liability.[701]

Section 230(c)(2) thus protects for liability arising from moderating decisions by the service provider.[702] In order to rely on the safe harbour offered by (c)(2), voluntary action has to be taken by the "provider or user of an interactive computer service" in "good faith" to "objectionable" content.[703] According to Goldman, Section 230(c)(2) added value for developers of anti-spyware software against claims of the developers of software that was (wrongfully) characterised as spyware.[704] As Goldman and Wilman note, providers typically rely on their terms of services that function as a contract between the provider and the user. In these terms of services, providers grant themselves an (almost) unlimited discretion regarding what content they allow and what not while exonerating themselves from any legal liability.[705] According to Wilman, usually, the only ones who sue providers for over-removal are users of the service. A contract governs the relationship between the user and the provider.[706] However, it is not unthinkable that a third party (which is not a user of the service) would sue for liability for the over-removal of user-provided information in which the third party appears.[707] In the case of such non-contractual liability claims, the safe harbour provided by Section 230(c)(2) could function as a backstop.

Section 230(c)(2) (together with (c)(1)) is put under pressure by lawmakers that seek to alter these provisions. These developments are discussed in Paragraph 3.3. For now, it is relevant to mention that Section 230(c)(2) is often mischaracterised as if this provision grants providers unlimited editorial discretion.[708] In part, this mischaracterised is caused by the wording of Section 230(c)(2)(A), which applies to moderation decisions "whether or not such material is constitutionally protected".[709] Would providers no longer be allowed to moderate user-provided information without Section 230(c)(2)? That, of course, is not the case. Section 230(c)(2) does not grant the right to moderate; it merely offers providers a safe harbour for liability arising from moderation. As discussed under Paragraph 3.2, editorial decisions are also protected under the

---

[700] Goldman, 'Why Section 230 Is Better Than the First Amendment', *Notre Dame Law Review*, 2019, p. 40.

[701] *Ebeid v. Facebook, Inc.*, 2019 WL 2059662 (N.D.Cal. 2019); *Lancaster v. Alphabet Inc.*, 2016 WL 3648608 (N.D.Cal. 2016); *Mezey v. Twitter, Inc.*, 2018 WL 5306769 (S.D.Fla. 2018). The applicability of Section 230(c)(1) to moderation efforts renders Section 230(c)(2) useless, see *e-ventures Worldwide, LLC v. Google, Inc.*, 2017 WL 2210029 (M.D.Fla. 2017); E. Goldman, 'Online User Account Termination and 47 U.S.C. §230(c)(2)', *UC Irvine Law Review*, Vol. 2, No. 2, 2012 (available at escholarship.org/uc/item/1hh0t3w6).

[702] Goldman, 'Online User Account Termination and 47 U.S.C. §230(c)(2)', *UC Irvine Law Review*, 2012, p. 662.

[703] 47 USCA § 230(c)(2)(A) (West 2018, Westlaw Next through PL 116-91); Goldman, 'Online User Account Termination and 47 U.S.C. §230(c)(2)', *UC Irvine Law Review*, 2012, p. 661.

[704] Goldman, 2020, 'An Overview of the United States' Section 230 Internet Immunity', p. 160.

[705] Par. 4.40 of Wilman, 2020, *The Responsibility of Online Intermediaries for Illegal User Content in the EU and the US*, p. 114; Goldman, 2020, 'An Overview of the United States' Section 230 Internet Immunity', p. 160.

[706] Par. 4.39-4.40 of Wilman, 2020, *The Responsibility of Online Intermediaries for Illegal User Content in the EU and the US*, p. 114.

[707] For example, in the Dutch case Rb. Amsterdam (vzr.), 9 September 2020, ECLI:NL:RBAMS:2020:4435 (*YouTube*).

[708] N. Hochman, 'Conservatives Should Support Section 230 Reform', *National Review,* 16 October 2021, available at nationalreview.com/2021/10/conservatives-should-support-section-230-reform (retrieved on 15 February 2022).

[709] 47 USCA § 230(c)(2)(A) (West 2018, Westlaw Next through PL 116-91).

First Amendment – even when made by providers. Without Section 230(c)(2), providers would not be required to carry all user-provided information irrespective of its content.

Statutory exceptions

In terms of the actual content of the user-provided information, the provider can only be imposed with liability when it fits in one of the few exceptions to Section 230.[710] For example, Section 230 does not stand in the way of enforcing federal criminal law. Service providers can be prosecuted for user-provided child sexual abuse material when they meet the statutory requirements.[711] Besides, Section 230 does not apply to the enforcement of intellectual property law, which means providers can also be exposed to liability for user-provided information with content infringing on intellectual property law.[712] The same is true for violations of privacy laws.[713] Privacy laws form a strange exception because it seems impossible for a provider to rely on Section 230 while violating privacy laws.[714] Next to these exceptions, Section 230 does not prohibit states from enforcing state law consistent with Section 230.[715] However, Section 230 stands in the way when state criminal law is not consistent with Section 230.[716]

While these exceptions seem broad, their legal meaning is very narrow. Goldman lists a few examples of the few federal criminal law cases involving the prosecution of providers. Examples are providers fined for allowing advertisements on their platforms that violated gambling and prescription drugs laws. Goldman also mentions the prosecution of the individual behind the internet marketplace Silk Road, known for the availability of illegal goods and services. The mind behind this marketplace was sentenced to life imprisonment.[717]

While the enforcement of federal criminal statutes is exempted from Section 230 immunities, this does not mean that the provider can no longer rely on Section 230 immunity for civil liability that may arise from the violation of the criminal law.[718] Service providers cannot be sued for civil damages by users over violations of federal criminal statutes.[719] An exception forms civil claims based on violations of sex-trafficking law that are expressly exempted from the immunities provided by Section 230.[720] Noteworthy is that this exception also applies to the enforcement of criminal state law comparable to this federal legislation.[721]

According to Goldman, this codified exception for a civil action based on violations of sex-trafficking law forms "new ground"[722] as the first real exception to Section 230 immunities

---

[710] Kosseff, 2019, *The Twenty-Six Words That Created the Internet*, p. 2.
[711] 47 USCA § 230(e)(1) (West 2018, Westlaw Next through PL 116-91).
[712] 47 USCA § 230(e)(2) (West 2018, Westlaw Next through PL 116-91).
[713] 47 USCA § 230(e)(4) (West 2018, Westlaw Next through PL 116-91).
[714] Goldman, 2020, 'An Overview of the United States' Section 230 Internet Immunity', p. 161.
[715] 47 USCA § 230(e)(3) (West 2018, Westlaw Next through PL 116-91).
[716] Goldman, 2020, 'An Overview of the United States' Section 230 Internet Immunity', p. 162.
[717] Goldman, 2020, 'An Overview of the United States' Section 230 Internet Immunity', p. 161; S. Thielman, 'Silk Road operator Ross Ulbricht sentenced to life in prison', *The Guardian,* 29 May 2015, available at theguardian.com/technology/2015/may/29/silk-road-ross-ulbricht-sentenced (retrieved on 15 February 2022).
[718] 47 USCA § 230(e)(1) (West 2018, Westlaw Next through PL 116-91).
[719] See, for example, *Jane Doe No. 1 v. Backpage.com, LLC*, 817 F.3d 12 (1st Cir. 2016).
[720] Goldman, 2020, 'An Overview of the United States' Section 230 Internet Immunity', pp. 161-162.
[721] 47 USCA § 230(e)(5)(A) (West 2018, Westlaw Next through PL 116-91).
[722] Goldman, 'The Complicated Story of FOSTA and Section 230', *First Amendment Law Review*, 2019, p. 280.

since its enactment in 1996.[723] Before the deimmunization of providers for violations of sex trafficking law, advertising illegal sex services could not lead to the service provider's liability. Because the user provided the advertisement, the provider could not be held liable as the publisher or the speaker of its content.[724]

Congress decided to add this exception after a 2016 ruling of the First Circuit.[725] In *Jane Doe No. 1 v. Backpage.com, LLC*, the First Circuit upheld Section 230 immunity for illegal sex trafficking advertisements that featured minors.[726] The argument was that Backpage.com offered features that enabled sex trafficking. For example, Backpage.com offered means to communicate anonymously, deleted advertisements that provided information against sex trafficking, and removed metadata (such as time and location) from photographs uploaded to Backpage.com.[727] At least, Backpage.com, according to the appellants, made it much easier to engage in sex trafficking. While the First Circuit found that "the appellants have made a persuasive case for that proposition", the court continued that "[s]howing that a website operates through a meretricious business model is not enough to strip away those protections."[728] The First Circuit concluded that "the remedy is through legislation, not through litigation." [729] The *Backpage* ruling was heavily criticised. Citron and Wittes, for example, argued that "[n]either the text of the statute nor its history requires sweeping immunity from liability for sites like Backpage."[730] However, the argument could be easily made the other way: what if Backpage.com would be held liable for these activities? What would that mean for other intermediaries? Unlike Roommates.com, it is hard to see how Backpage.com contributed to these advertisements other than offering the means to do so.

Sex trafficking advertisements were on the mind of legislators before *Backpage*. Before FOSTA, the SAVE Act (enacted in 2015) criminalised the advertising of sex trafficking victims by exposing providers to criminal liability when they are "knowingly assisting, supporting, or facilitating" such advertisements.[731] Congress, according to Goldman, had providers such as Backpage.com on its mind when it enacted the SAVE Act.[732] The SAVE Act did not contain a carve-out of Section 230 but relied on the pre-existing exception on Section 230 immunity for criminal law statutes.[733] The First Circuit – interpreting Section 230 before FOSTA-SESTA was enacted – held that Section 230(c)(1) offered immunity for civil liability as a speaker or publisher

---

[723] Goldman, 'The Complicated Story of FOSTA and Section 230', *First Amendment Law Review*, 2019, p. 280.
[724] Goldman, 'The Complicated Story of FOSTA and Section 230', *First Amendment Law Review*, 2019, p. 280; *Dart v. Craigslist, Inc.*, 655 F.Supp.2d 961, 967-969 (N.D.Ill. 2009).
[725] 47 USCA § 230(e)(5) (West 2018, Westlaw Next through PL 116-91); Kosseff, 2019, *The Twenty-Six Words That Created the Internet*, p. 270.
[726] *Jane Doe No. 1 v. Backpage.com, LLC*, 817 F.3d 12, 20-24 (1st Cir. 2016).
[727] *Jane Doe No. 1 v. Backpage.com, LLC*, 817 F.3d 12, 20 (1st Cir. 2016); Citron & Wittes, 'The Internet Will Not Break: Denying Bad Samaritans § 230 Immunity', *Fordham Law Review*, 2017, p. 407.
[728] *Jane Doe No. 1 v. Backpage.com, LLC*, 817 F.3d 12, 29 (1st Cir. 2016).
[729] *Jane Doe No. 1 v. Backpage.com, LLC*, 817 F.3d 12, 29 (1st Cir. 2016).
[730] Citron & Wittes, 'The Internet Will Not Break: Denying Bad Samaritans § 230 Immunity', *Fordham Law Review*, 2017, p. 409.
[731] §1591. Sex trafficking of children or by force, fraud, or coercion, 18 USCA § 1591(a) and (b) (West 2018, Westlaw Next through PL 116-193); Goldman, 'The Complicated Story of FOSTA and Section 230', *First Amendment Law Review*, 2019, p. 282.
[732] Goldman, 'The Complicated Story of FOSTA and Section 230', *First Amendment Law Review*, 2019, p. 282.
[733] 47 USCA § 230(e)(1) (West 2018, Westlaw Next through PL 116-91).

leaving the victims of sex trafficking emptyhanded even when a violation of a criminal law statute could be established.[734]

Digital Millennium Copyright Act (DMCA)

As noted, Section 230 does not apply to intellectual property law. Violations of intellectual property law are regulated in 17 USCA § 512.[735] The DMCA, which codified this Section 512, is seen as one of the (possible) inspirators for the e-Commerce Directive,[736] which may explain why both rely on a distinction between mere conduit, caching and hosting providers.[737] The DMCA, however, adds a fourth one called "information location tools".[738]

Similar to the e-Commerce Directive discussed in Chapter 4, the DMCA does not offer a liability regime of itself. Section 512, thus, does not contain provisions of when a provider becomes liable – it only sets out when a provider can rely on the safe harbour. Besides, the safe harbour regime offered in Section 512 is – like the e-Commerce Directive – voluntary. Non-compliance may only lead to legal liability for the user-provided information that would otherwise fall within the safe harbour protections.[739]

The conditional nature of the DMCA relies on actual knowledge or awareness of the existence of the infringing content of user-provided information of the internet intermediary service provider.[740] According to Wilman, evidence of actual knowledge under the DMCA is hard to prove by any other means than a receipt of a takedown notice.[741] The awareness test is, in fact, a "red flag" test.[742] Proof of awareness requires that the provider has specific awareness. General awareness that there is somewhere on the service user-provided information with infringing content is not enough now that such information is not identifiable. The third requirement of the DMCA requires expeditious removal of user-provided information that is infringing.[743] How fast removal must be to count as "expeditious" is not codified.[744] According to Wilman, seven days after gaining knowledge or awareness seems to count as expeditiously.[745]

Unlike the liability regime laid down in Article 14 of the e-Commerce, the DMCA regime, as Van Hoboken and Keller put it, "creates a detailed 'notice-and-takedown' system for content

---

[734] *Jane Doe No. 1 v. Backpage.com, LLC*, 817 F.3d 12, 23 (1st Cir. 2016).

[735] 17 USCA § 512 (West 2010, Westlaw Next through PL 116-179).

[736] Par. 2.15 of Wilman, 2020, *The Responsibility of Online Intermediaries for Illegal User Content in the EU and the US*, pp. 17-18.

[737] Compare 17 USCA § 512(a), (b) and (c) (West 2010, Westlaw Next through PL 116-179); Directive 2000/31/EC (*Directive on electronic commerce*), p. Article 12 to 14 of

[738] 17 USCA § 512(d) (West 2010, Westlaw Next through PL 116-179).

[739] Par. 5.12-5.13 of Wilman, 2020, *The Responsibility of Online Intermediaries for Illegal User Content in the EU and the US*, pp. 136-137.

[740] 17 USCA § 512(c)(1)(A)(i) and (ii) (West 2010, Westlaw Next through PL 116-179).

[741] Par. 5.28 of Wilman, 2020, *The Responsibility of Online Intermediaries for Illegal User Content in the EU and the US*, p. 145.

[742] Note 13 of Van Hoboken & Keller, 2019, 'Design Principles for Intermediary Liability Laws', p. 8; Par. 5.29 of Wilman, 2020, *The Responsibility of Online Intermediaries for Illegal User Content in the EU and the US*, p. 145.

[743] Par. 5.28-5.30 of Wilman, 2020, *The Responsibility of Online Intermediaries for Illegal User Content in the EU and the US*, pp. 145-146.

[744] 17 USCA § 512(c)(1)(A)(iii) (West 2010, Westlaw Next through PL 116-179).

[745] Par. 5.31 of Wilman, 2020, *The Responsibility of Online Intermediaries for Illegal User Content in the EU and the US*, p. 146.

alleged to infringe copyright."[746] For starters, the DMCA requires the provider to designate an agent to receive notices. This agent must be registered by the Copyright Office.[747] Notifications of infringing content must be sent to this agent. These notifications must fill certain substantial criteria as well. For example, the rightsholder (or person authorised by the rightsholder) must provide identification of the work(s) that are infringed, identification of the infringing user-provided information on the service and information of the location on the service of the provider.[748] A notice that fulfils the criteria laid down in Section 512(c)(3) can amount to actual knowledge, which requires providers to take down infringing user-provided information.[749] A notice that does not fulfil the criteria laid down in this article, however, does not.[750]

Providers that takedown or disable access to user-provided information because they (in good faith) believe that the content of the information was infringing are shielded from claims that may arise from this action.[751] Section 512(g)(2) sets out an exception from this exemption from liability for notice-based removal. In this case, the provider must inform the user who provided the information with infringing content.[752] The user may submit a counter-notice in which the user reveals his or her identity.[753] In the case of a counter-notice, the provider must forward this counter-notice to the original notifier with the announcement that it will restore access to the allegedly infringing information in ten business days.[754] The provider is required to restore access after ten but no later than 14 working days unless the submitter of the notification has notified the provider that it seeks a court order to prevent the user who provided the counter-notification from the infringing activity. In this case, the user-provided information is not restored.[755] A provider that follows the counter-notice is exempted from liability – even when it restores user-provided information with infringing content after a counter-notice because the notifier does not follow up with a legal procedure.[756]

These statutory exceptions show that the broad immunity offered by Section 230 is not a given. With respect to copyrighted content, the DMCA has a different approach that does not offer such immunity to providers. Of course, it is this broad immunity that makes Section 230 worthwhile. As discussed in the following paragraphs, this broad immunity is put under pressure.

3.1.3 What does Section 230 encourage?

Section 230 offers providers the possibility to moderate the content of user-provided information without fearing that they could be held liable for all content of user-provided information on their service.[757] Besides, Section 230 includes a provision about 'good faith' moderating which offers a safe harbour for liability for the moderation of user-provided information. As noted, Section

---

[746] Van Hoboken & Keller, 2019, 'Design Principles for Intermediary Liability Laws', p. 2.
[747] 17 USCA § 512(c)(2) (West 2010, Westlaw Next through PL 116-179).
[748] 17 USCA § 512(c)(3)(A) (West 2010, Westlaw Next through PL 116-179).
[749] 17 USCA § 512(c)(1)(C) (West 2010, Westlaw Next through PL 116-179).
[750] 17 USCA § 512(c)(3)(B)(i) (West 2010, Westlaw Next through PL 116-179).
[751] 17 USCA § 512(g)(1) (West 2010, Westlaw Next through PL 116-179).
[752] 17 USCA § 512(g)(2)(A) (West 2010, Westlaw Next through PL 116-179).
[753] 17 USCA § 512(g)(3) (West 2010, Westlaw Next through PL 116-179).
[754] 17 USCA § 512(g)(2)(B) (West 2010, Westlaw Next through PL 116-179).
[755] 17 USCA § 512(g)(2)(C) (West 2010, Westlaw Next through PL 116-179).
[756] 17 USCA § 512(g)(4) (West 2010, Westlaw Next through PL 116-179).
[757] Fishback, 'How the Wolf of Wall Street Shaped the Internet: A Review of Section 230 of the Communications Decency Act', *Texas Intellectual Property Law Journal*, 2020, pp. 284-285.

230(c)(2) is far less absolute formulated than Section 230(c)(1) by only protecting "good faith" moderation.[758]

Noteworthy is that Section 230 immunities do not actually require or encourage providers to moderate. Section 230 does not offer any stimulation to providers to engage in content moderation. Moderation, in other words, is not a requirement under the US liability regime laid down in Section 230. Section 230(c)(1) immunities for liability for the content of user-provided information also applies to providers that do not choose to moderate. Section 230 thus equally applies to services that do moderate and services that do not.[759]

Both Section 230 features, in terms of moderation, merely take away the incentive for providers to not moderate by offering two exemptions from liability.[760] Section 230(c)(1) and (2) both offer protection for liability. Section 230(c)(1) prevents providers from becoming liable for the content of user-provided information they did not or failed to moderate.[761] Section 230(c)(2) protects providers that moderate by offering a safe harbour for claims over erroneous removal of user-provided information because of its content. Section 230(c)(2) thus protects from liability for what is moderated, while Section 230(c)(1) protects providers from liability for what is not moderated. Section 230, thus, does not contain any (political) neutrality requirement in the moderation process.[762] Nor does exercising editorial control that can be equated with traditional publishers cause providers to forfeit their immunities.[763] Service providers that intervene in user-provided information because of their own preferences are thus also protected under Section 230. Fishback: "The goal was never to create an internet that was politically diverse, as some Congressmen currently believe."[764]

As Goldman notes, Section 230(c)(2) offers a solution for the so-called "Moderator's Dilemma", where moderating content would (potentially) lead to liability for the content of user-provided information that is mistakenly moderated. Service providers fearing being held liable for erroneous removal may conclude that it is better not to moderate.[765] Section 230 – of course – does not derogate from the illegal character of the content in question. Section 230 does not contain a "legalisation" of illegal or unlawful conduct on the internet. Section 230 merely shields providers from legal responsibility for this content as publishers or speakers themselves.[766] Section 230 merely takes away the fear of liability arising from moderation but does not encourage

[758] 47 USCA § 230(c)(2)(A) (West 2018, Westlaw Next through PL 116-91).

[759] Kosseff, 2019, *The Twenty-Six Words That Created the Internet*, p. 2; Gillespie, 2018, *Custodians of the Internet*, pp. 30-31.

[760] Par. 4.39-4.41 of Wilman, 2020, *The Responsibility of Online Intermediaries for Illegal User Content in the EU and the US*, pp. 114-115.

[761] As was the case with Prodigy in *Stratton Oakmont, Inc. v. Prodigy Services Co.*, 23 Media L Rep 1794 (N.Y. Sup. Ct. 1995).

[762] Fishback, 'How the Wolf of Wall Street Shaped the Internet: A Review of Section 230 of the Communications Decency Act', *Texas Intellectual Property Law Journal*, 2020, p. 277.

[763] Goldman, 2020, 'An Overview of the United States' Section 230 Internet Immunity', p. 159.

[764] Fishback, 'How the Wolf of Wall Street Shaped the Internet: A Review of Section 230 of the Communications Decency Act', *Texas Intellectual Property Law Journal*, 2020, p. 295.

[765] Goldman, 2020, 'An Overview of the United States' Section 230 Internet Immunity', pp. 157-158.

[766] 47 USCA § 230(c)(1) (West 2018, Westlaw Next through PL 116-91).

providers to moderate. As noted, Section 230(c)(1) even protects providers that refuse to take measures after notification of defamatory content.[767]

While the user providing the information with illegal or unlawful content can still be held legally liable,[768] it is hard to sue a (pseudo)anonymous user.[769] Holding the user that provided the content legally liable may especially be burdensome in the case of user-provided information with defamatory content. As discussed under Paragraph 5.2.3, this is perhaps the most cogent criticism directed against Section 230. Section 230 has little to offer in terms of legal protections for those hurt by the content of user-provided information.

Section 230 offers a legal environment where providers can moderate user-provided information with content considered harmful but not illegal under US law without fearing liability. Section 230, however, does not obligate nor encourage providers in any way to do so. Service providers that take an active approach to safeguard the quality of the discussion on their services are generally considered preferable to a completely hands-off approach.[770] However, the First Amendment makes it hard to obligate providers to engage in such content moderation. The relationship between Section 230 and the First Amendment is central to the next paragraph.

3.2 Section 230 and the First Amendment

The First Amendment may offer protections that are not, or only partially, covered by Section 230. In addition, the First Amendment also provides protections that materially overlap with Section 230. Therefore, it is essential to discuss how the First Amendment relates to Section 230. As far relevant here, the First Amendment reads that "Congress shall make no law […] abridging the freedom of speech". As discussed above, some proposals are made to amend or abolish Section 230 protections. Such proposals raise some interesting First Amendment issues. First, I discuss what additional protections the First Amendment offers on top of Section 230 protections. Then I discuss the overlap between Section 230 and the First Amendment. Thirdly, I turn to what the First Amendment and related case law would protect without Section 230.

3.2.1 The First Amendment complementing Section 230

First, providers may automatically generate some information of their own, which thus is not "provided by another information content provider"[771] or even considered developed by the service provider. Therefore, the provider could be held liable as a "publisher" or a "speaker" for the content of information that the provider commissioned or developed itself.[772] For example, providers that function as search engines take a more active stance towards content by aggregating,

---

[767] C. Omer, 'Intermediary Liability for Harmful Speech: Lessons from Abroad', *Harvard Journal of Law & Technology*, Vol. 28, No. 1, 2014, p. 304; Perry & Zarsky, 'Who Should Be Liable for Online Anonymous Defamation?', *University of Chicago Law Review Dialogue*, 2015, pp. 165-167.

[768] *Zeran v. America Online, Inc.*, 129 F.3d 327, 330 (3rd Cir. 1997).

[769] Omer, 'Intermediary Liability for Harmful Speech: Lessons from Abroad', *Harvard Journal of Law & Technology*, 2014, p. 319; Perry & Zarsky, 'Who Should Be Liable for Online Anonymous Defamation?', *University of Chicago Law Review Dialogue*, 2015, pp. 165-167.

[770] Fishback, 'How the Wolf of Wall Street Shaped the Internet: A Review of Section 230 of the Communications Decency Act', *Texas Intellectual Property Law Journal*, 2020, p. 280.

[771] 47 USCA § 230(c)(1) (West 2018, Westlaw Next through PL 116-91).

[772] 47 USCA § 230(c)(1) (West 2018, Westlaw Next through PL 116-91).

excerpting, and actively recommending content by deploying algorithms.[773] The results do not – necessarily – fall within the immunity provided by Section 230. Service providers, however, did successfully claim First Amendment protection for the output of their algorithms.[774] In some cases, the court may have (perhaps wrongly) denied Section 230 immunity but considered the output of the algorithms protected speech under the First Amendment.[775]

Especially important to note is that the First Amendment limits the possibility of regulating the content of user-provided information. For example, it is likely that the First Amendment would bar the government from regulating COVID-19 misinformation.[776]

3.2.2 Overlap between Section 230 and the First Amendment

Increasingly popular is the argument that the providers that are offering intermediary functions interfere in the freedom of expression of their users. As noted in Paragraph 3.1.2, Section 230(c)(2) is often slammed for granting providers unlimited editorial discretion.[777] As noted, Section 230(c)(2)(A) protections for moderation are extended to content "whether or not such material is constitutionally protected".[778] This provision is often misread or misunderstood as that without Section 230(c)(2), providers could no longer moderate user-provided information with constitutionally protected content.[779] Such a view on Section 230(c)(2)(A), however, could not be more wrong.

As Justice Kavanaugh argued in the majority opinion of *Manhattan Community Access Corporation v. Halleck* (2019): "the Free Speech Clause prohibits only governmental abridgment of speech. The Free Speech Clause does not prohibit private abridgment of speech."[780] According to Pollicino and Bietti, such a state action doctrine is "somewhat problematic when it comes to speech harms that occur in a highly privatized digital public sphere."[781] Pollicino and Bietti argue that such a doctrine shields a mixture of actors against liability for, for example, harmful disinformation.[782] Keller warns that it is not always evident whether the private provider or the state is responsible. States may "launder" their involvement because interventions are attributed to the service provider.[783] Even when state actors call upon private actors to take action against

---

[773] E. Volokh & D.M. Falk, 'Google: First Amendment Protection for Search Engine Search Results', *Journal of Law, Economics & Policy*, Vol. 8, No. 4, 2012, p. 884.

[774] Callamard, 'Are courts re-inventing Internet regulation?', *International Review of Law, Computers & Technology*, 2017, p. 332; *Search King Inc. v. Google Technology, Inc.*, 2003 WL 21464568 (W.D.Okla. 2003).

[775] For example, *e-ventures Worldwide, LLC v. Google, Inc.*, 2017 WL 2210029 (M.D.Fla. 2017); E. Goldman, 'First Amendment Protects Google's De-Indexing of "Pure Spam" Websites–e-ventures v. Google', *Technology & Marketing Law Blog,* 9 February 2017, available at blog.ericgoldman.org/archives/2017/02/first-amendment-protects-googles-de-indexing-of-pure-spam-websites-e-ventures-v-google.htm (retrieved on 14 February 2022).

[776] *Cohoon v. Konrath*, 2021 WL 4356069 (E.D.Wis. 2021).

[777] Hochman, 2021, 'Conservatives Should Support Section 230 Reform'.

[778] 47 USCA § 230(c)(2)(A) (West 2018, Westlaw Next through PL 116-91).

[779] See, for an overview of common Section 230 misunderstandings, Masnick, 2020, 'Hello! You've Been Referred Here Because You're Wrong About Section 230 Of The Communications Decency Act'.

[780] *Manhattan Community Access Corporation v. Halleck*, 139 S.Ct. 1921, 1928 (2019).

[781] O. Pollicino & E. Bietti, 'Truth and Deception across the Atlantic: A Roadmap of Disinformation in the US and Europe', *Italian Journal of Public Law*, Vol. 11, No. 1, 2019, p. 55.

[782] Pollicino & Bietti, 'Truth and Deception across the Atlantic: A Roadmap of Disinformation in the US and Europe', *Italian Journal of Public Law*, 2019, p. 55.

[783] Keller, 2019, 'Who Do You Sue? State and Platform Hybrid Power over Online Speech'.

(constitutionally protected) content provided by users, this does not render the provider a state actor.[784]

Irrespective of these concerns, the First Amendment, as interpreted by the SCOTUS, is unlikely to harm editorial decisions by providers.[785] Even without Section 230(c)(2)(A), providers are likely to still moderate user-provided information as they see fit.[786] In the US, it is not possible to force providers to either carry user-provided information irrespective of its content or to take down user-provided information containing protected content.[787] Of course, Section 230, in a way, bypassed (not violated) First Amendment concerns by signalling (not obligating) providers that they could also moderate harmful content that is constitutionally protected while offering some exemptions from liability for such efforts.[788] At the same time, Section 230 represents many of the same values as incorporated in the First Amendment.[789]

However, it must be kept in mind that the First Amendment does not bind providers. Unlike federal or state actors, providers are, for example, able to respond to false speech with regulation (including removal).[790] The First Amendment is one of the causes why (new) federal regulation of providers in the US does not succeed.[791]

3.2.3 Liability for providers without Section 230

Providers that rely on the First Amendment to prevent liability for the moderation of user-provided information expose themselves to criticism. As Pasquale notes, Section 230 shields providers from liability as a publisher or speaker, while First Amendment protections, in contrast, apply to providers that do function as a publisher or speaker.[792] Of course, it must be noted that relying on Section 230 immunities is not made dependent on whether the provider is indeed not functioning as publisher or speaker. Section 230 merely shields the provider from liability for the content of user-provided information as a publisher or speaker – even when the provider functions as a publisher.

Without Section 230, providers are no longer shielded from publisher liability for the content of user-provided information. Abolishing Section 230 would bring back the dilemma that arose under *Stratton Oakmont*.[793] Providers that moderate may be exposed to liability for the content of all user-provided information because they function as a publisher and not merely a distributor.

---

[784] *Doe v. Google LLC*, 2021 WL 4864418 (N.D.Cal. 2021).

[785] Moderation decisions that are also protected under 47 USCA § 230(c)(2)(a) (West 2018, Westlaw Next through PL 116-91).

[786] See, for, example, *e-ventures Worldwide, LLC v. Google, Inc.*, 2017 WL 2210029 (M.D.Fla. 2017).

[787] D. Keller, 'The Future of Platform Power: Making Middleware Work', *Journal of Democracy*, Vol. 32, No. 3, 2021 (available at muse.jhu.edu/article/797795), doi:10.1353/jod.2021.0043, p. 169.

[788] Par. 6.50 of Wilman, 2020, *The Responsibility of Online Intermediaries for Illegal User Content in the EU and the US*, p. 191.

[789] Par. 7.30 of Wilman, 2020, *The Responsibility of Online Intermediaries for Illegal User Content in the EU and the US*, p. 209.

[790] C.R. Sunstein, 'Falsehoods and the First Amendment', *Harvard Journal of Law & Technology*, Vol. 33, No. 2, 2020, p. 418.

[791] Transatlantic High Level Working Group on Content Moderation Online and Freedom of Expression, 2020, 'Freedom and Accountability: A Transatlantic Framework for Moderating Speech Online'.

[792] Pasquale, 'Platform Neutrality: Enhancing Freedom of Expression in Spheres of Private Power', *Theoretical Inquiries in Law*, 2016, p. 494.

[793] *Stratton Oakmont, Inc. v. Prodigy Services Co.*, 23 Media L Rep 1794 (N.Y. Sup. Ct. 1995).

Publisher liability may provide an incentive to providers to refrain from moderation and stimulate them to function as mere distributors for user-provided information.

Without Section 230, providers that qualify as a distributor are not liable for all user-provided information. Imposing distributors with strict liability is regarded as unconstitutional because of its chilling effects on freedom of expression rights.[794] Although the responsibilities of distributors do not go as far as those of publishers, distributors have some legal obligations with respect to the content of user-provided information that they distribute.

According to Kosseff, distributors that 1) have knowledge of the content, 2) should know about the content, or 3) distributed the content with "reckless disregard" may become liable for this content.[795] As Kosseff notes, Section 230 was enacted quickly after *Stratton Oakmont*. There is not much case law to go on regarding provider liability as content distributors.[796] However, it is not unlikely, according to Kosseff, that "the First Amendment and common law distributor caselaw likely would provide limited protection to many current online platforms."[797] Without Section 230, various providers would face new obligations regarding how they handle the content of user-provided information. Because of such obligations, they may be required to alter or cease the services they are offering. For example, providers notified of defamatory content of user-provided information would face liability when they do not take down this content because such a notification would clearly amount to knowledge. Service providers affected by such an obligation are (amongst others) services that allow reviews or social media networks. When these providers refuse to take down the content, they must go to court and argue that it is not defamatory.[798] While the provider may win the case, the risks and costs involved with such a procedure are not bearable for many providers. As Goldman notes, Section 230, in this respect, offers better protection to providers than the First Amendment because Section 230 often allows for a quick dismissal of the case. At the same time, constitutional lawsuits are often lengthy and pricy.[799]

For all providers, including search engines, the burden to moderate user-provided information would significantly increase because providers must demonstrate that they are not reckless with respect to (for example) defamatory content.[800] Apart from distributor liability, Section 230 also offers more certainty to providers. In the first place, because Section 230 declares state regulation incompatible with Section 230 inapplicable. In the second place, Section 230 applies to all common law claims based on publisher (and thus also a distributor) liability. Of course, the provider may also win publisher liability cases on First Amendment grounds, but this would (again) lead to lengthy procedures and uncertain outcomes because the First Amendment requirements differ from these common law claims.[801]

As noted, Section 230 shields providers that are functioning as publishers from liability as such. Without Section 230, the question is whether some providers (such as Wikipedia) could rely on distributor protections or whether they would be considered a publisher faced with strict

---

[794] *Smith v. People of the State of California*, 80 S.Ct. 215, 219 (1959).

[795] J. Kosseff, 'First Amendment Protection for Online Platforms', *Computer Law & Security Review*, Vol. 35, No. 5, 2019, doi:10.1016/j.clsr.2019.105340, p. 5.

[796] Kosseff, 'First Amendment Protection for Online Platforms', *Computer Law & Security Review*, 2019, p. 15.

[797] Kosseff, 'First Amendment Protection for Online Platforms', *Computer Law & Security Review*, 2019, p. 2.

[798] Kosseff, 'First Amendment Protection for Online Platforms', *Computer Law & Security Review*, 2019, p. 13 and 15.

[799] Goldman, 'Why Section 230 Is Better Than the First Amendment', *Notre Dame Law Review*, 2019, pp. 40-42.

[800] Kosseff, 'First Amendment Protection for Online Platforms', *Computer Law & Security Review*, 2019, p. 14.

[801] Goldman, 'Why Section 230 Is Better Than the First Amendment', *Notre Dame Law Review*, 2019, pp. 43-44.

liability.[802] Repealing or amending Section 230 in such a way that its immunities would be severely limited would throw many providers back in a state of nature.

Why seek to limit the immunities provided by Section 230? Why seek to abolish Section 230? Politicians on both the left and the right of the political spectrum have their reasons to do so. Left politicians seek to expose providers to liability for the content of user-provided information that they deem harmful. Politicians on the right seek to force providers to carry content they are arguing is being censored by providers. As noted, it is improbable that the SCOTUS would interpret the First Amendment to allow for such content-based restrictions or force a private provider to carry content it does not want to.[803]

3.3 Developments

3.3.1 Allow States and Victims to Fight Online Sex Trafficking Act (FOSTA)

As noted, the immunity offered by Section 230 was first meaningfully altered in 2018 by the FOSTA bill.[804] The House Committee on the Judiciary deemed it undesirable that Section 230 shields providers from civil and criminal liability for violations of the sex trafficking law.[805] Besides, the Committee argued that it is objectionable that Section 230 shields providers from the enforcement of criminal state law.[806] While Section 230 does not exempt providers from the enforcement of federal criminal law, the Committee argued that the tools available for prosecutors are not enough because it is hard to "demonstrate beyond a reasonable doubt that the website operators knew that the advertisements involved sex trafficking."[807] The FOSTA bill, which compromised between the (initial) House initiative FOSTA and the Senate initiative SESTA, was adopted by the House in February 2018 by the Senate in March 2018 and signed into law by President Trump on 11 April 2018.[808] FOSTA is criticised as offering little remedies for victims of sex trafficking. At the same time, it is argued to have devastating effects (both financially and in terms of safety) on sex workers who are deprived of the possibility of advertising their services online.[809]

As noted, before FOSTA, it was required to prove that an internet intermediary was "knowingly advertising sex trafficking".[810] Therefore the Committee concluded that "[a] new statute that instead targets promotion and facilitation of prostitution is far more useful to

---

802 Kosseff, 'First Amendment Protection for Online Platforms', *Computer Law & Security Review*, 2019, p. 14.
803 Keller, 2021, 'Six Constitutional Hurdles for Platform Speech Regulation'.
804 Goldman, 'The Complicated Story of FOSTA and Section 230', *First Amendment Law Review*, 2019, p. 280. However, in 2006 there were made some adjustments to Section 230 immunities for online gambling which had in the words of Goldman an "ambiguous effect", see Note 7 of Goldman, 'The Complicated Story of FOSTA and Section 230', *First Amendment Law Review*, 2019, p. 280. Goldman argues that "on balance, it looks like this law may have slightly expanded ICS [interactive computer service, MK] immunization by providing some limits on ICS liability for third party criminal gambling activities.", see E. Goldman, 'Unlawful Internet Gambling Enforcement Act of 2006', *Technology & Marketing Law Blog,* 13 December 2006, available at blog.ericgoldman.org/archives/2006/12/unlawful_intern.htm (retrieved on 14 February 2022).
805 H.R. Rep. No. 115–572, pt. 1, at 5 (2018).
806 H.R. Rep. No. 115–572, pt. 1, at 5 (2018).
807 H.R. Rep. No. 115–572, pt. 1, at 5 (2018).
808 Goldman, 'The Complicated Story of FOSTA and Section 230', *First Amendment Law Review*, 2019, pp. 283-284.
809 Goldman, 'The Complicated Story of FOSTA and Section 230', *First Amendment Law Review*, 2019, pp. 290-292.
810 H.R. Rep. No. 115–572, pt. 1, at 5 (2018).

prosecutors."[811] FOSTA introduced two new federal crimes.[812] The new 18 USCA §2421A introduces a new federal criminal crime for whoever

> owns, manages, or operates an interactive computer service (as such term is defined in defined in section 230(f) the Communications Act of 1934 (47 U.S.C. 230(f))), or conspires or attempts to do so, with the intent to promote or facilitate the prostitution of another person shall be fined under this title, imprisoned for not more than 10 years, or both.[813]

This statute allows for a more aggregative penalty (25 years) when whoever "owns, manages, or operates an interactive computer service" uses this service to

> (1) promotes or facilitates the prostitution of 5 or more persons; or
>
> (2) acts in reckless disregard of the fact that such conduct contributed to sex trafficking, in violation of 1591(a)[814]

FOSTA also introduced a new federal crime by changing 18 USCA §1591. Section 1591 already criminalised "sex trafficking of children or by force, fraud, or coercion". However, Section 1591 was amended to include whoever "benefits, financially or by receiving anything of value, from participation in a venture",[815] which is defined as "knowingly assisting, supporting, or facilitating"[816] sex trafficking. Participating in a venture that "recruits, entices, harbors, transports, provides, obtains, advertises, maintains, patronizes, or solicits by any means a person"[817] while knowing "that such means will be used to cause the person to engage in a commercial sex act, or that the person has not attained the age of 18 years and will be caused to engage in a commercial sex act"[818] can be fined by a prison sentence from 10 years to life.[819] In the case of advertisement, actual knowledge is not required, but it is enough to demonstrate "reckless disregard".[820] Reckless disregard is a significantly lower bar for providers to become criminally liable for sex trafficking advertisements than knowledge.

Section 230 did not shield providers from federal prosecution before FOSTA. Even without amending Section 230, providers could not rely on Section 230 to shield them from federal prosecution.[821] Section 230, however, did shield providers from some criminal prosecution at the state level. To recall, Section 230 provides immunity for legal actions and liability under state law when these laws are inconsistent with Section 230.[822] Section 230, however, did shield providers

---

811 H.R. Rep. No. 115–572, pt. 1, at 5 (2018).
812 Goldman, 'The Complicated Story of FOSTA and Section 230', *First Amendment Law Review*, 2019, p. 284.
813 §2421A. Promotion or facilitation of prostitution and reckless disregard of sex trafficking, 18 USCA § 2421A(a) (West 2018, Westlaw Next through PL 116-193).
814 18 USCA § 2421A(b) (West 2018, Westlaw Next through PL 116-193).
815 18 USCA § 1591(a)(1) (West 2018, Westlaw Next through PL 116-193).
816 18 USCA § 1591(e)(4) (West 2018, Westlaw Next through PL 116-193).
817 18 USCA § 1591(a)(1) (West 2018, Westlaw Next through PL 116-193).
818 18 USCA § 1591(a) (West 2018, Westlaw Next through PL 116-193).
819 18 USCA § 1591(b) (West 2018, Westlaw Next through PL 116-193).
820 18 USCA § 1591(a) (West 2018, Westlaw Next through PL 116-193).
821 47 USCA § 230(e)(1) (West 2018, Westlaw Next through PL 116-91).
822 47 USCA § 230(e)(3) (West 2018, Westlaw Next through PL 116-91).

from civil liability since most of the exemptions (until 2018) saw to (federal) enforcement of criminal law.[823]

FOSTA changed this by directly amending Section 230 by carving out immunity for state prosecution for "any charge in a criminal prosecution brought under State law if the conduct underlying the charge would constitute a violation of section 1591 of Title 18".[824] The same applies to criminal charges based on "a violation of section 2421A of Title 18".[825] Also new is that civil claims based on violations of Section 1591 are no longer immunised by Section 230.[826] According to Goldman, the Section 230 carve-out with respect to civil claims does not apply to Section 2421A, "even though that seems inconsistent with FOSTA's purposes."[827]

Goldman and other commentators question the necessity and desirability of FOSTA. While FOSTA targets providers as Backpage.com, the brand-new Section 2421A and the modified Section 1591 were not used to prosecute the management of Backpage.com.[828] Besides, Goldman questions the necessity of amending Section 230 to enable victims to sue providers for damages because of the violations of Section 2421A and Section 1591. As Goldman notes, a successful prosecution under Section 1591 requires mandatory restitution of victims under Section 1593.[829] Goldman also points out that Section 230 protection of Backpage.com before FOSTA was not a given now evidence was provided that Backpage.com was involved in developing the content of the published advertisements.[830] To recall, Section 230 does not protect providers involved in developing the content of user-provided information.[831]

Before FOSTA became law, two court rulings denied Section 230 immunity for civil action direct at Backpage.[832] FOSTA, thus, seems to do very little to protect victims of sex trafficking, while the potential adverse effects on providers are severe. According to Goldman, FOSTA leaves providers with three options: perfect moderation, no moderation at all to prevent knowledge, or cease to offer these services.[833] Of course, near-perfect moderation is only a potential possibility for the larger providers, while the smaller ones would choose to cease their services to prevent liability.[834]

3.3.2 The Trump-administration

Not only FOSTA may introduce the risk that providers refrain from moderating or stopping their services. On 28 May 2020, the Trump administration issued an executive order discussing this

---

[823] New is 47 USCA § 230(e)(5) (West 2018, Westlaw Next through PL 116-91). Of course, civil action for copyright violations was already possible since Section 230 does not see to intellectual property violations, see 47 USCA § 230(e)(1) (West 2018, Westlaw Next through PL 116-91).

[824] 47 USCA § 230(e)(5)(b) (West 2018, Westlaw Next through PL 116-91).

[825] 47 USCA § 230(e)(5)(c) (West 2018, Westlaw Next through PL 116-91).

[826] 47 USCA § 230(e)(5)(a) (West 2018, Westlaw Next through PL 116-91).

[827] Goldman, 'The Complicated Story of FOSTA and Section 230', *First Amendment Law Review*, 2019, p. 285.

[828] Goldman, 'The Complicated Story of FOSTA and Section 230', *First Amendment Law Review*, 2019, p. 286.

[829] §1593. Mandatory restitution, 18 USCA § 1593 (West 2018, Westlaw Next through PL 116-193); Goldman, 'The Complicated Story of FOSTA and Section 230', *First Amendment Law Review*, 2019, p. 287.

[830] Goldman, 'The Complicated Story of FOSTA and Section 230', *First Amendment Law Review*, 2019, p. 287.

[831] *Batzel v. Smith*, 333 F.3d 1018, 1031 (9th Cir. 2003); *Fair Housing Council of San Fernando Valley v. Roommates.com, LLC*, 521 F.3d 1157, 1166-1167 (9th Cir. 2008).

[832] Goldman, 'The Complicated Story of FOSTA and Section 230', *First Amendment Law Review*, 2019, p. 288.

[833] Goldman, 'The Complicated Story of FOSTA and Section 230', *First Amendment Law Review*, 2019, p. 288.

[834] Goldman, 'The Complicated Story of FOSTA and Section 230', *First Amendment Law Review*, 2019, pp. 288-289.

'Good Samaritan' clause. The Executive order stated that protection for liability from moderating content should only apply to "obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable"[835] content. Service providers that moderate other content should lose their immunity and be regarded as editors and thus publishers of this content.[836] Besides, the executive order sought to reinterpret Section 230 in such a way that good faith moderation cannot be "deceptive, pretextual, or inconsistent with a provider's terms of service" and that providers must "provide adequate notice, reasoned explanation, or a meaningful opportunity to be heard".[837]

The Trump administration proposed legislation expanding categories of content that do not fall under the immunity provided by Section 230. The proposal, for example, included violations of anti-terrorism, child abuse and cyber-stalking laws as carved out from the immunity provided by Section 230. Besides, the proposed legislation also deimmunizes providers for civil lawsuits based on these exceptions.[838] Under current law, victims of sexual child abuse material suing providers for damages would be left empty-handed. Section 230 merely exposes providers to civil lawsuits based on criminal law violations of sex trafficking law.[839]

In September 2020, the Trump Administration also proposed legislation to codify this new understanding of 'Good Samaritan' moderation in Section 230.[840] This proposal seeks to limit this exception to the content that the provider "has an objectively reasonable belief is obscene, lewd, lascivious, filthy, excessively violent, promoting terrorism or violent extremism, harassing, promoting self-harm, or unlawful".[841] Section 230 in its current form also includes "otherwise objectionable" as a ground for removal.[842] The proposal removes this exception, severely limiting the possibility for providers to take down legal but undesirable content.

Besides, the draft includes what is referred to as a "'Bad Samaritan' carve-out" deimmunizing providers that

> acted purposefully with the conscious object to promote, solicit, or facilitate material or activity by another information content provider that the service provider knew or had reason to believe would violate Federal criminal law, if knowingly disseminated or engaged in.[843]

The draft legislation does not alter the immunity provided to providers for defamation claims.[844] For other categories of content, the draft establishes a conditional liability approach. Service

---

[835] 47 USCA § 230(c)(1)(a) (West 2018, Westlaw Next through PL 116-91).

[836] The White House, 'Executive Order on Preventing Online Censorship', *The White House,* 28 May 2020, available at trumpwhitehouse.archives.gov/presidential-actions/executive-order-preventing-online-censorship (retrieved on 15 February 2022).

[837] The White House, 2020, 'Executive Order on Preventing Online Censorship'.

[838] § 230(f) US Department of Justice, 'Department of Justice's Review of Section 230 of the Communications Decency Act of 1996 (Redline)', *US Department of Justice,* 23 September 2020, available at justice.gov/file/1319331/download (retrieved on 15 February 2022).

[839] 47 USCA § 230(e)(1) and (e)(5) (West 2018, Westlaw Next through PL 116-91).

[840] § 230(c)(2)(a) US Department of Justice, 2020, 'Department of Justice's Review of Section 230 of the Communications Decency Act of 1996 (Redline)'.

[841] § 230(c)(2)(a) US Department of Justice, 2020, 'Department of Justice's Review of Section 230 of the Communications Decency Act of 1996 (Redline)'.

[842] 47 USCA § 230(c)(1)(a) (West 2018, Westlaw Next through PL 116-91).

[843] § 230(d)(1) US Department of Justice, 2020, 'Department of Justice's Review of Section 230 of the Communications Decency Act of 1996 (Redline)'.

[844] US Department of Justice, 'Department of Justice's Review of Section 230 of the Communications Decency Act of 1996 (Cover Letter)', *US Department of Justice,* 23 September 2020, available at justice.gov/file/1319346/download

providers that have actual knowledge may become liable when they "had actual notice of that material's or activity's presence on their service and its illegality" and failed to "expeditiously remove, restrict access to or availability of, or prevent dissemination" of user-provided information with such content.[845]

The draft legislation also proposed to codify what 'good faith' moderation is. According to the draft proposal, providers must issue clear content moderation policies which form the basis of content moderation practices of the service provider. Besides, providers are, under this proposal, obligated to notify the users of the restriction while mentioning the grounds on which this restriction took place. The user must be offered the opportunity to respond. Only user-provided information containing terrorist content or content related to criminal activities does not require such notification. Such notification is also not mandatory if it "would risk imminent harm to others".[846]

Of course, with the 2020 presidential elections, these Section 230 reform plans are archived with the rest of the administrations' websites. Although these proposals are no longer up to date, they give an idea of what adjustments to Section 230 are being considered.

3.3.3 The Biden-administration

After Biden became president-elect, the question arose about what this would mean for Silicon Valley and Section 230.[847] After all, Biden has also taken a firm stance on Section 230 in the past, including the statement that Section 230 protection "should be revoked because it [Facebook, MK] is not merely an internet company."[848] According to Biden, Mark Zuckerberg "should be submitted to civil liability and his company [Facebook, MK] to civil liability, just like you [the editorial board, MK] would be here at The New York Times."[849] Revoking or amending Section 230 protection for companies such as Google, Twitter and Meta is thus a recurring theme both amongst democrats and republicans in the US.

In May 2021, Biden revoked "Executive Order 13925 of May 28, 2020 (Preventing Online Censorship)" [850] issued by the Trump Administration.[851] Besides, Biden nominated Jessica Rosenworcel as the new chair of the Federal Communications Commission. Rosenworcel,

---

(retrieved on 15 February 2022); § 230 (d) (3) US Department of Justice, 2020, 'Department of Justice's Review of Section 230 of the Communications Decency Act of 1996 (Redline)'.

[845] § 230(d)(2) US Department of Justice, 2020, 'Department of Justice's Review of Section 230 of the Communications Decency Act of 1996 (Redline)'.

[846] § 230(g)(5) US Department of Justice, 2020, 'Department of Justice's Review of Section 230 of the Communications Decency Act of 1996 (Redline)'.

[847] K. Stacey & R. Waters, 'What can Silicon Valley expect from Joe Biden?', *Financial Times,* 8 November 2020, available at ft.com/content/44f738e8-eb6f-4394-b833-6b3207ce31bf (retrieved on 20 May 2021).

[848] The editorial board of The New York Times, 'Joe Biden: Former vice president of the United States', *The New York Times,* 17 January 2020, available at nytimes.com/interactive/2020/01/17/opinion/joe-biden-nytimes-interview.html (retrieved on 15 February 2022).

[849] The Editorial Board of the New York Times, 2020, 'Joe Biden: Former vice president of the United States'.

[850] The White House, 2020, 'Executive Order on Preventing Online Censorship'.

[851] The White House, 'Executive Order on the Revocation of Certain Presidential Actions and Technical Amendment', *The White House,* 14 May 2021, available at whitehouse.gov/briefing-room/presidential-actions/2021/05/14/executive-order-on-the-revocation-of-certain-presidential-actions-and-technical-amendment (retrieved on 15 February 2022).

according to commentators, opposes Section 230 reform.[852] On 7 December 2021, Rosenworcel was confirmed by the Senate as the first female chair of the FCC.[853] Of course, the Biden Administration may also seek to directly amend Section 230 instead of reforming Section 230 by seeking reinterpretation by the FCC. As the Washington Post wrote in January 2021, Section 230 was "a favorite punching bag of President Trump's in the past year when social media companies removed posts and accounts", but "[d]emocrats also think the law should be amended, but for different reasons: Tech companies should be held more responsible for moderating content on their sites."[854]

Conclusion

Section 230 does not grant providers unlimited discretion to moderate the content of user-provided information as they see fit. Section 230 does not grant providers the right to moderate constitutionally protected speech. Nor does Section 230 exempt providers from liability for lawful but harmful content disseminated through their services. Any Section 230 carve-out or other proposed amendment to change this will ultimately stumble upon First Amendment concerns. The First Amendment guarantees that providers are allowed to make editorial decisions on their services. However, the First Amendment does not apply between the provider and the user of the service. The First Amendment, however, does apply to the relationship between the government and the service provider. Therefore, the user nor the government can force providers to carry the content of user-provided information. At the same time, the government also cannot force the provider to take down user-provided information with constitutionally protected content.

Section 230(c)(1) does not grant rights to providers; it merely grants immunities for liability that may arise from how they offer their services. In doing so, Section 230(c)(1) applies to almost all providers and users of these services ("provider or user of an interactive computer service"). Section 230(c)(1), thus, bypasses debates over what an "internet intermediary service provider" is and offers legal certainty to all providers that they can rely on this protection. Besides, Section 230(c)(1) applies to "any information provided by another information content provider." Section 230(c)(1) immunities thus merely apply to the content of user-provided information. When the provider can be regarded as the developer of the content, Section 230(c)(1) immunities do not apply. As discussed, the courts are not very inclined to assume that this is the case. Only in fairly clear-cut cases has a provider been considered as the content developer of user-provided information. Section 230(c)(1) shields providers against publisher or speaker liability which, as shown, involves the majority of the legal grounds on which users or third parties could base their claims. When a provider is held responsible for the content of user-provided information it did not create or develop itself, the chances are good that Section 230(c)(1) bars the claim. As noted,

---

[852] N. Krishan, 'FCC nominee's record is at odds with Biden censorship goals', *The Washington Examiner,* 30 November 2021, available at washingtonexaminer.com/policy/fcc-nominees-record-is-at-odds-with-biden-censorship-goals (retrieved on 15 February 2022); W. Kimball, 'Biden Nominates Net Neutrality Champion Jessica Rosenworcel to Head the FCC', *Gizmodo,* 26 October 2021, available at gizmodo.com/biden-nominates-net-neutrality-champion-jessica-rosenwo-1847938641 (retrieved on 15 February 2022).

[853] R. Brandom, 'Jessica Rosenworcel confirmed by Senate to lead the FCC', *The Verge,* 7 December 2021, available at theverge.com/2021/12/7/22820873/jessica-rosenworcel-fcc-chair-confirmed-biden-net-neutrality (retrieved on 28 January 2022).

[854] R. Lerman, 'Social media liability law is likely to be reviewed under Biden', *The Washington Post,* 18 January 2021, available at washingtonpost.com/politics/2021/01/18/biden-section-230 (retrieved on 28 March 2022).

Section 230(c)(1) often allows providers to be granted a motion to dismiss, which avoids lengthy and expensive trials.

The protection offered by Section 230(c)(2) is less clear and more open to discussion in a court procedure. To successfully rely on Section 230(c)(2), the provider must demonstrate that it moderated in "good faith" and that its moderation efforts saw to "obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable" content of user-provided information. Section 230(c)(2), thus, offers a safe harbour and does not completely immunise the provider for liability from moderation efforts. Section 230(c)(2) protections are extended over constitutionally protected content of user-provided information. Section 230(c)(2) thus requires providers to demonstrate that they 1) "restrict access to or availability of material" that it 2) "considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable". As noted, providers typically rely on the contract with their users, which allows them to moderate user-provided information as they see fit, which makes Section 230(c)(2) less powerful of an instrument than Section 230(c)(1).

The question is how proposals to repeal or amend Section 230 will turn out. Smith & Alstyne, for example, propose to update Section 230 because they view the immunities as disincentivizing providers to moderate the content of user-provided information. Requiring some duty of care to rely on Section 230 protections, according to them, may remedy this situation.[855] Even when it is true that Section 230 leads to underregulation of harmful or illegal and otherwise unlawful content of user-provided information, less absolute internet intermediary liability regimes, however, may have their disadvantages.

The balance will be made up in Chapter 5. Before the pros and cons of different liability regimes can be discussed, it is necessary to discuss how the European approach toward internet intermediary liability for the content of user-provided information seeks to address these issues.

[855] M.D. Smith & M. van Alstyne, 'It's Time to Update Section 230', *Harvard Business Review,* 12 August 2021, available at hbr.org/2021/08/its-time-to-update-section-230 (retrieved on 15 February 2022).

4 The European Approach

Uncertainty about the liability of providers for the content of user-provided information was feared to hamper innovation and threaten the competitiveness of (new-found) internet companies.[856] In the EU, providers are (and were) regulated by a patchwork of legislation. The 27 Member States of the EU criminalize and impose civil liability for distinct expressions, also applying to internet content. Therefore, the e-Commerce Directive offers some harmonization of when 'Information Society services' that handle content originating from others can become liable. Not by harmonizing the substantive provisions within the EU but by exempting them from liability for illegal content when fulfilling a set of conditions.[857] After the e-Commerce Directive, other directives and regulations explicitly targeted providers with new substantive provisions, often harmonising their liability for specific types of content.[858] These Regulations and Directives will not be extensively discussed in this chapter. The focus lies on Articles 14 and 15 of the e-Commerce Directive, which set out the safe harbours and exceptions for providers that have direct involvement with the content of user-provided information.

The general rule seems simple and elegant. The e-Commerce Directive exempts intermediaries from liability for the content provided by others as long as they are not aware of its illegal character.[859] Providers are, additionally, not obligated to search for illegal content.[860] While the e-Commerce Directive prohibits a general obligation to monitor for illegal content, this does not stand in the way of court orders to take down (and keep down) specific content.[861] Providers are thus not entirely exempted from legal obligations. Unlike the US approach laid down in Section 230,[862] the immunity of Information Society services is conditional upon fulfilling a set of predefined criteria.[863] These conditions differ between various categories of providers based on their technological and functional involvement in the content of the information provided by users.

To clarify the difference between providers, I first set out what actors the Directive protects. As noted, the focus of this dissertation is on providers that have direct involvement in

---

856 Recital 2 and 60 of Directive 2000/31/EC (*Directive on electronic commerce*).

857 Articles 12 to 14 of Directive 2000/31/EC (*Directive on electronic commerce*).

858 For example, Directive 2005/29/EC of the European Parliament and of the Council of 11 May 2005 concerning unfair business-to-consumer commercial practices in the internal market and amending Council Directive 84/450/EEC, Directives 97/7/EC, 98/27/EC and 2002/65/EC of the European Parliament and of the Council and Regulation (EC) No 2006/2004 of the European Parliament and of the Council (*Unfair Commercial Practices Directive*), *OJ L 149*, *11.6.2005* (data.europa.eu/eli/dir/2005/29/oj); Directive (EU) 2019/2161 of the European Parliament and of the Council of 27 November 2019 amending Council Directive 93/13/EEC and Directives 98/6/EC, 2005/29/EC and 2011/83/EU of the European Parliament and of the Council as regards the better enforcement and modernisation of Union consumer protection rules, *OJ L 328*, *18.12.2019* (data.europa.eu/eli/dir/2019/2161/oj); Directive 2010/13/EU (*Audiovisual Media Services Directive*); Directive (EU) 2018/1808; Regulation (EU) 2021/784.

859 Recital 5 and 8 of Directive 2000/31/EC (*Directive on electronic commerce*).

860 Article 15 of Directive 2000/31/EC (*Directive on electronic commerce*).

861 Article 15 of the Directive does not stand in the way of a court order to monitor for specific infringements as long "[d]ifferences in the wording of that equivalent content, compared with the content which was declared to be illegal, must not, in any event, be such as to require the host provider concerned to carry out an independent assessment of that content." see Judgement of the Court (Third Chamber) of 3 October 2019 in *C-18/18, Glawischnig-Piesczek*, ECLI:EU:C:2019:821, in particular Rec. 45.

862 47 USCA § 230 (West 2018, Westlaw Next through PL 116-91).

863 17 USCA § 512 (West 2010, Westlaw Next through PL 116-179).

the content of user-provided information. Therefore, the rest of the chapter continues with these providers to discuss what the Directive protects. As noted, regulation is enacted to foster policy goals that are either implicit or explicit laid down in the chosen instrument. The third part of this chapter thus discusses what the Directive seeks to encourage.

In the European context, the Directive sets out the internet intermediary liability regime for the EU. Of course, (non-)compliance of Member States with this liability regime laid down in transposed national legislation may also raise freedom of expression concerns. The European Court of Human Rights may hear cases in which the petitioner argues that the state violated their freedom of expression rights. Therefore, the ECtHR was able to discuss the relation between internet intermediary liability and freedom of expression rights. In doing so, the ECtHR set out case law that may (partly) contrast with the liability regime set out in the Directive.

With the DSA as proposed by the EC in December 2020, the EC seeks to lay down numerous new obligations for providers. The DSA proposes new norms for the liability of providers for the content of user-provided information. In this chapter, merely the internet intermediary liability regime under the DSA is discussed.

4.1 The EU approach in the e-Commerce Directive

4.1.1 Which providers does the Directive protect?

As mentioned above, internet intermediaries are not defined in EU legislation as a legal category. While the e-Commerce Directive mentions internet intermediaries,[864] it relies on the much broader Information Society service.[865] An Information Society service is "any service normally provided for remuneration, at a distance, by electronic means and at the individual request of a recipient of services".[866] An Information Society service is thus provided at a distance, electronically, and upon individual request. The definition of 'Information Society service', for example, does not include radio transmissions because these are not offered on individual request.[867] Services fitting this definition are, for example, services such as social media platforms (such as Facebook) and streaming (such as Netflix). While individual users are usually not required to pay for access to social media services, they are considered offered for remuneration. These services do not require users to pay for their services directly for this criterion to be fulfilled. Because providers offering social media platforms show (personalised) advertisements to their users based on their usage, these services are offered for remuneration indirectly.[868] In other words, it can be argued that the user of these services pays with their data.

"Information Society service" is, thus, a broad category. Within this category, these services differ in what influence they have over the content distributed by them. Some Information Society services have a decisive say in what content they offer (for example, Netflix), but many providers function as internet intermediaries for user-provided information.[869] In this respect, the

---

[864] Intermediaries are, for example, mentioned in Recital 14, 40, 45, and 50, in Article 1 under paragraph 2 and in the header of Section 4 of Directive 2000/31/EC (*Directive on electronic commerce*).

[865] Article 1 and Article 2 under a and b of the Directive 2000/31/EC (*Directive on electronic commerce*).

[866] Article 1 under paragraph 1 sub b of Directive (EU) 2015/1535.

[867] Annex 1 of Directive (EU) 2015/1535.

[868] Par. 2.06 of Wilman, 2020, *The Responsibility of Online Intermediaries for Illegal User Content in the EU and the US*, p. 14.

[869] 'User-created content' is content that 1) is published on a (public) network, 2) requires a creative effort, and 3) is created outside the professional sphere, see OECD, 2007, 'Participative Web and User-Created Content'.

e-Commerce Directive defines three categories of internet services: mere conduit, caching, and hosting services. The Directive, thus, does not shield providers but merely some of the activities undertaken by them. The Directive offers protection for liability arising from offering services – but only in the capacity of offering this service.[870] Wilman refers to this as a "functional approach".[871] Not the actor, but the function providers fulfil is decisive for a successful reliance on the safe harbours.

I set out these three activities briefly because mainly the last category of providers (hosting services) is essential for discussing liability for the content of user-provided information.

Mere conduit

The first service activity distinguished by the Directive is so-called 'mere conduit' services. The best example of a 'mere conduit' service provider is an access provider. For example, ISPs that offer users an internet connection by, for example, cable, (optic)fibre or wireless signals function as a 'mere conduit'. A mere conduit service provider benefits from legal immunity of Article 12 of the Directive as long the provider

> (a) does not initiate the transmission;
>
> (b) does not select the receiver of the transmission; and
>
> (c) does not select or modify the information contained in the transmission.[872]

According to Article 12 of the Directive, a mere conduit service provider cannot be involved in the content of the information, nor can a mere conduit service initiate the transmission, nor select who receives the information. An easy comparison is postal services: the postal service does not proofread the content of a message, nor does it decide who receives postal pieces and who does not. The postal service merely delivers the message where the sender wants it to be delivered, and the postal service does not send postal pieces out on its own. Mere conduit services thus only undertake activities that are of "a mere technical, automatic and passive nature".[873] Mere conduit services thus provide technical means to transmit information, but it does stop there. Mere conduit service providers do not manually select information or recipient of this information, nor are they otherwise actively involved in the content of the transmitted information.

Caching service providers

A mere conduit service provider only offers the means to transmit information – irrespective of the sender, receiver, or its content. Mere conduits, as the name suggests, only carry information. Caching service providers do not store information. The second category of services also includes some storage of information. Caching providers store information available on other services on other (local) servers. Because the internet is a global network, this may mean that a user seeks to access a website hosted on the other side of the world. Retrieving information from the other side of the world means that the information must be transmitted to and from servers located there. The function caching service providers fulfil is temporarily storing the content available on another

---

[870] Par. 2.35 of Wilman, 2020, *The Responsibility of Online Intermediaries for Illegal User Content in the EU and the US*, p. 26.
[871] Par. 2.35 of Wilman, 2020, *The Responsibility of Online Intermediaries for Illegal User Content in the EU and the US*, p. 26.
[872] Article 12, paragraph 1 of Directive 2000/31/EC (*Directive on electronic commerce*).
[873] Recital 42 of Directive 2000/31/EC (*Directive on electronic commerce*).

service on local endpoints of the caching service to increase the services' availability (and thus speed). In caching information, providers may also – unintentionally – cache information with illegal or unlawful content. Article 13 of the Directive shields caching service providers from liability for this content as long as

> (a) the provider does not modify the information;
>
> (b) the provider complies with conditions on access to the information;
>
> (c) the provider complies with rules regarding the updating of the information, specified in a manner widely recognised and used by industry;
>
> (d) the provider does not interfere with the lawful use of technology, widely recognised and used by industry, to obtain data on the use of the information; and
>
> (e) the provider acts expeditiously to remove or to disable access to the information it has stored upon obtaining actual knowledge of the fact that the information at the initial source of the transmission has been removed from the network, or access to it has been disabled, or that a court or an administrative authority has ordered such removal or disablement.[874]

While caching service providers have access to and, therefore, control over the content of the information they cache, they are exempted from liability that may arise from the potentially illegal or unlawful content. However, the caching service provider cannot modify the content of the cached information and should comply with the conditions set out by the provider from which the cached information originates. The caching service provider must also comply with industry standards in how often the caching provider refreshes the cached information and cannot impose restrictions on the usage of the cached information contrary to these standards. Similar to the mere conduit service provider, the caching by the service provider must be of "a mere technical, automatic and passive nature".[875] The caching service provider can be forced into a more active role when the origin of the cached information deletes information with illegal or unlawful content. When this is the case, the caching service is also obligated to do so. The caching provider must act expeditiously on gaining knowledge that the source deleted or disabled access or when a court or administrative authority passed an order to this effect.[876]

Necessary to remark is that the conditional immunity provided by the safe harbours laid down in Article 12 (mere conduit) and Article 13 (caching) does not stand in the way of court orders "to terminate or prevent an infringement."[877] While mere conduit and caching service providers have no general obligation to monitor for illegal or unlawful content information, a court can order the provider to end a specific violation.[878]

Hosting service providers

The third (and for regulating the content user-provided information most important) category is hosting service providers. Article 14 of the Directive offers protection to information society services from liability for activities "that consists of the storage of information provided by a

---

[874] Article 13 of Directive 2000/31/EC (*Directive on electronic commerce*).

[875] Recital 42 of Directive 2000/31/EC (*Directive on electronic commerce*).

[876] Article 13 of Directive 2000/31/EC (*Directive on electronic commerce*).

[877] Articles 12(3) and 13(2) of Directive 2000/31/EC (*Directive on electronic commerce*).

[878] Articles 12(3), 13(2) and 15 of Directive 2000/31/EC (*Directive on electronic commerce*).

recipient of the service".[879] Hosting service providers, thus, store the information provided by the "recipient of the service", which can be a user of flesh and blood or another service provider.

The exemption of hosting service providers from liability is made conditional. The hosting service provider can only rely on the safe harbour of Article 14(1) as long as the service provider

> does not have actual knowledge of illegal activity or information and, as regards claims for damages, is not aware of facts or circumstances from which the illegal activity or information is apparent[880]

A hosting service provider that, for example, examines all content of user-provided information before admission cannot rely on Article 14(1) when this causes the provider to "have actual knowledge of illegal activity" or becomes "aware of facts or circumstances from which the illegal activity or information is apparent". However, the provider is only exposed to liability when it does not "acts expeditiously to remove or to disable access to the information."[881]

According to Wilman, the ECJ laid down in *L'Oréal v. eBay*[882] a test consisting of three steps to determine if a provider could qualify as a hosting service provider under Article 14. First, it is necessary to establish whether the intermediary qualifies as an information society service provider. Second, the activity of the intermediary must consist of hosting. Third, the hosting provider must act neutral.[883] According to Wilman, only after meeting these three criteria, Article 14 should be applied. If these three criteria are not met, the provider cannot call in the safe harbour of Article 14.[884] As noted, an Information Society service is "any service normally provided for remuneration, at a distance, by electronic means and at the individual request of a recipient of services".[885] Hosting service providers would typically qualify as an Information Society service. More interesting is to look at when a provider would qualify as a hosting service and when a provider is "neutral".

The activity of hosting

Hosting, at least, is offering storage of information. According to Riordan, Article 14 of the Directive strictly protects "storage activities".[886] However, such a strict reading of hosting raises questions about the meaning of Article 14 for services connected to hosting activities, such as social media functionalities. Wilman suggests that storage encompasses more than only storing information. The clearest example is that hosting services providers necessarily also provide access to content. Without the ability to provide access, hosting as a service would be rendered completely useless. Wilman, "providing […] access seems an almost inherent part of the concept of 'hosting'."[887] Riordan argues that the safe harbour of Article 14 may be extended to other

---

[879] Article 14(1) of Directive 2000/31/EC (*Directive on electronic commerce*).

[880] Article 14(1)(a) of Directive 2000/31/EC (*Directive on electronic commerce*).

[881] Article 14(1)(b) of Directive 2000/31/EC (*Directive on electronic commerce*).

[882] Judgement of the Court (Grand Chamber) in *C-324/09* (*L'Oréal v. eBay*).

[883] Par. 2.56 of Wilman, 2020, *The Responsibility of Online Intermediaries for Illegal User Content in the EU and the US*, p. 35.

[884] Par. 2.56-2.57 and 2.101 of Wilman, 2020, *The Responsibility of Online Intermediaries for Illegal User Content in the EU and the US*, p. 35 and 55. However, Riordan views "neutrality and passivity" as a test whether the protections of Article 14 apply to a hosting service that qualifies as such, see Par. 12.154 of Riordan, 2016, *The Liability of Internet Intermediaries*, p. 409.

[885] Article 1 under paragraph 1 sub b of Directive (EU) 2015/1535.

[886] Par. 12.113 of Riordan, 2016, *The Liability of Internet Intermediaries*, p. 401.

[887] Par. 2.37 of Wilman, 2020, *The Responsibility of Online Intermediaries for Illegal User Content in the EU and the US*, p. 27.

"technical and automatic processes that are applied neutrally to hosted information."[888] However, this only applies to "limited forms of network-layer assistance" and not to "more interventionist application-layer activities".[889] As discussed in Paragraph 1.3.1, the application layer encompasses all presentation and manipulation of the information shown to users.[890] Network layer activities, in contrast, encompass transmitting information from one computer to another.[891] Riordan, thus, argues that Article 14 of the Directive does not apply to activities such as moderation and curation.[892] The conclusion that immunity of hosting services should be limited to network layer activities is not unreasonable per se when considering the role hosting services fulfil on the internet. However, from a functional viewpoint, limiting the safe harbour to activities that do not involve the application layer level drastically restricts the legal meaning of Article 14 of the Directive because many providers that utilise hosting capabilities combine this with more interventionist activities. While technically correct, such a limited reading would render the safe harbour useless for a broad range of providers.

In contrast to Riordan, Wilman argues that the Directive gives reason for a broad reading of the concept of hosting. Wilman offers a few arguments for this position. Most notable is that Article 14 does not refer to *mere* hosting, while Article 12 only applies to a *mere* conduit.[893] Besides, the ECJ already held on multiple occasions that Article 14 also applies to the activities of online marketplaces and social media platforms. Most important, however, is the argument that a strict, mere technical interpretation of Article 14 would render the safe harbour useless. Offering exemptions from liability for storing information but not for auxiliary activities such as providing access to information would mean that there is no safe harbour at all. This argument also applies to other activities such as (automatic) content curation or search activities.[894] Article 14(1), as Advocate General Saugmandsgaard Øe puts it, "does not require that that storage is the *sole object, or even the main object*."[895] The second criterion, hosting, and the third, neutrality, are an extension of each other. Some activities that can be more remote from 'core' hosting activities also raise questions about the service provider's neutrality.

The neutrality of hosting service providers: the legal standard

Recital 42 led to a heavy debate about the applicability of the safe harbour laid down in Article 14 to hosting service providers that take on a more active approach to the content of user-provided information. Recital 42 states:

> The exemptions from liability established in this Directive cover only cases where the activity of the information society service provider is limited to the technical process of operating and giving access to a communication network over which information made available by third

888 Par. 12.114 of Riordan, 2016, *The Liability of Internet Intermediaries*, pp. 401-402.

889 Par. 12.113 of Riordan, 2016, *The Liability of Internet Intermediaries*, p. 401.

890 Par. 2.29 of Riordan, 2016, *The Liability of Internet Intermediaries*, p. 34.

891 Par. 2.50 of Riordan, 2016, *The Liability of Internet Intermediaries*, pp. 38-39.

892 Par. 12.113 of Riordan, 2016, *The Liability of Internet Intermediaries*, p. 401.

893 Note 80 of Par. 2.38 of Wilman, 2020, *The Responsibility of Online Intermediaries for Illegal User Content in the EU and the US*, p. 27.

894 Par. 2.37-2.39 of Wilman, 2020, *The Responsibility of Online Intermediaries for Illegal User Content in the EU and the US*, pp. 27-28.

895 Opinion of Advocate General Saugmandsgaard Øe of 16 July 2020 in *C-682/18, C-683/18, Frank Peterson v. Google LLC, YouTube LLC, YouTube Inc., Google Germany GmbH (C-682/18) and Elsevier Inc. v. Cyando AG (C-683/18)*, ECLI:EU:C:2020:586, in particular Rec. 145.

> parties is transmitted or temporarily stored, for the sole purpose of making the transmission more efficient; this activity is of a mere technical, automatic and passive nature, which implies that the information society service provider has neither knowledge of nor control over the information which is transmitted or stored.[896]

Breaking down Recital 42, the activities offered a safe harbour are easy the spot. The phrase "technical process of operating and giving access to a communication network over which information made available by third parties is transmitted" clearly sees to mere conduit activities. The second phrase, "temporarily stored, for the sole purpose of making the transmission more efficient", clearly refers to caching activities. The third fragment, however, raises some questions. Recital 42 states that "this activity is of a mere technical, automatic and passive nature, which implies that the information society service provider has neither knowledge of nor control over the information which is transmitted or stored." This fragment seems to apply to all activities: mere conduit, caching – and hosting activities – at least, this is how the ECJ interpreted the recital in *Google France*.[897]

Interesting to note is that the ECJ summarises "mere technical, automatic and passive nature" as a requirement "to examine whether the role played by that service provider is neutral".[898] Advocate General Poiares Maduro made a similar but even more far-stretching argument for the neutrality requirement in the opinion. The Advocate General argued that displaying advertisements by a search engine provider is not a neutral activity because the provider "has a direct interest in internet users clicking on the ads' links".[899] The ECJ did not adopt this argument in *Google France* but introduced the "mere technical, automatic and passive nature"-test for the safe harbour offered to hosting activities under Article 14.[900]

In *L'Oréal v. eBay* (2011), the ECJ made a minor correction to the *Google France* ruling. Central to *L'Oréal v. eBay* was whether an online marketplace such as eBay could benefit from the safe harbour offered by Article 14. The ECJ held that Article 14 of the e-Commerce Directive does not apply to a hosting provider that has "played an active role of such a kind as to give it knowledge of, or control over, the data relating to those offers for sale."[901] Especially important was that eBay fulfilled an active role with respect to the content of user-provided information. Since eBay "processes the data entered by its customer-sellers" while "[i]n some cases, eBay also provides assistance intended to optimise or promote certain offers for sale."[902]

Following the rule laid down in *Google France,* only neutral activities of a "mere technical, automatic and passive nature"[903] could rely on safe harbours.[904] In *L'Oréal v. eBay,* the ECJ refers

---

[896] Recital 42 of Directive 2000/31/EC (*Directive on electronic commerce*).

[897] Judgement of the Court (Grand Chamber) in *C-236/08, C-237/08 and C-238/08* (*Google France*), in particular Rec. 113-114.

[898] Judgement of the Court (Grand Chamber) in *C-236/08, C-237/08 and C-238/08* (*Google France*), in particular Rec. 114.

[899] Opinion of Advocate General Poiares Maduro of 22 September 2009 in *C-236/08, C-237/08 and C-238/08, Google France SARL and Google Inc. v. Louis Vuitton Malletier SA*, ECLI:EU:C:2009:569, in particular Rec. 143-145.

[900] Par. 2.40-2.41 of Wilman, 2020, *The Responsibility of Online Intermediaries for Illegal User Content in the EU and the US*, p. 28.

[901] Judgement of the Court (Grand Chamber) in *C-324/09* (*L'Oréal v. eBay*), in particular Rec. 116.

[902] Judgement of the Court (Grand Chamber) in *C-324/09* (*L'Oréal v. eBay*), in particular Rec. 114.

[903] Recital 42 of Directive 2000/31/EC (*Directive on electronic commerce*).

[904] Judgement of the Court (Grand Chamber) in *C-236/08, C-237/08 and C-238/08* (*Google France*), in particular Rec. 114.

to *Google France* as if it held that Article 14 requires a "merely technical and automatic processing". A hosting service provider may not play "an active role of such a kind as to give it knowledge of, or control over, those data".[905] Wilman notes that "passive" has disappeared in *L'Oréal v. eBay*. Wilman argues that the "passivity"-requirement is replaced by the requirement for the provider to refrain from an active role that gives "knowledge of, or control over" the information provided by users. While what Wilman calls "strict passivity" is not a prerequisite, a provider merely should not become too active that it gained "knowledge of, or control over" the content of user-provided information.[906] *L'Oréal v. eBay*, as a Grand Chamber ruling, is still the focal point of the safe harbour liability discussion within the European Union. Wilman notes that the arguments made in *L'Oréal v. eBay* seem to be followed but that there is a need for more clarity.[907]

In 2021, the ECJ was offered the opportunity to offer some clarification in a request for a preliminary ruling.[908] Advocate General Saugmandsgaard Øe delivered an opinion on the matter on 16 July 2020.[909] According to the Advocate General, the mere possibility of control is not enough. Instead, it refers to activities by which an internet intermediary "is deemed to acquire *intellectual control of that content*."[910] Saugmandsgaard Øe, for example, argues that this is the case when an intermediary selects the content, presents content as its own or is otherwise involved in the content of the user.[911] According to Saugmandsgaard Øe, this is not the case when an intermediary presents information that originated from others in a specific way. Only when the provider provides "individual assistance" is this different. In such a case, a hosting service provider takes on an active role and thus forfeits its safe harbour.[912] Likewise, providers that index, provide search functionalities, and recommend third-party content to users benefit from the safe harbour. Saugmandsgaard Øe stresses that service providers may decide when information is shown in the search functionality and when information is recommended to users. As long as this process is automated, a provider does not forfeit the safe harbour of Article 14.[913]

After this opinion, on 22 June 2021, the Grand Chamber of the ECJ ruled that a provider that deploys "technological measures aimed at detecting […] content which may infringe copyright, does not mean that, by doing so, that operator plays an active role giving it knowledge of and control".[914] Besides, the fact that a service provider

> automatically indexes content uploaded to that platform, that that platform has a search function and that it recommends videos on the basis of users' profiles or preferences is not a sufficient ground for the conclusion that that operator has 'specific' knowledge of illegal activities carried out on that platform or of illegal information stored on it.[915]

---

[905] Judgement of the Court (Grand Chamber) in *C-324/09* (*L'Oréal v. eBay*), in particular Rec. 113.

[906] Par. 2.45 of Wilman, 2020, *The Responsibility of Online Intermediaries for Illegal User Content in the EU and the US*, p. 30.

[907] Par. 2.49 of Wilman, 2020, *The Responsibility of Online Intermediaries for Illegal User Content in the EU and the US*, p. 32.

[908] Judgment of the Court (Grand Chamber) of 22 June 2021 in *C-682/18 and C-683/18, Frank Peterson v. Google LLC and Others and Elsevier Inc. v Cyando AG*, ECLI:EU:C:2021:503.

[909] Opinion of Advocate General Saugmandsgaard Øe in *C-682/18, C-683/18* (*YouTube*).

[910] Opinion of Advocate General Saugmandsgaard Øe in *C-682/18, C-683/18* (*YouTube*), in particular Rec. 152.

[911] Opinion of Advocate General Saugmandsgaard Øe in *C-682/18, C-683/18* (*YouTube*), in particular Rec. 152.

[912] Opinion of Advocate General Saugmandsgaard Øe in *C-682/18, C-683/18* (*YouTube*), in particular Rec. 156-159.

[913] Opinion of Advocate General Saugmandsgaard Øe in *C-682/18, C-683/18* (*YouTube*), in particular Rec. 160 and 162.

[914] Judgment of the Court (Grand Chamber) in *C-682/18 and C-683/18* (*YouTube*), in particular Rec. 109.

[915] Judgment of the Court (Grand Chamber) in *C-682/18 and C-683/18* (*YouTube*), in particular Rec. 114.

The ECJ, thus, reinforces the standard set out in *L'Oréal v. eBay* by providing clarity that activities such as indexing, search functionalities, and recommendations are also protected under the safe harbour of Article 14(1). Such conduct does not render them "active" in such a manner that they have "to give it knowledge of or control over the content" [916] of user-provided information. However, the ECJ also diverged from *L'Oréal v. eBay* by reintroducing or mentioning the passivity requirement once again. According to the ECJ in *YouTube* and *Cyando*:

> it is necessary to examine whether the role played by that operator is neutral, that is to say, whether its conduct is merely technical, automatic and passive, which means that it has no knowledge of or control over the content it stores, or whether, on the contrary, that operator plays an active role that gives it knowledge of or control over that content […].[917]

While the criterion is that the hosting service provider forfeits its safe harbour when its activities give "knowledge of or control over the content it stores", the ECJ does not clarify whether the Directive or the ECJ expects hosting service providers to remain neutral or passive.

Passive or neutral hosting service providers: legal expectations

McGonagle summarises the e-Commerce Directive to establish "a 'safe harbour' regime for passive intermediaries".[918] Did the ECJ establish a rule of neutrality or a rule of passivity? The rule laid down in the court its case law based on the Directive is that hosting service providers cannot become too active that it gives them knowledge of or control over the content of user-provided information.[919] Is a requirement for passivity in this context the same as a requirement for neutrality? The Directive seems to adopt passivity as a requirement, while the ECJ refers to this requirement as neutrality. Neutrality suggests that the provider has to treat all user-provided information equally, while passivity sees how the provider handles the content of this information.

In *YouTube* and *Cyando*, the ECJ seems to require hosting service providers to remain neutral. However, the ECJ explains neutrality as "merely technical, automatic and passive" conduct which the ECJ explains as "that it has no knowledge of or control over the content it stores".[920] Neutrality thus does not mean that the provider is not allowed to make content-based choices concerning the information provided by its users by deploying technological or automatic means.[921] Neutrality also does not mean that the provider does not have an interest in the success of specific instances of information (such as advertisements).[922] Neutrality understood this way, in my view, has more to do with intellectual distance from specific, individual instances of information. The ECJ seems to use "neutrality" to refer to the use of technological and automatic

---

[916] Judgment of the Court (Grand Chamber) in *C-682/18 and C-683/18* (*YouTube*), in particular Rec. 117.

[917] Judgment of the Court (Grand Chamber) in *C-682/18 and C-683/18* (*YouTube*), in particular Rec. 106.

[918] McGonagle, 2020, 'Free Expression and Internet Intermediaries: The Changing Geometry of European Regulation', p. 475.

[919] See, for this development of this rule, Judgement of the Court (Grand Chamber) in *C-236/08, C-237/08 and C-238/08* (*Google France*), in particular Rec. 114; Judgement of the Court (Grand Chamber) in *C-324/09* (*L'Oréal v. eBay*), in particular Rec. 116; Judgment of the Court (Grand Chamber) in *C-682/18 and C-683/18* (*YouTube*), in particular Rec. 106.

[920] Judgment of the Court (Grand Chamber) in *C-682/18 and C-683/18* (*YouTube*), in particular Rec. 106.

[921] Judgment of the Court (Grand Chamber) in *C-682/18 and C-683/18* (*YouTube*), in particular Rec. 114.

[922] Judgement of the Court (Grand Chamber) in *C-236/08, C-237/08 and C-238/08* (*Google France*), in particular Rec. 111-116.

means and "passivity" to no human involvement.[923] Such an explanation, however, is difficult to understand because technical means are not necessarily neutral because, by the use of automatic means, there are recommendations made based on the content of the provided information.

Passivity seems to refer to human passivity with respect to individual instances of user-provided information. Human involvement with specific information leading to "intellectual control" or "knowledge of, or control over" may cause the provider to lose the safe harbour.[924] According to Wilman, the standard laid down in *L'Oréal v. eBay* is result-oriented: not what the provider does, but whether it leads to knowledge or control is decisive.[925] In *YouTube* and *Cyando,* the ECJ reinforced this standard.[926]

Expecting hosting service providers to be neutral, as discussed under the US Approach, suggests that providers may not intervene in user-provided information based on the content of the information.[927] Requiring neutrality, in this view, means that providers cannot moderate or curate information based on its content. Content-based interventions are the opposite of neutrality. Neutrality, thus, means passivity with respect to the content of user-provided information.

The Directive, nor the ECJ, seems to expect complete neutrality from hosting service providers. The EC even tries to reconcile neutrality with an active stance from providers to combat illegal content such as hate speech.[928] Hosting service providers, however, are expected to keep some distance from the content that users provide to the service. The Directive, thus, expects hosting service providers to remain passive with respect to the user-provided information. Active hosting service providers cannot count on the safe harbours of the Directive when this active role leads them to have knowledge or control over the content of user-provided information. Unlike Section 230, Article 14 thus does not codify the expectation that hosting service providers fulfil roles comparable with traditional editors. Hosting service providers are expected to remain far more passive in how they handle the content provided to them than traditional editors.

4.1.2 For what does Article 14 protect service providers?

Article 14, like Section 230, offers some protection for liability from the content of user-provided information. Unlike Section 230, possible exposure to liability for the content of user-provided information by losing the safe harbour is made dependent upon fulfilling multiple criteria. First, Article 14 requires that the provider is held liable for "information stored at the request of a recipient of the service",[929] which means that Article 14 does not offer safe harbour protection for the content of the information the provider created or commissioned itself. Besides, Article 14 requires that the hosting service provider "does not have actual knowledge of illegal activity or

---

[923] Judgment of the Court (Grand Chamber) in *C-682/18 and C-683/18* (*YouTube*), in particular Rec. 106.

[924] It is hard to see, how by automatic means, such involvement can be reached. For examples of too active involvement, see Par. 2.51-2.5.2 of Wilman, 2020, *The Responsibility of Online Intermediaries for Illegal User Content in the EU and the US*, p. 33.

[925] Par. 2.55 of Wilman, 2020, *The Responsibility of Online Intermediaries for Illegal User Content in the EU and the US*, pp. 34-35.

[926] Judgment of the Court (Grand Chamber) in *C-682/18 and C-683/18* (*YouTube*), in particular Rec. 106.

[927] Fishback, 'How the Wolf of Wall Street Shaped the Internet: A Review of Section 230 of the Communications Decency Act', *Texas Intellectual Property Law Journal*, 2020, p. 277.

[928] Communication COM(2017)555 final, pp. 7 and 10-12; Recital 26 of Commission Recommendation (EU) 2018/334.

[929] Article 14(1) of Directive 2000/31/EC (*Directive on electronic commerce*).

information" and, with respect to "claims for damages, is not aware of facts or circumstances from which the illegal activity or information is apparent".[930] A third requirement laid down in Article 14 is that "upon obtaining such knowledge or awareness, acts expeditiously to remove or to disable access to the information."[931]

Unlike Section 230, the e-Commerce Directive does not have clear protection in place for liability from moderation decisions. While Article 15 offers protection from a general obligation to monitor for user-provided information with illegal content, there is a lack of clarity on whether hosting service providers are protected from 'wrongful moderation':[932] for both under- and overregulating user-provided information.

Not for content of the provider created, commissioned, or requested itself

A provider cannot rely on the safe harbour offered by Article 14 for liability from the content that it created itself. For example, a newspaper that publishes the content of its offline newspaper also online does not magically become a provider because it publishes its content online. As the ECJ argued:

> since a newspaper publishing company which posts an online version of a newspaper on its website has, in principle, knowledge about the information which it posts and exercises control over that information, it cannot be considered to be an 'intermediary service provider' within the meaning of Articles 12 to 14 of Directive 2000/31[933]

When the provider has created the content of the information itself, it cannot rely on Article 14 because the provider does not seek to rely on the safe harbour "for the information stored at the request of a recipient of the service". Besides, the provider can hardly be argued to have no knowledge or awareness of the content of the provided information. More fundamental, however, is that the provider that creates information of its own, as Riordan puts it, "acting as a primary party" and not as a provider.[934] The same is, according to Article 14(2) true for "when the recipient of the service is acting under the authority or the control of the provider."[935] Wilman, for example, argues that this might be the case when the illegal content of user-provided information is a result of the contractual relationship between the provider and the user that provided the information.[936]

Due to the neutrality requirement as discussed in Paragraph 4.1.1, hosting service providers that become too active in requesting or stimulating information with illegal content may forfeit their safe harbour. As Riordan notes, a provider that actively requests or stimulates users to provide information with illegal or unlawful content would not be regarded as a neutral service provider.[937] So-called bad actors that dedicate themselves to the ugly and the bad therefore are unlikely to rely

---

[930] Article 14(1)(a) of Directive 2000/31/EC (*Directive on electronic commerce*).

[931] Article 14(1)(b) of Directive 2000/31/EC (*Directive on electronic commerce*).

[932] Klos, "Wrongful moderation': Aansprakelijkheid van internetplatforms voor het beperken van de vrijheid van meningsuiting van gebruikers', *Nederlands Juristenblad*, 2020/2976.

[933] Judgment of the Court (Seventh Chamber) of 11 September 2014 in *C-291/13, Sotiris Papasavvas v. O Fileleftheros Dimosia Etaireia Ltd and Others*, ECLI:EU:C:2014:2209, in particular Rec. 45.

[934] Par. 12.129 of Riordan, 2016, *The Liability of Internet Intermediaries*, pp. 404-405.

[935] Article 14(2) of Directive 2000/31/EC (*Directive on electronic commerce*).

[936] Par. 2.25 of Wilman, 2020, *The Responsibility of Online Intermediaries for Illegal User Content in the EU and the US*, p. 21.

[937] Par. 12.154 of Riordan, 2016, *The Liability of Internet Intermediaries*, p. 409.

on the safe harbour. A provider actively promoting its service for illegal goods and services or having in place a business model for such content may forfeit its safe harbour under Article 14.[938]

Knowledge and awareness-based based-liability

Article 14 distinguishes between knowledge and awareness. Awareness, as Article 14(1)(a) reads, only applies to claims for damages. For other legal action, such as criminal charges, only actual knowledge suffices according to Article 14(1)(a).[939] The distinction is thus between the type of remedies (damages and criminal law remedies) and not what kind of violation the content causes. For example, an individual may sue successfully for damages, while a prosecutor might not be as successful. Due to the consequences that are tied to a criminal conviction, such a heightened standard for criminal charges is only logical.[940] The safe harbour of Article 14 extends to all types of categories of content while distinguishing between actual knowledge and – in the case of damages – awareness.

According to Wilman, the 'actual knowledge' test is subjective, while the 'awareness' test is objective. The knowledge test requires proving what the hosting service provider did know (subjective), while the awareness test also requires proving what the provider should know (objective).[941] A provider can be said to have actual knowledge when the provider has sufficient knowledge (subjective test) that a specific instance of user-provided information contains illegal content. Normally, a hosting service provider gains actual knowledge through either 1) moderation efforts out of its own initiative or 2) through a notice pointing out that the illegal content exists.[942] In other words, for "actual knowledge", the provider must have "specific" knowledge. A provider that has knowledge that the service is used to spread illegal content has only general knowledge.[943] Whether the provider has actual knowledge after encountering user-provided information with illegal content or after receiving a notice depends on whether the illegality can be determined by the service provider. In the context of the hosting provider engaging in its own investigations, the EC argues that (voluntary) measures against content are at least fitting

> where the illegal character of the content has already been established or where the type of content is such that contextualisation is not essential. It can also depend on the nature, scale and purpose of the envisaged measures, the type of content at issue, on whether the content has been notified by law enforcement authorities or Europol and on whether action had already been taken in respect of the content because it is considered to be illegal content.[944]

Advocate General Saugmandsgaard Øe, in the context of another Directive, argues that hosting service providers should "only be required to filter and block information which has first been established by a court as being illegal or, otherwise, information the unlawfulness of which is obvious from the outset, that is to say, it is manifest, without, inter alia, the need for

---

[938] Van Hoboken, et al., 2018, *Hosting intermediary services and illegal content online*, pp. 24-25.

[939] Par. 2.59 of Wilman, 2020, *The Responsibility of Online Intermediaries for Illegal User Content in the EU and the US*, p. 36.

[940] Par. 2.62-2.63 of Wilman, 2020, *The Responsibility of Online Intermediaries for Illegal User Content in the EU and the US*, pp. 37-38.

[941] Par. 2.60-2.61 of Wilman, 2020, *The Responsibility of Online Intermediaries for Illegal User Content in the EU and the US*, p. 37.

[942] Par. 2.67-2.70 of Wilman, 2020, *The Responsibility of Online Intermediaries for Illegal User Content in the EU and the US*, pp. 39-40.

[943] Van Hoboken, et al., 2018, *Hosting intermediary services and illegal content online*, p. 38.

[944] Recital 25 of Commission Recommendation (EU) 2018/334, p. 54.

contextualisation."[945] As Saugmandsgaard Øe suggests, the hosting service provider must be able to establish whether the content of the information is illegal in such a manner that it raises actual knowledge. Actual knowledge, thus, is not easy to reach. A lower standard is offered by "awareness".

According to Wilman, "awareness" can be understood as "constructive knowledge: not only what the intermediary did know counts, but also what the intermediary should have known.[946] The objective "awareness" test is derived from *eBay* in which the ECJ laid down the standard that Article 14 does not shield service providers that are "aware of facts or circumstances on the basis of which a diligent economic operator should have identified the illegality in question".[947] According to Riordan, this standard means that a hosting service provider cannot call in the safe harbour against repeated infringements when the provider has knowledge of the facts and circumstances that this content is, in fact, illegal.[948] According to Riordan, "actual knowledge of the facts or circumstances" holds a "hybrid subjective-objective test". The "hybrid subjective-objective test" means that Article 14 does not offer a safe harbour to a hosting service provider that subjectively identified some facts or circumstances on which the provider objectively, as "a diligent economic operator", could have assessed the illegality of the content of user-provided information.[949]

A service provider, however, cannot be obligated to search for illegal content or for facts and circumstances which may help to determine whether the content is illegal. Article 15 of the Directive prohibits such a general obligation to monitor. However, Article 14 does not stand in the way of court orders "to terminate or prevent an infringement", nor does Article 14 "affect the possibility for Member States of establishing procedures governing the removal or disabling of access to information."[950] However, Article 15 limits this possibility to the extent that hosting service providers cannot be required to actively search for illegal content or for facts or circumstances that may provide an indication for such illegality. Article 15(1) in full reads that:

> Member States shall not impose a general obligation on providers, when providing the services covered by Articles 12, 13 and 14, to monitor the information which they transmit or store, nor a general obligation actively to seek facts or circumstances indicating illegal activity.[951]

Both Wilman and Riordan argue that Article 14(1) must be read in conjunction with Article 15(1).[952] Article 14, in terms of awareness, requires that the "facts or circumstances" make the illegal nature of information "apparent".[953] According to Riordan, "awareness" does not mean constructive knowledge.[954] Constructive knowledge would leave too much to the service provider and indirectly impose an obligation to look for facts and circumstances. Such a standard would be

---

[945] Opinion of Advocate General Saugmandsgaard Øe of 15 July 2021 in *C-401/19, Poland v. Parliament and Council*, ECLI:EU:C:2021:613, in particular Rec. 198.

[946] Par. 2.61 of Wilman, 2020, *The Responsibility of Online Intermediaries for Illegal User Content in the EU and the US*, p. 37.

[947] Judgement of the Court (Grand Chamber) in *C-324/09* (*L'Oréal v. eBay*), in particular Rec. 120. Also see, Wilman, 2020, *The Responsibility of Online Intermediaries for Illegal User Content in the EU and the US*, p. 37.

[948] Par. 12.137 of Riordan, 2016, *The Liability of Internet Intermediaries*, p. 406.

[949] Par. 12.137 of Riordan, 2016, *The Liability of Internet Intermediaries*, p. 406.

[950] Article 14(3) of Directive 2000/31/EC (*Directive on electronic commerce*).

[951] Article 15(1) of Directive 2000/31/EC (*Directive on electronic commerce*).

[952] Par. 2.64 of Wilman, 2020, *The Responsibility of Online Intermediaries for Illegal User Content in the EU and the US*, p. 38.

[953] Article 14(1)(a) of Directive 2000/31/EC (*Directive on electronic commerce*).

[954] Par. 12.130 of Riordan, 2016, *The Liability of Internet Intermediaries*, p. 405.

contrary to Article 15(1). Riordan lays the threshold at actual knowledge of unlawful activities that have already occurred or are still ongoing.[955] A "case of borderline illegality does generally not seem sufficient to lead to "awareness" within the meaning of Article 14(1)(a)", Wilman argues.[956] While contested,[957] establishing that the content of information is, for example, defamatory or copyright protected is not enough to assume knowledge. According to Riordan, the hosting service provider must become knowledgeable about the fact and circumstances that makes it possible for the provider to assess its illegality.[958] Assuming "awareness" based on a standard that is lower than knowledge of manifestly illegal content of user-provided information would impose an obligation that would contradict Article 15(1) because the provider is then required to engage in information-gathering, which holds a general obligation to monitor.[959]

A notification, to give either knowledge or awareness, must be precise and substantiated.[960] In other words, the notification must point out where the information can be found and why this information should be considered to be illegal content.[961] When a notice is considered sufficient substantiated is subject to discussion. Wilman notices that a trade-off must be made between the position of the notifier and the position of the service provider. For notice and takedown procedures to be effective, it can be hardly expected from users or organisations submitting notices that they adhere to the same criteria as a public prosecutor. However, hosting service providers should not be considered to have the time and resources to launch an elaborate investigation. Article 15(1) of the Directive stands in the way of requiring intermediaries to fill the blanks in insufficiently substantiated notices.[962] For imprecise notifications, a similar argument can be made. When is a notification sufficient precise? According to Wilman, "the most practical and logical manner seems the provision of either a link to the item of content or the relevant uniform resource locator (URL)."[963] In other words, when the service provider must search its service for the user-provided information with illegal content, a notification is clearly not sufficiently precise. When the notifier fails to make clear why content should be considered illegal or unlawful, the notice is not sufficiently substantiated.

Article 15, however, does not prohibit courts from obligating service providers to monitor for new instances of information that contains identical or equivalent illegal content.[964] In *Glawischnig-Piesczek*, the ECJ ruled that Article 15 does not stand in the way of a court order requiring to prevent new instances of user-provided information with defamatory content from being uploaded as long

---

[955] Par. 12.130 of Riordan, 2016, *The Liability of Internet Intermediaries*, p. 405.

[956] Par. 2.63 of Wilman, 2020, *The Responsibility of Online Intermediaries for Illegal User Content in the EU and the US*, p. 38.

[957] For example, for violations of intellectual property rights, some authors argue that general knowledge is sufficient to establish its illegality, see Van Hoboken, et al., 2018, *Hosting intermediary services and illegal content online*, pp. 38-39.

[958] Par. 12.131-132 of Riordan, 2016, *The Liability of Internet Intermediaries*, p. 405.

[959] Par. 2.64 of Wilman, 2020, *The Responsibility of Online Intermediaries for Illegal User Content in the EU and the US*, p. 38.

[960] Judgement of the Court (Grand Chamber) in *C-324/09* (*L'Oréal v. eBay*), in particular Rec. 122.

[961] Par. 2.69 of Wilman, 2020, *The Responsibility of Online Intermediaries for Illegal User Content in the EU and the US*, pp. 39-40.

[962] Par. 2.71-2.73 of Wilman, 2020, *The Responsibility of Online Intermediaries for Illegal User Content in the EU and the US*, pp. 40-42.

[963] Par. 2.74 of Wilman, 2020, *The Responsibility of Online Intermediaries for Illegal User Content in the EU and the US*, p. 42.

[964] Opinion of Advocate General Saugmandsgaard Øe in *C-401/19* (*Poland v. Parliament and Council*), in particular Rec. 200.

> the monitoring of and search for information which it requires are limited to information containing the elements specified in the injunction, and its defamatory content of an equivalent nature does not require the host provider to carry out an independent assessment, since the latter has recourse to automated search tools and technologies.[965]

In other words, Article 15 does not stand in the way of a court order that requires the hosting service provider to take down – and keep down – user-provided information with defamatory content as long as the provider is not required to review each instance of content independently. The provider must be able to rely on automated filtering techniques without the obligation to assess the illegality of the content of user-provided information piece by piece.

In sum, to forfeit Article 14 safe harbour protections, the hosting service provider must have some (subjective) actual knowledge of the illegal content of user-provided information and, in the case of "claims for damages", awareness of facts or circumstances (subjective) from which "a diligent economic operator" (objective) could make up that the content of user-provided information is indeed illegal. This, however, does not mean that hosting service providers can be obligated to search for instances of information with illegal content or facts and circumstances that are required to assess its illegality.

Prevent liability by removing user-provided information

The second shield offered by the safe harbour is thus the possibility to prevent liability by removing or disabling access to the content of user-provided information after gaining knowledge or awareness of its illegality. Article 14(1)(b) extends the safe harbour to hosting service providers that obtain "such knowledge or awareness" as long as the provider "acts expeditiously to remove or to disable access to the information".[966]

According to Riordan, this second part of the Directive leaves some uncertainty in place for hosting service providers. At first, it is not clear what "expeditious" removal means. Multiple factors related to the content and the hosting service providers are relevant to discussing whether the action undertaken by the provider was indeed "expeditious". A very precise and substantiated notification pointing out illegal content that causes enormous harm every minute it remains up has a quicker removal timeframe than content that is not as harmful. Besides, the capacity and the conduct of the service provider may count towards the question of whether the removal was "expeditious".[967] According to Riordan, the standard is objective: the question is what can "be expected of a reasonable economic operator supplying the same information society services."[968] According to Wilman, this normally comes down to a few days.[969]

The second uncertainty the Directive leaves is whether the action or the result of the action counts. Must the removal be successful, or is it required that, as Riordan puts it, "the service provider must take reasonable steps"?[970] According to Wilman, the standard laid down in *L'Oréal v. eBay*[971] is also applicable here. Service providers are not required to take down every instance of

---

[965] Judgement of the Court (Third Chamber) in *C-18/18* (*Glawischnig-Piesczek*), in particular Rec. 46.

[966] Article 14(1)(b) of Directive 2000/31/EC (*Directive on electronic commerce*).

[967] Par. 12.144-145 of Riordan, 2016, *The Liability of Internet Intermediaries*, p. 408; Par. 2.76 of Wilman, 2020, *The Responsibility of Online Intermediaries for Illegal User Content in the EU and the US*, p. 43.

[968] Par. 12.145 of Riordan, 2016, *The Liability of Internet Intermediaries*, p. 408.

[969] Par. 2.76 of Wilman, 2020, *The Responsibility of Online Intermediaries for Illegal User Content in the EU and the US*, p. 43.

[970] Par. 12.146 of Riordan, 2016, *The Liability of Internet Intermediaries*, p. 408.

[971] Judgement of the Court (Grand Chamber) in *C-324/09* (*L'Oréal v. eBay*), in particular Rec. 124.

information of which someone argues that its content is illegal but are as "diligent economic operators" required to review whether the notification is sufficiently substantiated.[972] Providers are, as Wilman puts it, "not expected to carry out *broad and burdensome* investigations in situations where the notices are not precise or substantiated."[973] Tied to this question is whether these reasonable steps should see to illegal content that is pointed out to the hosting service provider or also repeated violations when these violations see to the same right and are violated by the same user.[974] Requiring expeditious removal of newly uploaded content could violate Article 15 (general prohibition to monitor)[975] therefore, Riordan views newly uploaded information as "a separated unlawful act […] and […] a fresh legal wrong."[976] Wilman notes that *Glawischnig-Piesczek*[977] suggests a different standard. In this case, the ECJ argued that the hosting service provider could be required to take down similar instances of information with defamatory content as long as the provider was not required to review each instance of content independently.[978] *Glawischnig-Piesczek*, however, deals with a court order which knows its own regime under Article 14(3) "requiring the service provider to terminate or prevent an infringement".[979] This standard thus does not apply to notices from normal users under Article 14(1)(b).[980]

The third uncertainty is that Article 14 refers to illegal content of information but also to "unlawful activity". Article 14(1)(b), however, only refers to the removal or inaccessibility of information, which leaves the question of how the hosting service provider should handle unlawful activities that cannot be ended by taking such action.[981] A reasonable explanation of what Article 14(1)(b) aims to express is that hosting service providers can only end illegal activity in case they can remove or make inaccessible the content of the information in question. Besides, a reasonable explanation is that the illegality must be tied to a specific instance of information. Sometimes the content itself is not illegal (for example, in the case of copyright infringement), but the fact that the legal content is shared without the authorisation of the rightsholder makes not the content itself illegal but the activities tied to the information with copyright-protected content.

4.1.3 What does the Directive encourage?

As noted, Article 14, in conjunction with Article 15 of the Directive, protects hosting service providers from liability for the content of user-provided information. Article 15 prohibits EU member states from imposing a general obligation to monitor for illegal content. Service providers are thus not required to actively monitor for illegal content on their service. Article 14(1)(a), in its turn, protects hosting service providers from liability arising from the content of user-provided information as long as the hosting service provider does not have any knowledge or awareness of

---

[972] Par. 2.72 of Wilman, 2020, *The Responsibility of Online Intermediaries for Illegal User Content in the EU and the US*, pp. 41-42.

[973] Par. 2.71 of Wilman, 2020, *The Responsibility of Online Intermediaries for Illegal User Content in the EU and the US*, pp. 40-41.

[974] Par. 12.146 of Riordan, 2016, *The Liability of Internet Intermediaries*, p. 408. See, Judgement of the Court (Third Chamber) in *C-18/18* (*Glawischnig-Piesczek*), in particular Rec. 41.

[975] Par. 12.147 of Riordan, 2016, *The Liability of Internet Intermediaries*, p. 408.

[976] Par. 12.148 of Riordan, 2016, *The Liability of Internet Intermediaries*, p. 408.

[977] Judgement of the Court (Third Chamber) in *C-18/18* (*Glawischnig-Piesczek*).

[978] Judgement of the Court (Third Chamber) in *C-18/18* (*Glawischnig-Piesczek*), in particular Rec. 45-46.

[979] Article 14(3) of Directive 2000/31/EC (*Directive on electronic commerce*).

[980] Par. 2.73 of Wilman, 2020, *The Responsibility of Online Intermediaries for Illegal User Content in the EU and the US*, p. 42.

[981] Par. 12.149 of Riordan, 2016, *The Liability of Internet Intermediaries*, p. 408.

its illegal content. When the provider does encounter user-provided information with illegal content, it still can rely on the safe harbour since Article 14(1)(b) offers an exemption when the provider "expeditiously" removes or disable access to the user-provided information with illegal content. Like Section 230, the service provider, however, cannot be active in such a manner that it has control over the content of user-provided information. Unlike Section 230, such an active role should also not give knowledge of the content of user-provided information.

Another way hosting service providers can encounter illegal content raising questions about actual knowledge of awareness is through their own voluntary investigations. As mentioned above, providers, under the Directive, are not required to monitor their services for illegal content. Article 15(1) of the Directive even forbids the Member States to impose such a requirement.[982] While Section 230 was (partly) enacted to take down the barriers for hosting service providers to moderate user-provided information, the question is what Article 14 stimulates hosting service providers to do. What does Article 14 encourage in terms of moderation?

Notice and takedown procedures and removal as a default

As noted, when hosting service providers encounter or are notified of illegal content on its service, this may lead to actual knowledge or awareness of illegal content.[983] Hosting service providers may become liable by losing their safe harbour if they do not expeditiously remove or disable access to the information containing this content.[984] As discussed, Article 14 has some gaps. Article 14 does not specify *how* providers gain such knowledge of illegal content (the Directive does not establish so-called notification and takedown procedures), nor does the Directive specify what should count as "expeditious" removal of information containing illegal content.

As Kuczerawy notices, the e-Commerce Directive does not explicitly provide a notice and takedown procedure.[985] Notice and takedown procedures know some significant benefits over requiring a court order to take down information with illegal content. These procedures are often cheaper, faster, and less burdensome for both the party that submits the notice and the service provider. A notice and takedown procedure, thus, can be viewed as an alternative to the more expensive court order.[986] Notice and takedown procedures know some significant risks as well. The provider has to decide whether a notice is indeed targeting information with illegal content, which introduces the risk of overregulation.[987] A notice in a notice and takedown procedure thus can be abused to target perfectly legal content. Notice and takedown procedures, especially when not legally sanctioned, may fail because the provider simply refuses to follow up on a notice which may lead to underregulation.[988]

In the EU context, the Directive, as Van Hoboken and Keller argue, established such a procedure by making the liability of the hosting service provider dependent on knowledge while

---

[982] Article 15(1) of Directive 2000/31/EC (*Directive on electronic commerce*).

[983] Article 14(1) of Directive 2000/31/EC (*Directive on electronic commerce*).

[984] Article 14(1)(b) of Directive 2000/31/EC (*Directive on electronic commerce*).

[985] A. Kuczerawy, 'From 'Notice and Takedown' to 'Notice and Stay Down': Risks and Safeguards for Freedom of Expression', *in* G. Frosio (Ed.) *The Oxford Handbook of Online Intermediary Liability*, Oxford, Oxford University Press, 2020, doi:10.1093/oxfordhb/9780198837138.013.27, p. 528.

[986] Par. 13.145 of Riordan, 2016, *The Liability of Internet Intermediaries*, p. 443.

[987] Kuczerawy, 2020, 'From 'Notice and Takedown' to 'Notice and Stay Down': Risks and Safeguards for Freedom of Expression', p. 528.

[988] Par. 13.145 of Riordan, 2016, *The Liability of Internet Intermediaries*, p. 443.

leaving the procedure over to the member states since the EU does not harmonise such a procedure.[989] As Kuczerawy notes, only a few member states in the EU provided additional legislation based on Article 14(3) of the Directive[990] (which states that Article 14(1) does not "affect the possibility for Member States of establishing procedures governing the removal or disabling of access to information.").[991] Providing a notice and takedown framework is, according to Kuczerawy, a necessity since this contributes to the foreseeability of the application of legislation as to legal certainty. Uncertainty may pose a risk to the freedom of expression rights of users.[992]

Article 16 of the Directive states that the Member States and the European Commission should encourage the drafting of codes of conduct.[993] While the Directive does not name notice and takedown procedures in Article 16, Article 21(2) names "notice and take down" procedures as necessary to address in the reports in evaluating the Directive.[994] In these codes of conduct, the signees agree to adopt, for example, a notice and takedown procedure for removing or making inaccessible illegal content of user-provided information, and they agree upon removal timeframes.[995] While codes of conduct containing such procedures impose requirements on hosting service providers for when to remove user-provided information with illegal content, these codes of conduct tend to have a blind spot for user rights. These codes of conduct nor the Directive contain a procedure for users that wish to argue against such moderation decisions,[996] while there are concerns that such notice and takedown procedures lead to overregulation.[997]

The Directive leaves little room in what instruments hosting service providers can wield in terms of content moderation remedies. The lack of explicit notice and takedown procedures provides an incentive to impose a remedy that sees to content removal of user-provided information. Goldman, therefore, views Article 14 of the e-Commerce Directive as an example of the binary approach to content moderation remedies: the provider has to choose between leaving the content of user-provided information up or taking the content down.[998] According to Van Hoboken and Keller, hosting service providers could also respond in other ways than removal (for example, by making the content inaccessible), meaning that Article 14 does provide the possibility for a "notice-and-action" regime.[999] However, it is not unlikely that following notification, the provider will turn to removal because this is the safest option.

---

989 Van Hoboken & Keller, 2019, 'Design Principles for Intermediary Liability Laws', p. 2.

990 Kuczerawy, 2020, 'From 'Notice and Takedown' to 'Notice and Stay Down': Risks and Safeguards for Freedom of Expression', p. 530.

991 Article 14(3) of Directive 2000/31/EC (*Directive on electronic commerce*).

992 Kuczerawy, 2020, 'From 'Notice and Takedown' to 'Notice and Stay Down': Risks and Safeguards for Freedom of Expression', p. 530.

993 P.P. Polański, 'Rethinking the notion of hosting in the aftermath of Delfi: Shifting from liability to responsibility?', *Computer Law & Security Review*, Vol. 34, No. 4, 2018, doi:10.1016/j.clsr.2018.05.034, p. 872.

994 Article 21(2) of Directive 2000/31/EC (*Directive on electronic commerce*).

995 For an example on the EU-level, see European Commission, 2016, 'Code of Conduct on Countering Illegal Hate Speech Online', p. 2. For an example on the level of the member states (the Netherlands) see ECP, 'NTD code', *ECP,* 2018, available at noticeandtakedowncode.nl/ntd-code (retrieved on 14 February 2022).

996 Polański, 'Rethinking the notion of hosting in the aftermath of Delfi: Shifting from liability to responsibility?', *Computer Law & Security Review*, 2018, p. 872.

997 McGonagle, 2020, 'Free Expression and Internet Intermediaries: The Changing Geometry of European Regulation', p. 483; Van Hoboken & Keller, 2019, 'Design Principles for Intermediary Liability Laws', p. 4.

998 Goldman, 'Content Moderation Remedies', *Michigan Technology Law Review*, 2021, p. 13.

999 Van Hoboken & Keller, 2019, 'Design Principles for Intermediary Liability Laws'.

Citron, for example, argues that hosting service providers following EU rules may turn to global removal of, for example, hate speech and terrorist content, while these categories of content are not necessarily illegal in other jurisdictions. Citron argues that hosting service providers are not only required to remove illegal content but also to do this quickly.[1000] The Code of Conduct on Countering Illegal Hate Speech Online requires hosting service providers to "review the majority of valid notifications for removal of illegal hate speech in less than 24 hours and remove or disable access to such content, if necessary."[1001] In the case of terrorist content, a competent authority[1002] may issue a removal order to take down this content within an hour.[1003] According to Citron, the requirement of speedy removals – especially in the case of hate speech – contributes to the fact that hosting service providers turn to removal as a default.[1004]

A factor that contributes to global removal is that both the code of conduct and the Regulation require hosting service providers to include a prohibition in their "terms and conditions". The code of conduct, for example, requires hosting service providers to "educate and raise awareness with their users about the types of content not permitted under their rules and community guidelines."[1005] The Regulation requires to "include in its terms and conditions and apply provisions to address the misuse of its services for the dissemination to the public of terrorist content."[1006] As Citron notes, these terms and conditions are often applied globally.[1007]

While notice and takedown procedures, as Van Eecke argues, "involves a moment of reflection, which avoids information always being taken down without further investigation",[1008] the question is whether the e-Commerce Directive offers protection against overregulation by hosting service providers.

(In)voluntary monitoring

Hosting service providers are, as argued, best in the position to regulate the content of user-provided information. The EC acknowledges this in its *Recommendation on measures to effectively tackle illegal content online* (2018) by stating that hosting service providers

> play a particularly important role in tackling illegal content online, as they store information provided by and at the request of their users and give other users access thereto, often on a large scale.[1009]

The EC acknowledges that hosting service providers are in a unique position to counter user-provided information with illegal content. While Article 15 prohibits general obligations to

---

[1000] Citron, 'Extremist Speech, Compelled Conformity, and Censorship Creep', *Notre Dame Law Review*, 2018, p. 1055.

[1001] European Commission, 2016, 'Code of Conduct on Countering Illegal Hate Speech Online', p. 2.

[1002] As designated by the member state according Article 12 of Regulation (EU) 2021/784.

[1003] Article 3 and Annex I of Regulation (EU) 2021/784.

[1004] Citron, 'Extremist Speech, Compelled Conformity, and Censorship Creep', *Notre Dame Law Review*, 2018, p. 1055.

[1005] European Commission, 2016, 'Code of Conduct on Countering Illegal Hate Speech Online', p. 2.

[1006] Article 5(1) of Regulation (EU) 2021/784.

[1007] Citron, 'Extremist Speech, Compelled Conformity, and Censorship Creep', *Notre Dame Law Review*, 2018, pp. 1055-1056.

[1008] Van Eecke, 'Online service providers and liability: A plea for a balanced approach', *Common Market Law Review*, 2011, pp. 1480-1481.

[1009] Commission Recommendation (EU) 2018/334, p. 4.

monitor for illegal or unlawful content, this does not mean that hosting service providers cannot be stimulated (or at least not disincentivised) to screen their services on a voluntary basis.

L'Oréal v. eBay, in this respect, raises some questions. Do providers that screen their services for illegal content become active in such a way that they "have knowledge or control of the data stored"?[1010] While the EC argues that voluntary monitoring for illegal content does not mean that the hosting service provider forfeits its safe harbour altogether, the EC repeats the rule that hosting service providers can maintain their safe harbour when they remove or disable access to the illegal content upon gaining knowledge or awareness.[1011] Kuczerawy rightfully argues that this offers little to no protection to hosting service providers acting as "Good Samaritans" actively screening for illegal content. Hosting service providers that are acting as "Good Samaritans" but fail to (correctly) identify illegal content or remove illegal content are not protected under the safe harbour.[1012]

The question, however, is whether the provider can gain knowledge or awareness by deploying automatic machine monitoring. According to Wilman, the means are not decisive in this respect. According to Wilman, if knowledge or awareness would mean human knowledge or awareness, hosting service providers could simply prevent gaining knowledge or awareness by merely deploying automatic means.[1013] However, it is hard to see how knowledge or awareness can be gained through automatic means. Of course, a computerised filter may malfunction and thus may correctly identify (earlier encountered) illegal content but fail to remove these instances of content. A computer filter that assesses the context of a defamatory statement or hate speech seems years away. Therefore, it is unlikely that a provider can gain "actual knowledge" or "awareness" through automatic computer means for new instances of illegal content or content in which illegality is dependent on a contextual assessment.[1014]

Overregulation and underregulation

In sum, Articles 14 and 15 read in conjunction offer a legal incentive for hosting service providers to refrain from voluntary monitoring. A hosting provider that does not monitor its service for illegal content does not expose itself to liability. When hosting service providers do monitor or receive a notification of illegal content, which gives the provider knowledge of or awareness of information with illegal content, Article 14 may cause hosting service providers to overregulate. Service providers may interpret legal norms overbroad to be on the safe side and may provide an incentive to hosting service providers to remove content globally while disabling access to the content in a specific jurisdiction may be enough.

4.2 Relation with the European Convention on Human Rights[1015]

While in the US, SCOTUS can interpret federal legislation and assess its constitutionality, the relationship between the ECtHR and EU legislation is complicated. The ECtHR does not interpret

---

[1010] Judgement of the Court (Grand Chamber) in *C-324/09* (*L'Oréal v. eBay*), in particular Rec. 123.

[1011] Communication COM(2017)555 final, pp. 11-12.

[1012] Kuczerawy, 2018, 'The EU Commission on voluntary monitoring: Good Samaritan 2.0 or Good Samaritan 0.5?'.

[1013] Par. 2.78 of Wilman, 2020, *The Responsibility of Online Intermediaries for Illegal User Content in the EU and the US*, pp. 44-45.

[1014] See also, Citron, 'Extremist Speech, Compelled Conformity, and Censorship Creep', *Notre Dame Law Review*, 2018, pp. 1054-1055.

[1015] This paragraph is based on Klos, 2020, 'Tackling Online Freedom of Expression: the European Approach'.

EU law. Nor does the ECtHR review how domestic courts apply domestic provisions.[1016] Instead, the ECtHR reviews whether the outcome in a given case is compliant with the ECHR. For the liability of providers, especially Article 10 of the Convention is relevant. Article 10, seeing to freedom of expression rights, reads:

> 1. Everyone has the right to freedom of expression. This right shall include freedom to hold opinions and to receive and impart information and ideas without interference by public authority and regardless of frontiers. This Article shall not prevent States from requiring the licensing of broadcasting, television or cinema enterprises.
>
> 2. The exercise of these freedoms, since it carries with it duties and responsibilities, may be subject to such formalities, conditions, restrictions or penalties as are prescribed by law and are necessary in a democratic society, in the interests of national security, territorial integrity or public safety, for the prevention of disorder or crime, for the protection of health or morals, for the protection of the reputation or rights of others, for preventing the disclosure of information received in confidence, or for maintaining the authority and impartiality of the judiciary.

In two cases, the ECtHR discussed the liability of providers of internet intermediary services under Article 10 of the Convention. In *Delfi AS v. Estonia* (2013), imposing liability on a professionally managed, commercial internet news portal liable for the defamatory content of user comments that were manifestly illegal (hate speech) was considered not a violation of Article 10. The fact that the provider removed the comments the same day upon notification made no difference for the ECtHR.[1017] The Grand Chamber left this verdict untouched in 2015.[1018] In *Magyar Tartalomszolgáltatók Egyesülete and Index.hu Zrt v. Hungary* (2016), the ECtHR revisited internet intermediary liability. In contrast to *Delfi AS v. Estonia*, imposing strict liability on a non-commercial provider for comments that were not clearly illegal while the provider had a notice and takedown procedure in place was considered a violation of Article 10.[1019] What made these cases different from each other?

In *Delfi AS v. Estonia,* the ECtHR weighted that Delfi was a large Estonian internet news portal. On a daily basis, 10,000 comments were posted by users on approximately 330 new articles.[1020] For the ECtHR, Delfi is to be considered a "professionally managed Internet news portal run on a commercial basis which sought to attract a large number of comments on news articles published by it".[1021] Delfi was argued to benefit from user comments because its revenues are based on the number of visits. Comments on news articles help to attract visitors and thus help to increase the revenue of Delfi.[1022] The professional and commercial nature of Delfi's activities is clearly important to the ECtHR.

Besides, the Grand Chamber discussed Delfi its (potential) involvement in the content of the user-provided comments. While Delfi monitored the comment section for inappropriate comments and obscene language, this monitoring was automatically applying a list of forbidden words. This means that comments with defamatory content were not automatically removed when

---

[1016] *Delfi AS v. Estonia* [GC], no. 64569/09, § 127, ECHR 2015-II, 16 June 2015.
[1017] *Delfi AS v. Estonia*, no. 64569/09, § 15, 84-92, 10 October 2013.
[1018] *Delfi AS v. Estonia* [GC], no. 64569/09, ECHR 2015-II, 16 June 2015.
[1019] *Magyar Tartalomszolgáltatók Egyesülete and Index.hu Zrt v. Hungary*, no. 22947/13, § 89-92, 2 February 2016.
[1020] *Delfi AS v. Estonia* [GC], no. 64569/09, § 11, 12-14, ECHR 2015-II, 16 June 2015.
[1021] *Delfi AS v. Estonia* [GC], no. 64569/09, § 144, ECHR 2015-II, 16 June 2015.
[1022] *Delfi AS v. Estonia* [GC], no. 64569/09, § 144, ECHR 2015-II, 16 June 2015.

these comments did not include the words in this list. Defamatory comments, however, would be removed upon receiving a notification.[1023] Delfi made clear that the comments on news articles posted by users did not reflect their own viewpoints and were the responsibility of the users.[1024] The ECtHR nevertheless concluded that Delfi "must be considered to have exercised a substantial degree of control over the comments published on its portal".[1025] Relevant in this respect was that Delfi could (and did) moderate the content of user-provided information. After publication, Delfi was the only one that could edit/remove the content of a comment.[1026] Because Delfi moderated its service, the ECtHR concluded that Delfi's "involvement in making public the comments on its news articles on the Delfi news portal went beyond that of a passive, purely technical service provider."[1027]

The case was resolved around an article titled "SLK Destroyed Planned Ice Road", which was published on 24 January 2006. Around twenty comments of the 185 comments that were posted over two days after publication contained threats or offences directed at the majority shareholder of SLK.[1028] These comments were visible for around six weeks until Delfi took these comments down at the request of the lawyers of the SLK shareholder. However, the SLK shareholder also demanded 32,000 euros in non-pecuniary damages. While Delfi expeditiously removed the defamatory comments, Delfi argued that it could not be required to pay the damages.[1029]

The conditional immunity regime provided by Article 14 of the Directive, in principle, seems to shield Delfi from liability by offering a safe harbour. Delfi cannot be argued to have knowledge of the content of the comments before it received the notification,[1030] while Article 15 of the Directive forbids member states to impose a general obligation to monitor *all* user comments.[1031] Besides, it is questionable whether the ECJ would deny Delfi immunity based on the fact that it was not "a mere technical, automatic and passive"[1032] provider but playing "an active role allowing it to have knowledge or control of the data stored."[1033] Delfi was not that involved in the content of the comments.[1034] The Grand Chamber (rightfully) argued that the ECtHR does not interpret national provisions (Art. 14 of the Directive is transposed in national legislation) – but merely assesses conformity with the Convention.[1035] The ECtHR thus reviews how the application of the Estonian implementation of the Directive holds up against Article 10 of the

---

[1023] Which, is, in principle in line with Article 14(1) and 15(1) of Directive 2000/31/EC (*Directive on electronic commerce*).

[1024] *Delfi AS v. Estonia* [GC], no. 64569/09, § 12-14, ECHR 2015-II, 16 June 2015.

[1025] *Delfi AS v. Estonia* [GC], no. 64569/09, § 145, ECHR 2015-II, 16 June 2015.

[1026] *Delfi AS v. Estonia* [GC], no. 64569/09, § 145, ECHR 2015-II, 16 June 2015.

[1027] *Delfi AS v. Estonia* [GC], no. 64569/09, § 146, ECHR 2015-II, 16 June 2015.

[1028] *Delfi AS v. Estonia* [GC], no. 64569/09, § 16-17, ECHR 2015-II, 16 June 2015.

[1029] *Delfi AS v. Estonia* [GC], no. 64569/09, § 18-20, ECHR 2015-II, 16 June 2015.

[1030] Article 14(1) of Directive 2000/31/EC (*Directive on electronic commerce*).

[1031] Article 15(1) of Directive 2000/31/EC (*Directive on electronic commerce*).

[1032] Recital 42 of Directive 2000/31/EC (*Directive on electronic commerce*).

[1033] Judgement of the Court (Grand Chamber) in *C-324/09* (*L'Oréal v. eBay*), in particular Rec. 123.

[1034] Not everyone agrees, see A. Kuczerawy & P.-J. Ombelet, 'Not so different after all? Reconciling Delfi vs. Estonia with EU rules on intermediary liability', *CiTiP Blog,* 2 July 2015, available at law.kuleuven.be/citip/blog/not-so-different-after-all-reconciling-delfi-vs-estonia-with-eu-rules-on-intermediary-liability (retrieved on 15 February 2022).

[1035] *Delfi AS v. Estonia* [GC], no. 64569/09, § 127, ECHR 2015-II, 16 June 2015.

ECHR. According to Brunner, this may be the reason the ECtHR does not discuss economic arguments for not imposing liability for user-provided information with illegal content on Delfi.[1036]

After a lengthy national procedure in Estonia, Delfi was required to pay 320 euros in non-pecuniary damages.[1037] Delfi, arguing that its freedom of expression rights was violated, brought the case to the ECtHR. The Chamber delivered its judgment on 10 October 2013, holding that there was no violation of Article 10 of the ECHR. Delfi appealed for a rehearing of the case by seventeen judges (known as the Grand Chamber). Delfi argued that it could not be held liable as a publisher of the content of the user-provided comments.[1038] The Grand Chamber delivered its judgment on the matter on 16 June 2015.[1039]

The ECtHR thus primarily balances the rights laid down in the ECtHR. In this case, Article 10 with Article 8 (the right to respect for private life).[1040] The ECtHR does not seem to fear the adverse effects on freedom of expression rights. The ECtHR argues that "a large news portal's obligation to take effective measures to limit the dissemination of hate speech and speech inciting violence – the issue in the present case – can by no means be equated to 'private censorship'."[1041] A notice and takedown procedure, as followed by Delfi, was not sufficient, according to the ECtHR because

> the ability of a potential victim of hate speech to continuously monitor the Internet is more limited than the ability of a large commercial Internet news portal to prevent or rapidly remove such comments.[1042]

While the ECtHR argues that in many cases, notice and takedown procedures would suffice, in *Delfi,* the comments contained "hate speech and direct threats to the physical integrity of individuals".[1043] In such a case

> the rights and interests of others and of society as a whole may entitle Contracting States to impose liability on Internet news portals, without contravening Article 10 of the Convention, if they fail to take measures to remove clearly unlawful comments without delay, even without notice from the alleged victim or from third parties.[1044]

To recall, the ECtHR argued that Delfi is a "large professionally managed Internet news portal run on a commercial basis which published news articles of its own and invited its readers to comment on them."[1045] The ECtHR mentions a few types of intermediaries to which the judgment does not apply:

> for example an Internet discussion forum or a bulletin board where users can freely set out their ideas on any topic without the discussion being channelled by any input from the forum's

---

[1036] L. Brunner, 'The Liability of an Online Intermediary for Third Party Content: The Watchdog Becomes the Monitor: Intermediary Liability after Delfi v Estonia', *Human Rights Law Review*, Vol. 16, No. 1, 2016, doi:10.1093/hrlr/ngv048, p. 169.
[1037] *Delfi AS v. Estonia* [GC], no. 64569/09, § 23-30, ECHR 2015-II, 16 June 2015.
[1038] *Delfi AS v. Estonia* [GC], no. 64569/09, § 68, ECHR 2015-II, 16 June 2015.
[1039] *Delfi AS v. Estonia* [GC], no. 64569/09, ECHR 2015-II, 16 June 2015.
[1040] *Delfi AS v. Estonia* [GC], no. 64569/09, § 138-139, ECHR 2015-II, 16 June 2015.
[1041] *Delfi AS v. Estonia* [GC], no. 64569/09, § 157, ECHR 2015-II, 16 June 2015.
[1042] *Delfi AS v. Estonia* [GC], no. 64569/09, § 158, ECHR 2015-II, 16 June 2015.
[1043] *Delfi AS v. Estonia* [GC], no. 64569/09, § 159, ECHR 2015-II, 16 June 2015.
[1044] *Delfi AS v. Estonia* [GC], no. 64569/09, § 159, ECHR 2015-II, 16 June 2015.
[1045] *Delfi AS v. Estonia* [GC], no. 64569/09, § 115, ECHR 2015-II, 16 June 2015.

> manager; or a social media platform where the platform provider does not offer any content and where the content provider may be a private person running the website or blog as a hobby.[1046]

This distinction is criticised in both the dissenting opinion and in the legal literature. Brunner, for example, points out that following the definition of the ECtHR, the only distinguishing feature between Delfi and social media platforms is that social media platforms, unlike Delfi, do not offer content of their own.[1047] As Brunner argues, "[a]s both a producer of content and a host of user-generated content, Delfi could be termed a hybrid intermediary."[1048] The dissenting opinion is even more discerning, arguing that "[i]t is hard to imagine how this 'damage control' will help. Freedom of expression cannot be a matter of a hobby."[1049]

Spano warns that in "assessing the precedential value of the Grand Chamber judgment in Delfi AS, it is important at the outset to bear in mind that the case was the first of its kind decided at Strasbourg."[1050] A second case in which the European Court of Human Rights had to render a decision about the liability of internet intermediaries was *Magyar Tartalomszolgáltatók Egyesülete and Index.hu Zrt v. Hungary* (2016).[1051] *Magyar* was different with respect to two things: 1) the nature of the speech and 2) the nature of the intermediary in question. According to the court, the "incriminated comments did not constitute clearly unlawful speech; and they certainly did not amount to hate speech or incitement to violence"[1052] and "while the second applicant is the owner of a large media outlet which must be regarded as having economic interests, the first applicant is a non-profit self-regulatory association of Internet service providers, with no known such interests."[1053]

Following *Delfi*, the Court sums up a few indicators to decide what the role of a service provider must be:

> the context of the comments, the measures applied by the applicant company in order to prevent or remove defamatory comments, the liability of the actual authors of the comments as an alternative to the intermediary's liability, and the consequences of the domestic proceedings for the applicant company[1054]

The Court repeats 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

---

[1046] *Delfi AS v. Estonia* [GC], no. 64569/09, § 116, ECHR 2015-II, 16 June 2015.

[1047] Brunner, 'The Liability of an Online Intermediary for Third Party Content: The Watchdog Becomes the Monitor: Intermediary Liability after Delfi v Estonia', *Human Rights Law Review*, 2016, p. 172.

[1048] Brunner, 'The Liability of an Online Intermediary for Third Party Content: The Watchdog Becomes the Monitor: Intermediary Liability after Delfi v Estonia', *Human Rights Law Review*, 2016, p. 168.

[1049] Joint dissenting opinion of judges Sajó and Tsotsoria of *Delfi AS v. Estonia* [GC], no. 64569/09, § 25, ECHR 2015-II, 16 June 2015.

[1050] R. Spano, 'Intermediary Liability for Online User Comments under the European Convention on Human Rights', *Human Rights Law Review*, Vol. 17, No. 4, 2017, doi:10.1093/hrlr/ngx001, p. 675.

[1051] *Magyar Tartalomszolgáltatók Egyesülete and Index.hu Zrt v. Hungary*, no. 22947/13, 2 February 2016.

[1052] *Magyar Tartalomszolgáltatók Egyesülete and Index.hu Zrt v. Hungary*, no. 22947/13, § 64, 2 February 2016.

[1053] *Magyar Tartalomszolgáltatók Egyesülete and Index.hu Zrt v. Hungary*, no. 22947/13, § 64, 2 February 2016.

[1054] *Magyar Tartalomszolgáltatók Egyesülete and Index.hu Zrt v. Hungary*, no. 22947/13, § 69, 2 February 2016.

> measures to remove clearly unlawful comments without delay, even without notice from the alleged victim or from third parties[1055]

The question is what the *Magyar* case would mean for providers. According to Spano, it can be argued that the Court distinguishes between non-profit intermediaries and large commercial intermediaries.[1056] The question, however, is what this distinction is worth when it comes to user-provided information that has clearly unlawful content such as hate speech. According to Polański, the ECtHR indeed seems to distinguish between small, non-commercial providers that deal with content that is not hate speech or incitement to violence; and larger, commercial providers that deal with such content.[1057] How would this apply to provider small non-commercial providers that would deal with hate speech? The ECtHR argued in *Magyar* 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 […][1058]

In the case of hate speech, it is not unlikely that the ECtHR would allow legislation imposing strict liability on small, non-commercial providers.

Noteworthy, in this respect, is that the Grand Chamber in Delfi emphasised how active Delfi was with respect to the content of user-provided information.[1059] In *Magyar,* the ECtHR reviewed the measures taken by the provider but did not characterise the editorial role nor framed the provider as active/passive.[1060] This distinction, however, remains pivotal to the DSA that is proposed as the successor (at least in part) of the e-Commerce Directive.

4.3 The Digital Services Act

On 15 December 2020, the EC launched the first draft of the DSA.[1061] In tandem with a second proposal, the Digital Markets Act (hereafter: DMA), the EC seeks to provide a regulatory framework to address the growing power of digital services. The proposals focus on consumer rights and how digital services impact fundamental rights such as the right to freedom of expression. Besides, the proposals also address the gatekeeping powers of providers.[1062] Since this chapter focuses on the regulatory framework for the liability of providers with respect to the content of user-provided information, I discuss how the DSA seeks to alter the EU liability regime.

The DSA is, regarding the liability of intermediaries, the successor of the e-Commerce Directive. However, as Articles 71(1) and (2) of the DSA state, only Articles 12 to 15 of the

---

[1055] *Magyar Tartalomszolgáltatók Egyesülete and Index.hu Zrt v. Hungary*, no. 22947/13, § 91, 2 February 2016.

[1056] Spano, 'Intermediary Liability for Online User Comments under the European Convention on Human Rights', *Human Rights Law Review*, 2017, p. 676.

[1057] Polański, 'Rethinking the notion of hosting in the aftermath of Delfi: Shifting from liability to responsibility?', *Computer Law & Security Review*, 2018, p. 874.

[1058] *Magyar Tartalomszolgáltatók Egyesülete and Index.hu Zrt v. Hungary*, no. 22947/13, § 91, 2 February 2016.

[1059] *Delfi AS v. Estonia* [GC], no. 64569/09, § 145-146, ECHR 2015-II, 16 June 2015.

[1060] *Magyar Tartalomszolgáltatók Egyesülete and Index.hu Zrt v. Hungary*, no. 22947/13, § 80-82, 2 February 2016.

[1061] Commission Proposal COM(2020) 825 final (*Digital Services Act*).

[1062] European Commission, 2020, 'Europe fit for the Digital Age: Commission proposes new rules for digital platforms'.

Directive are replaced by the DSA.[1063] Articles 12 to 15 are with a few adjustments included in the proposal of the DSA as Articles 3 to 7.[1064]

4.3.1 The applicability and scope of the DSA

Chapter I of the DSA sets out the definitions used in the regulation. New in this respect is that the DSA, unlike the e-Commerce Directive, sets out definitions, for example, "intermediary", "online platform", and "illegal content".

The objectives of the DSA are set out in Article 1(2) of the DSA. According to the DSA, its aims are to

> (a) contribute to the proper functioning of the internal market for intermediary services;
>
> (b) set out uniform rules for a safe, predictable and trusted online environment, where fundamental rights enshrined in the Charter are effectively protected.[1065]

Therefore, the DSA sets out, according to Article 1(1) of the DSA:

> (a) a framework for the conditional exemption from liability of providers of intermediary services;
>
> (b) rules on specific due diligence obligations tailored to certain specific categories of providers of intermediary services;
>
> (c) rules on the implementation and enforcement of this Regulation, including as regards the cooperation of and coordination between the competent authorities.[1066]

The DSA only applies to intermediary services.[1067] Like the e-Commerce Directive, the DSA thus starts by distinguishing between intermediary services and providers that do not offer such a service.

Furthermore, the DSA, building on the e-Commerce Directive, distinguishes between three types of intermediary services. These services are the same as defined under the e-Commerce Directive. According to article 2(f) of the DSA

> 'intermediary service' means one of the following services:
>
> – a 'mere conduit' service that consists of the transmission in a communication network of information provided by a recipient of the service, or the provision of access to a communication network;
>
> – a 'caching' service that consists of the transmission in a communication network of information provided by a recipient of the service, involving the automatic, intermediate and temporary storage of that information, for the sole purpose of making more efficient the information's onward transmission to other recipients upon their request;
>
> – a 'hosting' service that consists of the storage of information provided by, and at the request of, a recipient of the service;[1068]

[1063] Article 71(1) and (2) of Commission Proposal COM(2020) 825 final (*Digital Services Act*), p. 85.
[1064] Commission Proposal COM(2020) 825 final (*Digital Services Act*), pp. 46-47.
[1065] Article 1(2) of Commission Proposal COM(2020) 825 final (*Digital Services Act*), p. 43.
[1066] Article 1(1) of Commission Proposal COM(2020) 825 final (*Digital Services Act*), p. 43.
[1067] Article 1(4) of Commission Proposal COM(2020) 825 final (*Digital Services Act*), p. 43.
[1068] Article 2(f) of Commission Proposal COM(2020) 825 final (*Digital Services Act*), pp. 44-45.

When the DSA refers to an intermediary, it thus refers to the three intermediary services that can rely on the safe harbours under the e-Commerce Directive. However, merely offering an intermediary service is not enough. According to Article 1(3), the DSA applies "to intermediary services provided to recipients of the service that have their place of establishment or residence in the Union".[1069] Article 2(d) defines that "to offer service in the Union" means that there must be a "substantial connection" that is determined by the establishment of the provider in the Union. Besides, showing "a significant number of users in one or more Member States" or showing that the intermediary targets its "activities towards one or more Member States" is enough to assume this connection.[1070]

In sum, the DSA applies to intermediary services providers established in the EU or providers that have a large user base within the EU or target their activities to the EU – or one or more of its Member States.

4.3.2 New definitions in the DSA

As mentioned above, Chapter I of the DSA offers definitions of the most commonly used terminology in the DSA. The e-Commerce Directive relies heavily on the broad definition of "information society services"[1071] the DSA uses more specific terminology for the different intermediary functions while maintaining the distinction of intermediary services into mere conduit, caching, and hosting services.[1072]

The DSA defines a fourth category of intermediary services in Article 2(h) as a more specific form of hosting services. The DSA defines an online platform as "a provider of a hosting service which, at the request of a recipient of the service, stores and disseminates to the public information".[1073] In article 2(i), "dissemination to the public" is defined as "making information available, at the request of the recipient of the service who provided the information, to a potentially unlimited number of third parties". [1074] An online platform thus differs from the more generic hosting service in its public nature. An online platform offers a place for users to post information that is then made accessible to a broad and unrestricted audience. The determinedness of the audience is key to whether there is "dissemination to the public". If it is evident in advance how big a closed group is and whom it consists of, there is no dissemination to the public. Examples are closed private groups in a messaging application or an e-mail message.[1075]

Under these definitions, a wide range of activities may fall under the definition of 'online platform'. Therefore Article 2(h) adds that activities that only form a "minor and purely ancillary feature of another service" do not fall under the definition of an online platform as long as this service cannot be "for objective and technical reasons […] used without that other service".[1076] The DSA mentions the comment section of an online newspaper that publishes news under an

---

[1069] Article 1(3) of Commission Proposal COM(2020) 825 final (*Digital Services Act*), p. 43.
[1070] Article 2(e) of Commission Proposal COM(2020) 825 final (*Digital Services Act*), p. 44.
[1071] Article 2(a) of Directive 2000/31/EC (*Directive on electronic commerce*). The definition of 'Information Society service' is provided by Article 1(1)(b) of Directive (EU) 2015/1535.
[1072] Article 2(f) of Commission Proposal COM(2020) 825 final (*Digital Services Act*), pp. 44-45; Articles 12-14 of Directive 2000/31/EC (*Directive on electronic commerce*).
[1073] Article 2(h) of Commission Proposal COM(2020) 825 final (*Digital Services Act*), p. 45.
[1074] Article 2(i) of Commission Proposal COM(2020) 825 final (*Digital Services Act*), p. 45.
[1075] Recital 14 of Commission Proposal COM(2020) 825 final (*Digital Services Act*), p. 21.
[1076] Article 2(h) of Commission Proposal COM(2020) 825 final (*Digital Services Act*), p. 45.

editorial responsibility as an example of such an ancillary activity.[1077] Of course, an online news medium that is entirely dependent on news items submitted by users published without editorial control is a prime example of an online platform. In that case, the online newspaper cannot count on this exception. Article 2(h) continues that this exception cannot be used as "a means to circumvent the applicability of this Regulation."[1078] Service providers that qualify as an online platform, thus, are imposed with increased responsibility with respect to the content of user-provided information and in how they moderate such content.

So far, the DSA has defined intermediaries, which can be broken down into three categories of services: mere conduit, caching, and hosting services.[1079] Online platforms are defined as a specific class of hosting services. Online platforms are hosting services that allow users to provide information that is disseminated to a large – potentially infinite – number of users. The DSA ties the obligations under the proposal to the qualification of the intermediary. Partly this is tied to the possible control these intermediaries have over the information provided by users; partly, this is tied to the impact those services have when illegal content is provided to them. Illegal content is the second definitional innovation made by the DSA compared to the e-Commerce Directive.

Firstly, Article 2(g) of the DSA defines illegal content as "any information". Any information clarifies that the form in which illegal content is provided does not matter.[1080] The DSA even expresses that information related to illegal content also falls within the boundaries of this definition. The DSA gives examples of information that is in itself illegal content, such as "illegal hate speech" and "terrorist content", but also information that relates to illegal activities such as "child sexual abuse". However, selling "counterfeit products" or "activities involving infringement of consumer protection law" falls within this definition.[1081] Both information that is not compliant with EU law and the law of a Member State is considered illegal content under the DSA. The "precise subject matter or nature of that law" is, according to Article 2(g), not relevant. The DSA does not distinguish between, for example, criminal and civil law in this respect.[1082]

One of the questions in the preparation of the DSA was whether the DSA should contain provisions for information with content that is not illegal but harmful. During the consultation arose a general agreement that hosting service providers should not be obligated to remove harmful content since such an obligation raises freedom of expression concerns.[1083] Therefore there is no definition for "harmful content" in the DSA. However, there is no definition for content that is not illegal either, which is relevant because intermediaries are allowed to engage in content moderation for legal content next to some legal obligation to moderate illegal content.

Conclusion

Article 14 of the e-Commerce Directive enacted in 2000 offers a safe harbour for hosting service providers, which takes away the requirement for these providers to screen all user-provided information for illegal content. Article 15 even restricts states from imposing legislation that leads

[1077] Recital 13 of Commission Proposal COM(2020) 825 final (*Digital Services Act*), pp. 20-21.
[1078] Article 2(h) of Commission Proposal COM(2020) 825 final (*Digital Services Act*), p. 45.
[1079] Articles 3-5 of the Commission Proposal COM(2020) 825 final (*Digital Services Act*), pp. 46-47.
[1080] Article 2(g) of Commission Proposal COM(2020) 825 final (*Digital Services Act*), p. 45.
[1081] Recital 12 of Commission Proposal COM(2020) 825 final (*Digital Services Act*), p. 20.
[1082] Article 2(g) of Commission Proposal COM(2020) 825 final (*Digital Services Act*), p. 45.
[1083] Commission Proposal COM(2020) 825 final (*Digital Services Act*), p. 9.

to such a general obligation to monitor. The e-Commerce Directive shifts ex-ante editorial control to ex-post control – reviewing user-provided information before it appears on the service.

The liability regime laid down in Article 14 is knowledge-based. When the provider has knowledge or awareness of user-provided information with illegal content, the provider can escape liability by removing the information with the illegal content in question. In its case law, the ECJ, however, also limited the applicability to hosting service providers that are not active in such a manner that this gives rise to control over or knowledge of the illegal content. The ECJ made the applicability of the safe harbour as the conditions of liability conditional to the involvement of the hosting service providers in the content of the information.

The ECtHR contributed to this confusion with *Delfi*. Not only the involvement of the provider is decisive for assuming liability, but also the characteristics of the provider and the content of the information are decisive. Professional, for-profit online news platforms have a higher duty of care for manifestly illegal content (such as hate speech) than other providers. For these hosting service providers, a notice and takedown procedure is not enough. Instead, these hosting service providers must take down such manifestly illegal content without delay. Service providers that are not run for profit that deal with content that is not manifestly illegal can suffice with a notice and takedown procedure.

The relationship between the ECtHR and the eCommerce Directive is not necessarily clear. This lack of clarity may cause national courts to differ in whether they apply the safe harbours to a hosting service provider or whether they assume that the service provider had a higher standard of care, as was the case with Delfi.

In addition to the question of how the provider can prevent becoming liable for user-provided information with illegal content, the question arises whether the provider exposes itself to liability when it conducts voluntary monitoring for illegal content. As noted, the Directive does not have a clear exemption for hosting service providers from liability when they by mistake pass illegal content as legal. Do providers that engage in voluntary, proactive moderation expose themselves to liability? As noted, the DSA does not remedy this situation.

One of the main criticisms of the Directive is that it does not offer insight when the provider gains knowledge or awareness. Besides, the Directive does not offer any insight into what counts as expeditiously removal. Because the Directive does not elaborate on important core concepts, the liability regime laid down in the Directive potentially could incentivise overregulation and underregulation. How the different liability regimes may provide such a legal incentive is discussed in the following and last chapter.

Part 3: Content-based restrictions and internet intermediary service provider regulation

5 Regulatory regimes and incentives for under- and overregulation

Introduction

After discussing the liability regimes for providers regarding illegal or unlawful content of user-provided information in the EU and the US in Chapters 3 and 4, the fifth and last chapter discusses how different liability regimes provide a legal incentive that may cause underregulation and overregulation of user-provided information. As shown, providers are offered some legal protection for liability that may arise from the content of user-provided information. However, as discussed, these safe harbours and immunities are increasingly subjected to criticism. In discussing these regimes in the previous two chapters, it became clear that providers are increasingly viewed as the parties that should be responsible for taking down user-provided information with illegal or unlawful content.[1084] This responsibility sometimes stretches out over user-provided information that is not against the law (and thus not illegal) but is considered harmful.[1085] Providers are equally scrutinized for overregulating the content of user-provided information that is not strictly illegal or unlawful. Service providers are suspected of overregulating user-provided information to prevent liability[1086] or (while there is little proof) for wilfully censoring conservative political content.[1087] These suspicions aside, overregulation and underregulation are not without consequence for the users of these services.

Both overregulation and underregulation can be tied to the liability regimes in which service providers function. The previous two chapters set out the US and EU liability regimes. These liability regimes have their similarities. The US and the EU regime offer both protection against liability for illegal and unlawful content of user-provided information. Next to this, the US have for violations of intellectual property rights a similar regime as the generic liability regime for providers in the EU.[1088] These liability regimes, however, also have their differences. The US liability regime, at its core, has (with the exceptions mentioned) an absolute character, unlike the EU regime, which is of a conditional nature by providing immunity for liability to service providers as long as they fulfil a set of conditions.[1089]

Because of these differences, the proposed interventions on these liability regimes are different. As noted, the US proposals to increase the responsibility of providers for the content of user-provided information take the form of carving out the immunities provided to providers. An

---

[1084] For example in the US, violations of sex trafficking law, see FOSTA-SESTA, H.R. 1865, 115th Cong. (2018 through PL 115-164). In the EU, examples are hate speech and terrorist content, see European Commission, 2016, 'Code of Conduct on Countering Illegal Hate Speech Online'; Regulation (EU) 2021/784.

[1085] For example, this is the case with disinformation. When Biden said that providers offering platform functions were "killing people" by not doing enough against COVID-19 misinformation Biden directed the responsibility to Facebook, see Judd, Vazquez & O'Sullivan, 2021, 'Biden says platforms like Facebook are 'killing people' with Covid misinformation'. See also, European Commission, 2021, 'Code of Practice on Disinformation'.

[1086] However, there is some empirical evidence of such over-removal, see Keller, 2020, 'Empirical Evidence of "Over-Removal" by Internet Companies Under Intermediary Liability Laws'.

[1087] Hochman, 2021, 'Conservatives Should Support Section 230 Reform'; M.H. McGill & D. Lippman, 'White House Drafting Executive Order to Tackle Silicon Valley's Alleged Anti-Conservative Bias', *Politico,* 7 August 2019, available at politico.com/story/2019/08/07/white-house-tech-censorship-1639051 (retrieved on 15 February 2022).

[1088] 17 USCA § 512 (West 2010, Westlaw Next through PL 116-179).

[1089] Compare 47 USCA § 230(c) (West 2018, Westlaw Next through PL 116-91); Articles 12 to 14 of Directive 2000/31/EC (*Directive on electronic commerce*).

example of such a carve-out is exposing providers to claims based on sex trafficking law[1090] by the FOSTA-SESTA bill adopted in 2018.[1091] In the EU, such carve-outs are less popular. Content-based regulation of user-provided information through providers takes the form of voluntary codes[1092] and legal obligations that, in most cases,[1093] do not affect the safe harbours that protect providers from liability. An example forms the *Regulation on addressing the dissemination of terrorist content online,* which does not affect these safe harbours.[1094] Next to imposing liability for new categories of content, the EU regulatory regime takes the form of imposing duties of care to providers.[1095]

These approaches cause diverging consequences for how providers effectively regulate the different content categories. In this fifth and final chapter, the central question is to what extent the different regulatory regimes in the US and the European context provide a (legal) incentive for under- or overregulation of user-provided information based on its content.

The first paragraph of this chapter discusses how these liability regimes may cause providers to choose moderation practices or instruments that cause overregulation or underregulation. As set out in Chapters 1 and 2, overregulation or underregulation could be caused by numerous factors. Based on these chapters, I discuss overregulation and underregulation caused by ambiguity because of the legal terminology used to address illegal (or, in some cases, harmful) content.[1096] Besides, I discuss overregulation and underregulation by what (technological) measures providers are required to take. After that, I discuss over- and underregulation caused by limiting the remedies that providers can deploy in addressing rule violations by users.[1097]

The second paragraph of this chapter classifies the liability regimes discussed in Chapters 3 and 4 in the different theoretical regimes discussed in Paragraph 2.1.1. Occurrences of strict liability, conditional liability, and immunity regimes[1098] in the European context and the US are successively discussed. The expected under- or overregulation of the content of user-provided information tied to these regulatory regimes is discussed for the various regimes. While providers become in most regimes liable for user-provided information with illegal content, this is not the case when regulation is based on voluntary agreements or government requests. These non-liability regimes based on so-called soft law instruments are grouped as a fourth regime.

5.1 Overregulation and underregulation: ambiguity, means, and remedies

As mentioned, overregulation and underregulation can be either intentional or unintentional. Because numerous actors are involved in enacting, implementing, and applying content

---

[1090] Internet intermediaries are exposed to criminal prosecution and civil claims, see 47 USCA § 230(e)(5)(a) and (b) (West 2018, Westlaw Next through PL 116-91).

[1091] FOSTA-SESTA, H.R. 1865, 115th Cong. (2018 through PL 115-164).

[1092] European Commission, 2016, 'Code of Conduct on Countering Illegal Hate Speech Online'; European Commission, 2021, 'Code of Practice on Disinformation'.

[1093] An exception is Article 17(3) of Directive (EU) 2019/790. The obligations laid down in this article form a *lex specialis* which have precedent over Article 14 of Directive 2000/31/EC (*Directive on electronic commerce*).

[1094] See Recital 7 of Regulation (EU) 2021/784.

[1095] See Chapter III, Section 4 and 5 and Chapter IV of Commission Proposal COM(2020) 825 final (*Digital Services Act*).

[1096] See Citron, 'Extremist Speech, Compelled Conformity, and Censorship Creep', *Notre Dame Law Review*, 2018, pp. 1052-1055.

[1097] See Goldman, 'Content Moderation Remedies', *Michigan Technology Law Review*, 2021.

[1098] As distinguished by Gillespie, 2018, *Custodians of the Internet*, p. 33.

regulations, multiple actors can make decisions that lead to overregulation or underregulation. How I picture such a process: the state creates legislation, the legal team of the provider interprets this legislation, and the policy team converts this interpretation into guidelines for moderators, who apply this handbook to real-world cases. In all of these steps, the actor in question may interpret the output of the previous step more broadly than intended. Therefore, even the most careful drafted legislation may result in a wildly overbroad application by the moderation teams. This paragraph discusses the explanations of overregulation and underregulation based on the legislation discussed in Chapters 3 and 4.

5.1.1 Ambiguity and legal categories

Not all content that is prohibited is easy to spot. In other words, it may be hard for moderators of providers to recognise all prohibited content. Partly this is because both state as provider regulation has to be 'translated' for and applied to online content.[1099] Especially criminal law statutes limiting freedom of expression rights could be drafted decades ago and mainly fitting for criminal conduct by speech or the printing press. For example, Dutch criminal statutes concerning sedation or hate speech (group libel) require the criminalised expression to be done in "public".[1100] While it is clear that a seditious expression done on Twitter may count as public, it is less clear whether sedition and group libel in closed WhatsApp groups with a limited number of participants are done "in public". Besides, what counts as "seditious" or "insulting" must be derived from a substantial history of court rulings and parliamentary debates. An expression may fit a legal definition but, for example, may not be punishable because a restriction in such a given case is not "necessary in a democratic society" and thus incompatible with freedom of expression rights laid down in Article 10 of the ECHR.[1101] As noted in Paragraph 2.3.2, in moderating user-provided information, unclear definitions raise freedom of expression concerns. Ambiguity in the legislation or the community guidelines of service providers introduces the risk of overregulation of content of user-provided information that is not illegal.

Besides interpretation problems, many categories of regulated content require in-depth contextual review to establish whether the content falls under a regulated category. For example, child sexual abuse material is always illegal – independent of its context. Either the content qualifies as child sexual abuse material, or it does not. "Citing" an image containing child sexual abuse material does not derogate from its illegality. In contrast, functionally citing copyright-protected work is normally not considered an infringement. While hate speech may lose its illegal character when used in an academic, journalistic, or artistic context, such an exception does not exist for sexual child abuse material. Child sexual abuse material, thus, is binary of nature: the context does not matter when it concerns child sexual abuse material; its illegality is a given. For many, many other content categories, this is not the case. What is shared in one context may not be illegal in another and vice versa.[1102] Of course, the interpretation of ambiguous legislation is mainly of concern to conditional liability regimes, of which the EU forms an example.

[1099] Citron, 'Extremist Speech, Compelled Conformity, and Censorship Creep', *Notre Dame Law Review*, 2018, p. 1052.

[1100] See, for example, article 131 and 137c of Wetboek van Strafrecht (Dutch Criminal Code).

[1101] HR, 14 February 2017, ECLI:NL:HR:2017:220 (concl. P.C. Vegter), *Nederlandse Jurisprudentie* 2017/259, m.nt. E.J. Dommering.

[1102] douek, 2020, 'The Rise of Content Cartels', pp. 27-28; Citron, 'Extremist Speech, Compelled Conformity, and Censorship Creep', *Notre Dame Law Review*, 2018, pp. 1054-1055; Gillespie, 2018, *Custodians of the Internet*, p. 101;

Not all categories of content considered illegal or unlawful in the EU are extensively defined in EU legislation. Hate speech and (since 2021) terrorist content is defined in (binding) legislative instruments, while disinformation is not. While hate speech is not a legal category in itself, the *Council Framework Decision 2008/913/JHA* forms the basis for hate speech regulation. Hate speech is there (narrowly) defined as:

> publicly inciting to violence or hatred directed against a group of persons or a member of such a group defined by reference to race, colour, religion, descent or national or ethnic origin[1103]

The EU *Code of Conduct on Countering Illegal Hate Speech Online* does not introduce a new definition for hate speech but relies on the definition provided in the Framework Decision.[1104] The *Code of Conduct* mainly sees the service provider's measures to regulate hate speech content instead of new (material) definitions. Thus, providers are required to evaluate whether the content of user-provided information qualifies as "publicly inciting" and "violence or hatred", for which the context is crucial. Block listing forbidden words are not enough and pose a significant risk of overregulation. What is considered hatred in one context is not hateful in another. Therefore, a contextual assessment of whether the content amount to hate speech is an absolute requirement.

The regulation of terrorist content raises similar concerns. Similar to hate speech, terrorist content is also defined in EU legislation. The EU *Regulation on addressing the dissemination of terrorist content online* includes a (lengthy) definition of terrorist content. Terrorist content includes (for example) advocating (including glorification, "the commission of terrorist offences, thereby causing a danger that one or more such offences may be committed". Nex to this, solicitation of "a person or a group of persons to commit or contribute to the commission" of terrorist offences or "to participate in the activities of a terrorist group" is included. Lastly, providing instructions to commit terrorist offences (for example, bomb-making instructions) and content that "constitutes a threat to commit" terrorist offences are counted as terrorist content.[1105]

Especially the ambiguity that advocating and glorification of terrorist offences raises freedom of expression concerns. Is someone who shares terrorist content to inform the public about terrorist offences exposed to prosecution because sharing such content is considered advocating or glorification of terrorist offences? Because the draft proposal did not include an explicit exception for terrorist content used in the context in an educational, research, or journalistic context, "a group of pioneers, technologists, and innovators who have helped create and sustain today's internet" expressed their concerns to members of the EP.[1106] While journalists, researchers, nor lecturers will not necessarily create terrorist content themselves, they may include terrorist content in their reporting. In the original proposal, the EC included in the recitals that

---

Sander, 'Democratic Disruption in the Age of Social Media: Between Marketized and Structural Conceptions of Human Rights Law', *European Journal of International Law*, 2021, p. 15.

[1103] Article 1(1)(a) of Council Framework Decision 2008/913/JHA. Member States of the EU are required to transpose this in national law but may choose to "punish only conduct which is either carried out in a manner likely to disturb public order or which is threatening, abusive or insulting." see Article 1(2) of Council Framework Decision 2008/913/JHA.

[1104] European Commission, 2016, 'Code of Conduct on Countering Illegal Hate Speech Online', p. 1.

[1105] Article 2(7) of Regulation (EU) 2021/784.

[1106] M. Baker, T. Berners-Lee & V. Cerf, 'EU Terrorist Content regulation will damage the internet in Europe without meaningfully contributing to the fight against terrorism', *Politico,* 2 April 2019, available at politico.eu/wp-content/uploads/2019/04/TCO-letter-to-rapporteurs.pdf (retrieved on 14 February 2022).

> Content disseminated for educational, journalistic or research purposes should be adequately protected. Furthermore, the expression of radical, polemic or controversial views in the public debate on sensitive political questions should not be considered terrorist content.[1107]

In the Regulation itself, no clear exception was included. The final Regulation, however, included such an exception for "[m]aterial disseminated to the public for educational, journalistic, artistic or research purposes or for the purposes of preventing or countering terrorism".[1108] This category of content is thus explicitly excluded from the scope. The Regulation acknowledges that to "determine the true purpose of that dissemination and whether material is disseminated to the public for those purposes" requires a (contextual) assessment.[1109] The Regulation emphasises the role of public authorities in issuing removal orders which obligates the public authority to review the context before issuing an order. However, the Regulation also emphasises that hosting service providers have a role to "take specific measures to protect its services against the dissemination to the public of terrorist content".[1110] Hosting service providers are required to take "full account of the rights and legitimate interest of the users, in particular users' fundamental rights", including the right to freedom of expression.[1111] Especially when the hosting service provider deploys automatic means to counter terrorist content, the provider is required to have in place an "appropriate and effective safeguard […] to ensure accuracy and to avoid the removal of material that is not terrorist content."[1112]

As mentioned before, in contrast to the original proposal, the Regulation addresses the concerns expressed by critics by recognising the need for contextual assessment.[1113] Next to an exception for, for example, journalism, the Regulation has a different stance regarding automatic detection. While the proposal relied on automatic detection of terrorist content,[1114] the Regulation does not adopt such an approach because of the contextual nature of terrorist content. I discuss the use of automatic means in the next paragraph after discussing disinformation.

Disinformation is different from hate speech and terrorist content. Unlike hate speech and terrorist content, the closest thing to a legal definition of disinformation is in a non-binding code of practice and equally non-binding EC communications. The EC relies on the definition provided by the *High level Group on fake news and online disinformation*. This expert group defined disinformation as "false, inaccurate, or misleading information designed, presented and promoted to intentionally cause public harm or for profit."[1115] The EC adopted this definition in its *Code of Practice on Disinformation*.[1116] Disinformation thus does not include "misleading advertising, reporting errors,

---

[1107] Recital 9 of Commission Proposal COM(2018) 640 final of 12 September 2018 Regulation of the European Parliament and of the Council on preventing the dissemination of terrorist content online, p. 15.

[1108] Article 1(3) of Regulation (EU) 2021/784.

[1109] Article 1(3) of Regulation (EU) 2021/784.

[1110] Article 5(2) of Regulation (EU) 2021/784.

[1111] Article 5(3)(c) of Regulation (EU) 2021/784.

[1112] Article 5(3) of Regulation (EU) 2021/784.

[1113] Baker, Berners-Lee & Cerf, 2019, 'EU Terrorist Content regulation will damage the internet in Europe without meaningfully contributing to the fight against terrorism'.

[1114] Article 6(2)Commission Proposal COM(2018) 640 final, p. 26.

[1115] European Commission, 2021, 'Code of Practice on Disinformation'.
High level Group on fake news and online disinformation, *A multi-dimensional approach to disinformation: Report of the independent High level Group on fake news and online disinformation*, Luxembourg, Publications Office of the European Union, 2018, doi:10.2759/739290, p. 11.

[1116] European Commission, 2021, 'Code of Practice on Disinformation'.

satire and parody, or clearly identified partisan news and commentary". Next to this, disinformation is a residual category that does not see to illegal content.[1117] Service providers can only moderate information as disinformation when they carefully assess its content, the context in which it is shared, and even the user's intention to share this content. Not only is the provider following this definition required to review whether the content is factually untrue or at least misleading, but also whether the content is designed "to intentionally cause public harm" or is "for profit."[1118]

Because of these ambiguous definitions, providers are incentivised to adopt more explicit but often stricter rules regarding regulated content categories themselves. Service providers, for example, may regulate all misleading and false content for specific categories of disinformation.[1119] Service providers, however, do not offer much insight into applying these self-formulated definitions, while relying on their terms and conditions makes it harder to review the freedom of expression implications critically. As Keller argues it in the context of the *Code of Conduct*

> Since platforms are in theory enforcing only their own rules when they take down hate speech, users have no clear means to dispute legal interpretations or raise defenses based on European free-expression guarantees.[1120]

Even when providers copy-paste the legal definition of hate speech in their terms of service, there is, as Keller notes, little legal protection for users because providers are considered to enforce their terms and conditions and not (EU) legislation.[1121]

Ambiguity, in sum, knows three risks. The first risk is that providers apply the definitions overbroadly. The second risk is that providers are unable to review the context of the content. The third risk is that providers adopt stricter rules that are clearer and easier enforceable but against which the user has little or no legal instruments.

5.1.2 Passive/active measures and regulatory regimes

The previous paragraph discussed the first stage of moderation: interpretation and application of ambiguous concepts laid down in legislation. Service providers may implement legislation in multiple ways to prevent or remedy user-provided information with illegal content appearing on the service. Besides, these measures may be implemented for content that is not considered illegal but harmful by the provider itself. As shown in Chapters 3 and 4, legislation within the different regulatory regimes may require providers to 1) enact or alter terms and conditions prohibiting illegal content, 2) prevent illegal content from (re)appearing on their services, and 3) takedown illegal content when the provider fulfils a set of conditions. I will discuss these measures as 1) passive, 2) proactive, and 3) reactive.[1122]

[1117] European Commission, 2021, 'Code of Practice on Disinformation'.

[1118] High Level Group on Fake News and Online Disinformation, 2018, *A multi-dimensional approach to disinformation*, p. 11.

[1119] For example, LinekdIn, prohibits information with content that "directly contradicts guidance from leading global health organizations and public health authorities." See Linkedin, 'Professional community policies', *Linkedin*, available at linkedin.com/legal/professional-community-policies (retrieved on 15 February 2022).

[1120] Keller, 2019, 'Who Do You Sue? State and Platform Hybrid Power over Online Speech', p. 6.

[1121] Keller, 2019, 'Who Do You Sue? State and Platform Hybrid Power over Online Speech', p. 6.

[1122] This distinction is in part derived from E.J. Llansó, 'No amount of "AI" in content moderation will solve filtering's prior-restraint problem', *Big Data & Society*, Vol. 7, No. 1, 2020, doi:10.1177/2053951720920686, p. 1.

Passive measures

I understand passive measures as all are general measures taken for all users and all information without assessing the content of user-provided information. I distinguish passive measures from more active measures in terms of whether they are likely to cause the provider to gain knowledge of its content.[1123] In short, passive measures are all measures that are user, content, and context-independent. Because these measures do not directly counter illegal or (perceived) harmful content, they have a low direct impact on users' freedom of expression rights (they are passive towards the actual content of user-provided information) but are also less effective in countering illegal content such as hate speech. Passive measures aim to contribute to norm awareness by informing users about the applicable rules on the service.[1124] Besides, passive measures may aim to offer counter-speech or authoritative information about, for example, voting rights or COVID-19 measures. Such passive measures aim to relieve the (possible) effects of disinformation by directing users towards trustworthy information. Passive measures are, thus (from the viewpoint of what the user provided to the service) content, context, and user-independent and are meant to foster norm-conform behaviour and nudge users to authoritative sources.

Of course, informing users about the applicable norms or authoritative information does not remedy information with content that violates the rules. However, raising norm awareness is a necessity for active measures. As noted, the norms applicable to users of internet services must be sufficiently foreseeable and predictable. In other words, informing users is foundational for active moderation efforts. The norms in community guidelines may form a proxy for legislation enacted in some jurisdictions.[1125]

As noted, jurisdictions vary over what content categories are regulated. When jurisdictions regulate similar content categories, the legal definitions, however, may diverge. While there is an incentive for providers to globally regulate content categories because these categories are regulated in one or more jurisdictions, this does not necessarily mean that these categories of content are regulated in *all* jurisdictions.[1126] For example, the First Amendment forms a barrier to federal and state regulation regarding hate speech, terrorist propaganda, and disinformation in the US.[1127] Service providers, however, are not prohibited from including these categories of content in their terms and conditions themselves.[1128]

Because hosting service providers are stimulated[1129] or obligated[1130] in the EU to adopt hate speech and terrorist content regulation in their terms and conditions, EU content regulation may

---

[1123] Similar to the EU approach as discussed in Chapter 4.

[1124] While Goldman does not refer to passive measures, informing users is included in the content moderation remedies menu, see Goldman, 'Content Moderation Remedies', *Michigan Technology Law Review*, 2021, p. 37.

[1125] Keller, 2019, 'Who Do You Sue? State and Platform Hybrid Power over Online Speech', p. 6.

[1126] Citron, 'Extremist Speech, Compelled Conformity, and Censorship Creep', *Notre Dame Law Review*, 2018, pp. 1055-1056.

[1127] See, for disinformation: Sunstein, 'Falsehoods and the First Amendment', *Harvard Journal of Law & Technology*, 2020. See, for terrorist content: A. Tsesis, 'Terrorist Speech on Social Media', *Vanderbilt Law Review*, Vol. 70, No. 2, 2017 (available at scholarship.law.vanderbilt.edu/vlr/vol70/iss2/4), pp. 662-675. See for F. Schauer, 'The Exceptional First Amendment', *in* M. Ignatieff (Ed.) *American Exceptionalism and Human Rights*, Princeton, Princeton University Press 2005, doi:10.1515/9781400826889.29, pp. 32-38.

[1128] As noted in Chapter 3, the First Amendment offers editorial discretion to providers while Section 230 offers some immunity for liability.

[1129] European Commission, 2016, 'Code of Conduct on Countering Illegal Hate Speech Online', p. 2.

[1130] Article 5(1) of Regulation (EU) 2021/784.

be given global effect. Service providers enforce their terms and conditions globally for multiple reasons, including offering clarity and reducing the cost of enforcement and branding reasons.[1131] Content regulation originating from the EU thus may affect users in the US.[1132] Of course, the other way around may be equally true.

Reactive and proactive measures

Reactive and proactive measures are grouped together because of their active nature. Active measures include all these measures used to uncover prohibited content in user-provided information. Llansó distinguishes between "reactive" and "proactive" measures. Reactive measures depend on others pointing out prohibited content, while proactive measures include all measures carried out by providers to actively discover illegal or prohibited content.[1133] Gillespie, in this respect, distinguishes between editorial review, community flagging, and automatic detection.[1134] Editorial review by the provider itself has the goal of moderating information with prohibited content before it appears on the service.[1135] Gillespie:

> The dream of editorial review, if not always the reality, is perfect moderation. If it is done correctly, nothing that reaches the public violates the rules. Audiences would not be offended, regulations would be honored, corporate brands would be untarnished.[1136]

Editorial review, however, can easily lead to overregulation. As shown, ambiguous legislation can be overinterpreted and overapplied. Besides, the provider may also use this editorial review to impose judgements of their own.[1137] However, for very large online platforms, it is hard to see how such an editorial review process would function. Due to a large number of posts on these platforms, reviewing all user-provided information seems to be near impossible.[1138] According to Gillespie, very large online platforms, therefore, deploy reactive measures based on community flagging. Human moderators do not proactively review all content that is submitted to the service, but only what users of the platform point out as potentially violating.[1139]

Reactive measures, for the purpose of this chapter, can be grouped into three types of measures. Kuczerawy distinguishes between notice and takedown measures, notice and notice measures, and notice and stay down measures[1140] which will be described briefly. A notice and takedown measure means that the provider must take down prohibited content after receiving a notification. The third-party hurt by the illegal content (the notifier) notifies the provider of its existence. The provider must decide whether the user-provided information contains illegal

---

1131 Bradford, 2020, *The Brussels Effect*, p. 166.

1132 Citron, 'Extremist Speech, Compelled Conformity, and Censorship Creep', *Notre Dame Law Review*, 2018, pp. 1055-1057.

1133 Llansó, 'No amount of "AI" in content moderation will solve filtering's prior-restraint problem', *Big Data & Society*, 2020, p. 1.

1134 Gillespie, 2018, *Custodians of the Internet*, p. 77.

1135 Gillespie, 2018, *Custodians of the Internet*, p. 78.

1136 Gillespie, 2018, *Custodians of the Internet*, p. 79.

1137 Gillespie, 2018, *Custodians of the Internet*, p. 82.

1138 Gillespie, 2018, *Custodians of the Internet*, p. 86.

1139 Gillespie, 2018, *Custodians of the Internet*, pp. 86-87.

1140 Kuczerawy, 2020, 'From 'Notice and Takedown' to 'Notice and Stay Down': Risks and Safeguards for Freedom of Expression', p. 526.

content.[1141] A notice and notice measure seeks to (actively) educate a user before action is undertaken actively. A notice and notice measure means multiple warnings are issued before the provider imposes a remedy.[1142] Notice and stay down measures mean that the provider acts based on a notice (similar to the notice and takedown) but now has to ensure that information with violating content stays down. The provider ensures that information with this violating content is not re-uploaded.[1143] The "stay down"-element of this last measure also has a proactive component because the provider proactively prevents reuploading information with violating content. However, the starting point of notice and stay down measures are violating content encountered on the service by someone other than the service provider. Because of this starting point, this measure is considered reactive.

Reactive measures based on flags or notices do not mean that providers do not have to review enormous amounts of information.[1144] Tied to the problem of flagging or notifications is that it is not always clear how such a report must be understood. Are they merely individual complaints or more than that? Besides, such mechanisms can be and are abused by reporting content that is legal but deemed undesirable by the notifier.[1145] An example of such abuse is reporting political content that does not fit the beliefs of the notifier.

Automatic content moderation systems are proposed as an addition or even replacement for human moderators. Automatic content moderation systems, when functional, may, of course, be especially helpful in moderating the amount of information for illegal content that is reported or offered to the providers. However, not all moderation systems remove content.[1146] Llansó distinguishes automatic content moderation systems used to detect content that may violate the community guidelines and automatic filtering deployed to evaluate content. Detection leaves the decision to act over to (human) moderators. Detection thus merely means that a human moderator is notified. Evaluation goes further than merely detecting prohibited content. Evaluation means that an automatic content moderation system also evaluates the need for a remedy and what remedy is suitable.[1147] Gorwa, Binns and Katzenbach warn that automatic moderation offers little transparency. Automatic content moderation systems may even unequally impact (protected) minorities in how they can express themselves since these systems may be oversensitive to content concerning minority groups. This sensitivity also exists when minorities share information with such content themselves.[1148] Automatic detection systems are, as Gillespie notes, "just not very

---

[1141] Kuczerawy, 2020, 'From 'Notice and Takedown' to 'Notice and Stay Down': Risks and Safeguards for Freedom of Expression', p. 528.

[1142] Kuczerawy, 2020, 'From 'Notice and Takedown' to 'Notice and Stay Down': Risks and Safeguards for Freedom of Expression', p. 532.

[1143] Kuczerawy, 2020, 'From 'Notice and Takedown' to 'Notice and Stay Down': Risks and Safeguards for Freedom of Expression', p. 538.

[1144] Gillespie, 2018, *Custodians of the Internet*, p. 88.

[1145] Gillespie, 2018, *Custodians of the Internet*, pp. 91-93.

[1146] Gillespie, 2018, *Custodians of the Internet*, p. 97.

[1147] Llansó, 'No amount of "AI" in content moderation will solve filtering's prior-restraint problem', *Big Data & Society*, 2020, p. 2.

[1148] R. Gorwa, R. Binns & C. Katzenbach, 'Algorithmic content moderation: Technical and political challenges in the automation of platform governance', *Big data & society*, Vol. 7, No. 1, 2020, doi:10.1177/2053951719897945, pp. 10-11.

good yet."[1149] In other words, automatic content moderation is not usable for all prohibited content categories – especially not when the context is key.

However, not all automatic content moderation systems are equally risky. Gorwa, Binns and Katzenbach distinguish between automatic content moderation systems based on matching and systems that are based on prediction. Matching means that the information content is matched against a database with content known to violate the community guidelines. In contrast, automatic moderation systems based on prediction try to classify novel content that violates the community guidelines based on similarities with earlier encountered violating content while this content is different.[1150] While automatic content moderation systems are able to detect racial slurs, they are less suitable for evaluating context-depended content such as hate speech.[1151]

For some content categories, automatic moderation based on matching forms a suitable option, such as child sexual abuse material. Microsoft's PhotoDNA offers a scanning tool to check the hash (a digital fingerprint) of images against a database of hashes known to consist of child sexual abuse material. The hashes are provided by (amongst others) the National Center for Missing & Exploited Children (NCMEC).[1152] Child sexual abuse material, however, is different from, for example, terrorist content. Terrorist content, unlike child sexual abuse material, always requires a contextual assessment because the context in which it is shared is essential.[1153]

Automatic content moderation is argued to remedy the subjectivity of human moderators.[1154] Proactive automatic measures or automatic evaluation after a notice is most suitable for content that does not require a contextual assessment.[1155] As Gillespie puts it: "The most effective automatic detection techniques are the ones that know what they're looking for beforehand."[1156] For example, matching based on earlier encountered content does not work for content that is different from this content – a problem that was earlier discussed in the context of the *Glawischnig-Piesczek*-case[1157] in Paragraph 4.1.2. Because automatic content moderation tools are bad at context, these systems may lead to overregulation of user-provided information when they find a match at all.[1158]

Regardless of the remedies, the means could incentivise overregulation. Primarily proactive means (whether automatic or manual by human moderators) introduces the risk of prior constraints. Automatic means introduce the risk that content that is only under specific circumstances illegal or unlawful is always regarded as such because there is a match with earlier encountered violations (matching), or the content is similar to such content (prediction). However,

---

[1149] Gillespie, 2018, *Custodians of the Internet*, p. 98.

[1150] Gorwa, Binns & Katzenbach, 'Algorithmic content moderation: Technical and political challenges in the automation of platform governance', *Big data & society*, 2020, pp. 5-6.

[1151] Gillespie, 2018, *Custodians of the Internet*, p. 98.

[1152] Microsoft, 'PhotoDNA', *Microsoft*, available at microsoft.com/en-us/PhotoDNA (retrieved on 15 February 2022).

[1153] Baker, Berners-Lee & Cerf, 2019, 'EU Terrorist Content regulation will damage the internet in Europe without meaningfully contributing to the fight against terrorism'; Citron, 'Extremist Speech, Compelled Conformity, and Censorship Creep', *Notre Dame Law Review*, 2018, pp. 1054-1055.

[1154] Gillespie, 2018, *Custodians of the Internet*, p. 97.

[1155] Citron, 'Extremist Speech, Compelled Conformity, and Censorship Creep', *Notre Dame Law Review*, 2018, pp. 1053-1055.

[1156] Gillespie, 2018, *Custodians of the Internet*, p. 99.

[1157] Judgement of the Court (Third Chamber) in *C-18/18* (*Glawischnig-Piesczek*).

[1158] Gillespie, 2018, *Custodians of the Internet*, pp. 99-101.

also passive measures such as informing users about what norms are applicable may have overregulation as an (indirect) consequence because the communicated norms are often the basis for global enforcement. Next to this, users may perceive these standards to be much stricter than they actually are, which may also lead to a form of self-censorship.

5.1.3 Remedies and regulatory regimes

Regardless of the measure, content moderation often leads to removal as a remedy. Content moderation remedies are, as Goldman notes, guided by a "binary approach". User-provided information with rule-violating content is either left up or taken down. This approach makes removal the default option after a rule violation. This approach influenced both governmental and provider content regulation.[1159] As noted in Paragraph 2.3, such a non-differentiated approach may lead to overregulation.

After discussing the US and European approaches toward internet content regulation, it is not hard to see how these regimes influence content moderation remedies. In some liability regimes, removal is even included in the legislation as the only remedy. Hosting providers can only rely on the safe harbour offered by Article 14 of the e-Commerce Directive when they remove or disable access to illegal content when they gain knowledge or awareness.[1160] The ECtHR viewed a duty of care for a professionally managed news portal to take down manifestly illegal content "without delay", not as violating Article 10 ECHR.[1161] In other cases, a provider can be required to establish a notice and takedown mechanism that (of course) also results in the removal of the illegal content in question.[1162] How these remedies relate to liability regimes is discussed under Paragraph 5.2.

As already discussed, content-specific measures resulting in removal or inaccessibility or account level actions resulting in limitations for users to provide new information to the service are the most impactful for exercising freedom of expression rights. Therefore, as the CMSI notes, removal "needs to be done in a way which is predictable, legitimate, necessary and proportionate."[1163] Goldman, in this respect, argues that content removal as a remedy should not be the go-to solution. Other remedies may be preferable because they are better at addressing the harms caused by illegal (or otherwise harmful) content and less impactful on freedom of expression rights.[1164]

As noted, in choosing an appropriate remedy following a rule violation, providers may also deploy automatic means.[1165] The remedy is automatically chosen when automatic content moderation is used to evaluate user-provided information content.[1166] Gorwa, Binns and Katzenbach distinguish between hard and soft remedies. Hard remedies encompass remedies such as removal, while soft remedies alter the visibility of information by (for example) recommending

---

[1159] Goldman, 'Content Moderation Remedies', *Michigan Technology Law Review*, 2021, pp. 4-6 and 12-14.

[1160] Article 14 of Directive 2000/31/EC (*Directive on electronic commerce*).

[1161] *Delfi AS v. Estonia* [GC], no. 64569/09, § 162, ECHR 2015-II, 16 June 2015.

[1162] *Magyar Tartalomszolgáltatók Egyesülete and Index.hu Zrt v. Hungary*, no. 22947/13, § 91, 2 February 2016.

[1163] Council of Europe, 2021, 'Content moderation: best practices towards effective legal and procedural frameworks for self-regulatory and co-regulatory mechanisms of content moderation', p. 14.

[1164] Goldman, 'Content Moderation Remedies', *Michigan Technology Law Review*, 2021, p. 22.

[1165] Gorwa, Binns & Katzenbach, 'Algorithmic content moderation: Technical and political challenges in the automation of platform governance', *Big data & society*, 2020, p. 3.

[1166] Llansó, 'No amount of "AI" in content moderation will solve filtering's prior-restraint problem', *Big Data & Society*, 2020.

specific information to users.[1167] While the right of freedom of expression is the right to speak your mind, it does not encompass a right to reach an audience – the right to be heard on a specific forum.[1168] Soft remedies, in principle, do have a negligible impact on freedom of expression rights, while the (perceived) impact for individual users may be severe.

Removal, in contrast, imposes a hard limit on encountering information with its content on the service. In addition, remedies taken on the account level may also raise freedom of expression concerns. When account action is taken limiting the possibility of providing new information, the provider limits the user its possibility to "speak".[1169] Of course, it must be noted that soft remedies may have\ similar results as hard remedies. For example, soft remedies that limit the reach of user-provided information to only the user that provided the information have a similar effect as hard removal.

As shown, there are fears that governmental legislation backed by liability for providers leads to overregulation in the form of overremoval. Especially when legislation requires providers to impose remedies that see removal, the provider functions as a proxy between the state and the user. As discussed in the previous paragraphs, speech regulation laid down in legislation is often hard to interpret and even harder to apply. At the same time, liability regimes (normally) require swift remedies for illegal content. When legislation requires removal, this increases the risk of overinterpretation and overapplication. As shown, this could lead to overregulation of internet content.

In an ideal world, a content moderation remedy, as Goldman notes, would only be applied to user-provided information when it is certain that its content violates the rules. However, such a regulatory regime would almost certainly contribute to either very high costs for providers or severe underregulation. Goldman, therefore, proposes that non-removal remedies should be applied to user-provided information with content that is suspected (but not confirmed) to violate the rules.[1170] In a way, the chosen content moderation remedy may remedy or strengthen overregulation caused by ambiguity and the chosen means to encounter norm violations. The question may arise to whom this recommendation is directed. While in the US context, service providers have a lot of room to decide on the remedies they impose, service providers under EU legislation may be incentivised to remove questionable content. In such a case choosing non-removal remedies over removal may be too much to ask.

5.2 Regulatory regimes: incentives for under- or overregulation

As discussed in Chapter 2, the character of internet content regulation requires states to delegate carrying out this regulation to providers. As mentioned in Chapter 1, "old-school regulation" is complemented by what Balkin calls "new-school regulation".[1171] In Balkin's words, this form of regulation means that state actors attempt "to coerce or co-opt private owners of digital infrastructure to regulate the speech of private actors."[1172] In terms of cooperation, not all liability

---

[1167] Gorwa, Binns & Katzenbach, 'Algorithmic content moderation: Technical and political challenges in the automation of platform governance', *Big data & society*, 2020, pp. 3-4.

[1168] *Appleby and Others v. the United Kingdom*, no. 44306/98, § 47, ECHR 2003-VI, 6 May 2003.

[1169] However, not all account level remedies have this effect, see Goldman, 'Content Moderation Remedies', *Michigan Technology Law Review*, 2021, pp. 28-31.

[1170] Goldman, 'Content Moderation Remedies', *Michigan Technology Law Review*, 2021, pp. 41-43.

[1171] Balkin, 'Free Speech is a Triangle', *Columbia Law Review*, 2018, p. 2015.

[1172] Balkin, 'Free Speech is a Triangle', *Columbia Law Review*, 2018, p. 2016.

regimes are created equal. Following Gillespie, in this chapter, I group the different liability regimes into immunity regimes, strict liability regimes, and conditional liability regimes.[1173]

In some jurisdictions, it is hard for state actors to delegate internet content regulation to providers. In other words, not all regulatory regimes allow or facilitate such "cooperation". As noted, the US approach is characterised as almost absolute immunity for providers for (civil) liability. Besides, content-based restrictions may violate the First Amendment.[1174]

A strict liability regime in which the providers become directly liable for illegal content of user-provided information allows governments to intervene in the content of user-provided information firmly. However, strict liability regimes leave little room for nuances, especially when combined with weak (constitutional) freedom of expression rights. A conditional liability regime, in contrast, allows the state to set out the conditions on which providers may become liable for the content of user-provided information. These conditions can, as discussed in de following paragraphs, for example, be knowledge-based or negligence-based. Because the conditions are set out by the governments that seek to regulate the providers, such an approach leaves much room to tie regulation to different providers or distinct between categories of content based on their (perceived) harmfulness. Then again, such room is only granted in jurisdictions with no constitutional wall against such content-based restrictions.

Next to content regulation by making providers (direct or indirectly) liable for the content of user-provided information, there are regulatory instruments that seek to explicate the moral obligations of providers without making them legally liable. Such regulatory instruments are not meaningless despite their lack of direct legal bindingness. These non-binding voluntary regulatory instruments may influence how providers regulate user-provided information.[1175] Therefore, non-liability regimes are also included in this overview. As noted in Paragraph 5.1, factors that can be tied to overregulation and underregulation of user-provided information are ambiguity, the required means, and remedies. These three factors are for each liability regime discussed.

5.2.1 Strict liability and content regulation

Strict liability regimes are characterised by the fact that the plaintiff does not have to prove that the defendant was at fault or negligent regarding the illegal or unlawful content of user-provided information to establish liability. Providers simply become liable for the content of user-provided information because information with illegal content is published on the service. Strict liability regimes, as noted in Chapter 1, are rare. Most liability regimes offer some sort of escape from liability for providers. The clearest example of strict liability for providers can be found in China.[1176] However, there are (or were) some examples of regulatory regimes that have elements that can be characterised as strict internet intermediary liability to be found in the European context and the US as well. I first turn to these regimes and then to whether and how these regimes provide a legal incentive to overregulate or underregulate the content of user-provided information.

Strict liability regimes

As discussed, the US relied on a distinction between two roles providers can fulfil on the internet before Section 230. Does a provider qualify as a distributor or a publisher? In *Cubby v. CompuServe*

[1173] Gillespie, 2018, *Custodians of the Internet*, p. 33.

[1174] See Chapter 3.

[1175] Citron, 'Extremist Speech, Compelled Conformity, and Censorship Creep', *Notre Dame Law Review*, 2018.

[1176] Gillespie, 2018, *Custodians of the Internet*, p. 33.

(1991), distributor liability for internet providers was clarified as that the "distributor must have knowledge of the contents of a publication before liability can be imposed for distributing that publication".[1177] Because the provider exercised editorial control over user-provided information that can be compared with the editorial control of a distributor, the provider could not be held liable for the content of all user-provided information.[1178] In *Cubby,* the provider was thus not imposed with strict liability because its editorial control could not be equalised with a publisher. To be held liable as a distributor, the plaintiff had to show that CompuServe had "known or had reason to know" that this specific instance of user-provided information its content was defamatory.[1179]

Cubby can be contrasted with the earlier discussed ruling in *Stratton Oakmont v. Prodigy Services* (1995), in which the provider was regarded as "a publisher rather than a distributor."[1180] Because Prodigy exercised editorial control, which consisted of removing and filtering the content of user-provided information, Prodigy could be regarded as a publisher instead of a distributor.[1181] The provider becomes liable for all illegal content of user-provided information because the provider exercises editorial control. As discussed in Chapter 3, such a liability regime can be characterised as strict (publisher) liability. In the case of defamation, primary publishers are held liable by strict liability standards, while secondary publishers are only held liable when knowledge (at least constructive knowledge) is proven.[1182] Section 230 prevents the liability of providers for user-provided information by exempting them from (civil) liability from the content of user-provided information.

In the European context, the *Delfi* ruling by the Grand Chamber of the ECtHR in 2015 comes close to a strict liability regime, according to Keller.[1183] *Delfi* reads like the European version of *Prodigy*.[1184] As noted, Delfi did set the rules on its platform, had some automated moderation in place based on a list of forbidden words, and offered a notice and takedown procedure for comments that violated the rules but were not filtered out by the automatic moderation. Delfi exercised thus some editorial control before the publication of the comments and had a disclaimer in place that the content of user-provided comments does not reflect Delfi's viewpoints.[1185] As noted in Chapter 4, Delfi was denied conditional immunity under Article 14 of the e-Commerce Directive. The ECtHR, however, does not interpret the national law based on the Directive, nor does it answer the question of whether the Directive is interpreted or applied correctly. In other words, the ECtHR merely assessed how the given decision regarding Delfi its liability holds up to the ECHR.[1186]

[1177] *Cubby, Inc. v. CompuServe, Inc.*, 776 F.Supp. 135, 139 (S.D. New York 1991).

[1178] *Cubby, Inc. v. CompuServe, Inc.*, 776 F.Supp. 135, 140 (S.D. New York 1991).

[1179] *Cubby, Inc. v. CompuServe, Inc.*, 776 F.Supp. 135, 139-141 (S.D. New York 1991).

[1180] *Stratton Oakmont, Inc. v. Prodigy Services Co.*, 23 Media L Rep 1794 (N.Y. Sup. Ct. 1995).

[1181] Fishback, 'How the Wolf of Wall Street Shaped the Internet: A Review of Section 230 of the Communications Decency Act', *Texas Intellectual Property Law Journal*, 2020, p. 282.

[1182] *Barnes v. Yahoo!, Inc.*, 570 F.3d 1096, 1104 (9th Cir. 2009).

[1183] *Delfi AS v. Estonia* [GC], no. 64569/09, ECHR 2015-II, 16 June 2015; Note 19 of Keller, 2019, 'Who Do You Sue? State and Platform Hybrid Power over Online Speech', p. 4.

[1184] Omer, 'Intermediary Liability for Harmful Speech: Lessons from Abroad', *Harvard Journal of Law & Technology*, 2014, p. 313.

[1185] *Delfi AS v. Estonia* [GC], no. 64569/09, § 11-14, ECHR 2015-II, 16 June 2015.

[1186] *Delfi AS v. Estonia* [GC], no. 64569/09, § 123 and 127, ECHR 2015-II, 16 June 2015.

In *Delfi*, the ECtHR gave some guidance for states in imposing strict liability to providers that do not directly takedown (anonymous) user-provided information with unlawful content. The Grand Chamber argued

> that the rights and interests of others and of society as a whole may entitle Contracting States to impose liability on Internet news portals, without contravening Article 10 of the Convention, if they fail to take measures to remove clearly unlawful comments without delay, even without notice from the alleged victim or from third parties.[1187]

According to Keller, the ECtHR adopted in *Delfi* a form of strict liability.[1188] The strict nature of this liability regime is a given because providers that "fail to take measures to remove clearly unlawful comments without delay"[1189] become directly liable without the requirement to fulfil additional conditions. As noted, the ECtHR clarified a year later in *Magyar* that this strict approach could be differentiated between commercial and non-commercial providers. Besides, the ECtHR differentiated between the nature of the illegal content of the comments. While the domestic courts saw the mere fact that the provider disseminated user-provided information with unlawful content as sufficient for liability,[1190] the ECtHR observed that this was done "without embarking on a proportionality analysis of the liability of the actual authors of the comments and that of the [service provider, MK]."[1191]

As discussed in Paragraph 4.2, the facts of *Delfi* and *Magyar* are different. Delfi was held liable for user-provided information that "were of a clearly unlawful nature."[1192] In *Magyar,* the "comments did not constitute clearly unlawful speech; and they certainly did not amount to hate speech or incitement to violence."[1193] Besides, in *Delfi*, the provider was "professionally managed" and "run on a commercial basis". According to the ECtHR, Delfi "sought to attract a large number of comments on news articles published by it."[1194] In *Magyar*, the provider that was held liable had no such commercial interests and was run not for profit.[1195]

In *Magyar,* a notice and action mechanism (for example, removal upon receiving a notice) was seen as sufficient – at least for not clearly unlawful comments that are not considered hate speech.[1196] Strict liability for the content of user-provided comments would violate the freedom of expression rights of the service provider.[1197] The ECtHR, therefore, suggests that a strict liability regime may be suitable for user-provided information with clearly unlawful content such as "hate speech and direct threats to the physical integrity of individuals".[1198] When this is the case, the provider could be held liable according to the ECtHR

[1187] *Delfi AS v. Estonia* [GC], no. 64569/09, § 159, ECHR 2015-II, 16 June 2015.
[1188] Note 19 of Keller, 2019, 'Who Do You Sue? State and Platform Hybrid Power over Online Speech', p. 4.
[1189] *Delfi AS v. Estonia* [GC], no. 64569/09, § 159, ECHR 2015-II, 16 June 2015.
[1190] *Magyar Tartalomszolgáltatók Egyesülete and Index.hu Zrt v. Hungary*, no. 22947/13, § 42, 78-79, 2 February 2016.
[1191] *Magyar Tartalomszolgáltatók Egyesülete and Index.hu Zrt v. Hungary*, no. 22947/13, § 79, 2 February 2016.
[1192] *Delfi AS v. Estonia* [GC], no. 64569/09, § 140, ECHR 2015-II, 16 June 2015.
[1193] *Magyar Tartalomszolgáltatók Egyesülete and Index.hu Zrt v. Hungary*, no. 22947/13, § 64, 2 February 2016. See also, *Magyar Tartalomszolgáltatók Egyesülete and Index.hu Zrt v. Hungary*, no. 22947/13, § 63, 2 February 2016.
[1194] *Delfi AS v. Estonia* [GC], no. 64569/09, § 144, ECHR 2015-II, 16 June 2015.
[1195] *Magyar Tartalomszolgáltatók Egyesülete and Index.hu Zrt v. Hungary*, no. 22947/13, § 64, 2 February 2016.
[1196] *Magyar Tartalomszolgáltatók Egyesülete and Index.hu Zrt v. Hungary*, no. 22947/13, § 91, 2 February 2016.
[1197] Note 19 of Keller, 2019, 'Who Do You Sue? State and Platform Hybrid Power over Online Speech', p. 4.
[1198] *Magyar Tartalomszolgáltatók Egyesülete and Index.hu Zrt v. Hungary*, no. 22947/13, § 91, 2 February 2016.

> if they fail to take measures to remove clearly unlawful comments without delay, even without notice from the alleged victim or from third parties.[1199]

While *Delfi* has elements that can be characterised as an example of strict liability for users' comments with illegal (hate speech) content, *Magyar* forms an example in which strict liability for users-provided information with defamatory content is struck down.[1200]

The incentive that comes from strict liability regimes

As discussed in Chapter 3, in the US, the strict liability approach for providers that moderate the content of user-provided information had severe adverse effects. Providers that moderated the content of user-provided information were exposed to strict (publisher) liability, while providers that did not moderate could only be held liable as distributors. Distributors can only be held liable when the plaintiffs prove (constructive) knowledge of the illegal content. A first incentive that comes from strict liability is thus that the providers have an incentive to alter the services or conduct to fall within other legal categories that are not imposed with strict liability. Service providers that fear being held liable as publishers may alter how they moderate their services in order to qualify as a distributor. This way, the liability regime, in fact, stimulates the provider to underregulate.

Carving out the immunities provided by Section 230 would expose a provider to (strict) liability for user-provided information with illegal content.[1201] Such carve-outs expose providers to strict liability for the categories of content that are exempted from the immunity regime. Since the immunisation offered by Section 230 was to prevent overregulation, such a carve-out may precisely have the reverse effect. In other words, providers would be confronted with strict (publisher) liability for the exempted (unimmunised) categories of content when they carry out some moderation on their platforms.

For example, Section 230 carved out for user-provided information with content that violates sex trafficking law was followed with a policy change by providers. Craigslist,[1202] Reddit,[1203] and Google[1204] amended their policies or changed their enforcement practices regarding adult content. While Craigslist is most outspoken about the fact that this legislative change led to adopting its policy out of fear of liability,[1205] the policies of Reddit (prohibiting '[p]aid services involving physical sexual contact')[1206] and Google (suddenly removing videos containing adult content in accordance with their terms of service)[1207] may also have been influenced by the carve-

---

[1199] *Delfi AS v. Estonia* [GC], no. 64569/09, § 159, ECHR 2015-II, 16 June 2015.

[1200] See note 19 of Keller, 2019, 'Who Do You Sue? State and Platform Hybrid Power over Online Speech', p. 14.

[1201] See *Stratton Oakmont, Inc. v. Prodigy Services Co.*, 23 Media L Rep 1794 (N.Y. Sup. Ct. 1995).

[1202] Craigslist, 'FOSTA', *Craigslist*, available at craigslist.org/about/FOSTA (retrieved on 14 February 2022).

[1203] u/Reddit-Policy, 'New addition to site-wide rules regarding the use of Reddit to conduct transactions', *reddit.com/r/announcements,* 21 March 2018, available at reddit.com/r/announcements/comments/863xcj/new_addition_to_sitewide_rules_regarding_the_use (retrieved on 15 February 2022).

[1204] S. Cole, 'Sex Workers Say Porn on Google Drive Is Suddenly Disappearing', *Vice,* 21 March 2018, available at vice.com/en/article/9kgwnp/porn-on-google-drive-error (retrieved on 14 February 2022).

[1205] Craigslist, 'FOSTA'.

[1206] u/Reddit-Policy, 2018, 'New addition to site-wide rules regarding the use of Reddit to conduct transactions'.

[1207] Cole, 2018, 'Sex Workers Say Porn on Google Drive Is Suddenly Disappearing'.

out.[1208] FOSTA-SESTA may have led to a chilling effect with respect to adult content. Large providers seemed to be incentivised to remove content that they doubt is legal. At the same time, the less resourceful intermediaries closed functionalities for which the costs of moderation or the risk of liability would be too high.[1209]

Does *Delfi* has the same consequence as FOSTA-SESTA? While *Delfi* was met with heavy criticism,[1210] the (theoretical) consequence was less severe than *Prodigy*. However, that does not mean that the standard laid down in *Delfi* is not troubling. Balkin even mentions *Delfi* in the context of collateral censorship.[1211] However, the consequences of the strict liability elements of *Delfi* are limited because national courts are still able to rely on their constitutional safeguards.[1212] Besides, it is not unthinkable that in numerous cases, the courts apply and rely on the safe harbours provided by the e-Commerce Directive. The ECtHR does not require states to impose liability on providers as a positive obligation. The liability of *Delfi* in the given case was merely viewed as compatible with Article 10 of the Convention. Article 10 of the Convention even contains arguments against such strict liability elements. The CDMSI argues that states may have a positive obligation under the Convention to ensure that providers do not have an incentive to overregulate user-provided information content.[1213] A general obligation for providers to monitor for illegal content could violate Article 10 of the Convention.[1214] Not everyone views the *Delfi* ruling as a risk. Fishback, for example, is more favourable towards *Delfi*: "Delfi knew that there were going to be defamatory comments and failed to act. Not only do we want to promote moderation, but we want to promote proper moderation."[1215]

The question is whether the strict liability elements laid down in *Delfi* stimulate such moderation. Strict liability regimes should be considered notorious for a strong incentive to overregulate. Service providers would most likely choose measures that circumvent liability. Liability for a specific category of content would provide a powerful incentive to prohibit this category of content. Liability for anonymous comments would incentivise providers to restrict the anonymous publishing of comments which would require a careful balancing in the light of the

---

[1208] A. Romano, 'A new law intended to curb sex trafficking threatens the future of the internet as we know it', *Vox,* 2 July 2018, available at vox.com/culture/2018/4/13/17172762/fosta-sesta-backpage-230-internet-freedom (retrieved on 15 February 2022).

[1209] Goldman, 'The Complicated Story of FOSTA and Section 230', *First Amendment Law Review*, 2019, pp. 288-289.

[1210] Kosseff, 2019, *The Twenty-Six Words That Created the Internet*, p. 153; M. Masnick, 'Huge Loss For Free Speech In Europe: Human Rights Court Says Sites Liable For User Comments', *techdirt,* 16 June 2015, available at techdirt.com/articles/20150616/11252831361/huge-loss-free-speech-europe-human-rights-court-says-sites-liable-user-comments.shtml (retrieved on 15 February 2022); Note 282 of Yemini, 'The New Irony of Free Speech', *Columbia Science and Technology Law Review*, 2018, p. 175; J. Malcolm, 'As Threats to Korean and European Web Hosts Rise, the Manila Principles Go on Tour', *Electronic Frontier Foundation,* 10 July 2015, available at eff.org/deeplinks/2015/07/threats-korean-and-european-web-hosts-rise-manila-principles-go-tour (retrieved on 15 February 2022).

[1211] See note 80 of Balkin, 'Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation', *U.C. Davis Law Review*, 2018, p. 1179.

[1212] D. Voorhoof, 'De aansprakelijkheid van online nieuwsplatforms na Delfi', *MediaForum*, No. 6, 2015, p. 204.

[1213] Council of Europe, 2021, 'Content moderation: best practices towards effective legal and procedural frameworks for self-regulatory and co-regulatory mechanisms of content moderation', p. 24.

[1214] Brunner, 'The Liability of an Online Intermediary for Third Party Content: The Watchdog Becomes the Monitor: Intermediary Liability after Delfi v Estonia', *Human Rights Law Review*, 2016, p. 172.

[1215] Fishback, 'How the Wolf of Wall Street Shaped the Internet: A Review of Section 230 of the Communications Decency Act', *Texas Intellectual Property Law Journal*, 2020, p. 294.

ECHR.[1216] Especially when providers are offered no safe harbour, such liability regimes may have such an effect. Conditional liability regimes may take away the sharp edges of strict liability regimes.

5.2.2 Conditional liability and content regulation

Conditional liability regimes (or conditional immunity regimes) are characterised by the fact that the provider to become liable (or retain its immunity) has to fulfil a predefined set of conditions. Providers only become liable when they fail to meet the safe harbour (conditional immunity) requirements or when they fulfil a predefined set of requirements causing them to become liable (conditional liability). Conditional liability regimes are to be found in many jurisdictions. As noted, Article 14 of the e-Commerce Directive of the EU is a clear example of such a conditional immunity regime. In the US, the DMCA is an example of conditional liability. I first turn to these regimes and then turn to how these regulatory regimes provide a legal incentive to overregulate or underregulate the content of user-provided information.

Conditional liability (and immunity) regimes

As noted in Chapter 3, the US internet intermediary liability regime knows one major exception in Section 230: violations of intellectual property law.[1217] As shown in Chapters 3 and 4, the DMCA liability regime laid down in Section 512(c) has a lot in common with Article 14 of the e-Commerce Directive. As discussed in Chapter 4, the e-Commerce Directive offers a safe harbour to hosting service providers that store user-provided information. A similar liability regime is laid down in Section 512(c)(1)(A), which applies to "information residing on systems or networks at direction of users".[1218] These safe harbours shield providers from liability for illegal (in the case of the DMCA violations of intellectual property law) content of user-provided information. This safe harbour is conditional. Only a hosting service provider that does not have knowledge or awareness of the illegal or unlawful nature can rely on the safe harbour. When the provider obtains such knowledge or awareness, the provider can escape liability when the provider acts expeditiously and removes or disable access to this information.[1219]

As noted, the liability regime laid down in the DMCA requires removal upon receiving a notification. The provider is obligated to inform the users that provided the information subjected to removal.[1220] The user may submit a counter-notice.[1221] The provider forwards the counter-notice to the notifier, announcing that access to the allegedly infringing information will be restored after ten business days.[1222] When the original notifier notifies the provider that it will request a court order against the user, the information is not restored. When the notifier does not take the case to court, the provider must restore the information after ten but no later than fourteen business days[1223] The counter-notice procedure shields the providers from liability. The provider that follows this counter-notice procedure is exempted from legal liability – when the content is taken

---

[1216] *Delfi AS v. Estonia* [GC], no. 64569/09, § 147-149, ECHR 2015-II, 16 June 2015.
[1217] 47 USCA § 230(e)(2) (West 2018, Westlaw Next through PL 116-91).
[1218] 17 USCA § 512(c)(1)(A) (West 2010, Westlaw Next through PL 116-179).
[1219] Article 14 of Directive 2000/31/EC (*Directive on electronic commerce*).
[1220] 17 USCA § 512(g)(2)(A) (West 2010, Westlaw Next through PL 116-179).
[1221] 17 USCA § 512(g)(3) (West 2010, Westlaw Next through PL 116-179).
[1222] 17 USCA § 512(g)(2)(B) (West 2010, Westlaw Next through PL 116-179).
[1223] 17 USCA § 512(g)(2)(C) (West 2010, Westlaw Next through PL 116-179).

down but also when the content remains up.[1224] Similar to Section 230, the DMCA knows a safe harbour for good faith restrictions on information which the provider believed its content was infringing.[1225]

Unlike the DMCA, the e-Commerce Directive (as discussed) does not codify when a provider gains knowledge or awareness, nor does the e-Commerce Directive offer a notice and takedown procedure. By lacking such procedures, the e-Commerce Directive also does not offer a counter-notice procedure or an exception for liability from erroneous removal of user-provided information offered to the providers. Instead, the e-Commerce Directive suggests[1226] that hosting service providers that erroneously remove user-provided information that does not contain illegal or unlawful content violate the user's expression rights.[1227]

The DSA, as proposed by the EC in December 2020, offers both a notice and takedown procedure and remedies for users arguing that the information they provided was erroneously moderated. For starters, the DSA offers a notice and action mechanism that "shall be considered to give rise to actual knowledge or awareness".[1228] As noted, the DSA adds some requirements with respect to the quality of the notification, which must be "sufficiently precise and adequately substantiated […] on the basis of which a diligent economic operator can identify the illegality of the content in question."[1229] By receiving a notice, a hosting service provider may gain knowledge or awareness of the illegal or unlawful content of user-provided information which may lead to expeditious action resulting in removal or inaccessibility of user-provided information to prevent liability.[1230] In deciding about the action, the provider must consider the freedom of expression rights of the user.[1231] At this moment, it is unclear how the courts will (have to) balance the interests of the provider against the freedom of expression interests of different users.

While the recitals of the e-Commerce Directive state that the hosting service provider is required to take into account users' freedom of expression rights while moderating under Article 14,[1232] the DSA makes this far more explicit by establishing complaint and appeal procedures. This procedure consists of three parts. First, the hosting service provider must provide a statement of reasons to the user whose information is affected by moderation decisions that lead to removal or inaccessibility "to specific items of information" provided by the user.[1233] Additional rules apply when the hosting service provider offers a service that qualifies as an "online platform".[1234] In this case, the user, informed by this statement of reasons, can file a complaint to the provider within

---

[1224] 17 USCA § 512(g)(4) (West 2010, Westlaw Next through PL 116-179).

[1225] 17 USCA § 512(g)(1) (West 2010, Westlaw Next through PL 116-179).

[1226] Recital 46 of Directive 2000/31/EC (*Directive on electronic commerce*).

[1227] Van Eecke, 'Online service providers and liability: A plea for a balanced approach', *Common Market Law Review*, 2011, p. 1468.

[1228] Article 14(3) of Commission Proposal COM(2020) 825 final (*Digital Services Act*), p. 51.

[1229] Article 14(2) of Commission Proposal COM(2020) 825 final (*Digital Services Act*), p. 51.

[1230] Article 5(1) of Commission Proposal COM(2020) 825 final (*Digital Services Act*), p. 47.

[1231] Recital 22 of Commission Proposal COM(2020) 825 final (*Digital Services Act*), p. 22. See also, Article 12 of Commission Proposal COM(2020) 825 final (*Digital Services Act*), p. 50.

[1232] Recital 46 of Directive 2000/31/EC (*Directive on electronic commerce*).

[1233] Article 15(1) of Commission Proposal COM(2020) 825 final (*Digital Services Act*), p. 52.

[1234] Section 3 of the DSA only applies to providers offering an "online platform", see Commission Proposal COM(2020) 825 final (*Digital Services Act*), p. 53. Not all providers offering online platforms functionalities qualify as an online platform for the purpose of Section 3, see Recital 13 and 43 and Article 2(h) and Article 16 of Commission Proposal COM(2020) 825 final (*Digital Services Act*), pp. 20-21, 27, 44-45 and 54.

six months after the service provider removed or disabled access to information of the user, suspended specific services or terminated the user's account.[1235] If the user is still dissatisfied after the internal procedure, the user can use an out-of-court dispute settlement procedure.[1236] The DSA does not exclude the possibility of petitioning a court when allowed under national law.[1237]

Next to knowledge-based liability regimes, the second group of conditional liability regimes are negligence-based liability regimes. There are some arguments that can be made that next to strict liability, a negligence-based liability regime can be found in *Delfi*. As noted, scholars differ in their opinion of what *Delfi* means for the liability of providers. While *Delfi* has some elements that fit under a strict liability approach, the argument can be made that the ECtHR adopted a negligence-based liability approach. Delfi could be held liable because the provider did not take the necessary steps to prevent illegal content. In accessing its duty of care, the seriousness of the violation of the rights of others (content) as the characteristics of the provider are decisive.[1238] For example, as Fishback argued in the context of *Delfi*. Sometimes illegal content can be expected, and it is strange when service providers are not required to take action in such a case.[1239] Usually, negligence-based liability is complimented with a specific standard, a duty of care, that the provider must uphold. The standard that can be derived from *Delfi* is that a provider with similar characteristics as Delfi can filter all user-provided comments for clearly illegal content. Of course, the same line of argument can conclude that the ECtHR allows a strict liability regime for illegal content for providers with similar characteristics as Delfi. Because the standard does not mainly see to the duty of care but to the characteristics of the provider, I would say that *Delfi* holds a strict liability approach for a certain type of provider.

As noted, there are also conditional liability regimes that are tied to a specific circumstance. The third group of liability regimes discussed here are "residual liability regimes", in which the provider only becomes liable for the content of user-provided information when the responsible user is legally not reachable.[1240] These liability regimes (at this point) are theoretical but may become very real soon.[1241] In the *Delfi*-ruling, a liability regime that depends on disclosing the users'

[1235] Article 17(1) of Commission Proposal COM(2020) 825 final (*Digital Services Act*), p. 53.

[1236] Article 18(1) of Commission Proposal COM(2020) 825 final (*Digital Services Act*), pp. 53-54.

[1237] Article 18(1) of Commission Proposal COM(2020) 825 final (*Digital Services Act*), pp. 53-54.

[1238] Polański, 'Rethinking the notion of hosting in the aftermath of Delfi: Shifting from liability to responsibility?', *Computer Law & Security Review*, 2018, p. 874.

[1239] Fishback, 'How the Wolf of Wall Street Shaped the Internet: A Review of Section 230 of the Communications Decency Act', *Texas Intellectual Property Law Journal*, 2020, p. 294.

[1240] Perry & Zarsky, 'Who Should Be Liable for Online Anonymous Defamation?', *University of Chicago Law Review Dialogue*, 2015, p. 172.

[1241] For example, in Australia, see Prime Minister of Australia, 'Combatting online trolls and strengthening defamation laws', *Prime Minister of Australia,* 28 November 2021, available at pm.gov.au/media/combatting-online-trolls-and-strengthening-defamation-laws (retrieved on 15 February 2022); T. Lowrey, 'Social media companies could be forced to give out names and contact details, under new anti-troll laws', *ABC News,* 28 November 2021, available at abc.net.au/news/2021-11-28/social-media-laws-online-trolls/100657004 (retrieved on 15 February 2022); E. Roth, 'Australian PM proposes defamation laws forcing social platforms to unmask trolls', *The Verge,* 28 November 2021, available at theverge.com/2021/11/28/22806369/australia-proposes-defamation-laws-unmask-trolls (retrieved on 15 February 2022). In France similar proposals are made, see M. Pollet, 'French senator calls for creation of digital identity supervisory body', *Euractiv,* 21 October 2021, available at euractiv.com/section/digital/news/french-senator-calls-for-creation-of-digital-identity-supervisory-body (retrieved on 15 February 2022).

personalia is implicitly considered by the ECtHR. Knowing the responsible user does not make it easy for the victim to remove illegal content. According to the ECtHR

> […] the uncertain effectiveness of measures allowing the identity of the authors of the comments to be established, coupled with the lack of instruments put in place by the applicant company for the same purpose with a view to making it possible for a victim of hate speech to bring a claim effectively against the authors of the comments[1242]

If the authors of individual comments were both identifiable and reachable, *Delfi* could have turned a whole other direction. The ECtHR, however, did not endorse such a residual liability regime. The ECtHR "is mindful of the interest of Internet users in not disclosing their identity" since "[a]nonymity has long been a means of avoiding reprisals or unwanted attention."[1243] As Perry and Zarsky mention, a conditional liability that is dependent on the anonymity of the user "may jeopardize the right to speak with anonymity […] and the right to privacy".[1244] Residual liability regimes requiring disclosing the users' identity would require a careful balance between the right of (anonymous) freedom of expression[1245] and privacy (reputation) rights.[1246]

The result of a conditional liability regime

As discussed, strict liability regimes force providers to either cease offering the service or conduct perfect content moderation of the content of user-provided information. In strict liability regimes, some providers may choose not to moderate at all. Under the discussed strict liability regimes, only providers that do conduct moderation are held liable for the content of user-provided information (see the publisher/distributor distinction in the US). When providers stop moderating, user-provided information with illegal or unlawful content may become rampant on the service. When providers decide to cease offering their services, this may lead to fewer possibilities for users to express themselves. As Goldman notes, perfect moderation is not feasible for many providers.[1247]

Conditional liability regimes may pose less risk for providers in terms of liability for user-provided information with illegal content than these strict liability regimes. Under a conditional liability regime, a provider is not incentivised to cease offering its services or engage in perfect moderation. Conditional liability regimes rely on a specific set of conditions that the provider must fulfil to maintain or lose its immunity. Consequently, the focus shifts to these conditions. For example, conditional liability regimes based on knowledge or awareness may incentivise providers to (over-)remove grey-area content and for 'censorship creep'[1248] in which the legislation is also applied to content that was originally not foreseen. A broader category of content than intended is also included to prevent liability.[1249] Based on the foregoing, it is likely that knowledge-based liability regimes provide an incentive to remove user-provided information with alleged illegal

---

[1242] *Delfi AS v. Estonia* [GC], no. 64569/09, § 151, ECHR 2015-II, 16 June 2015.

[1243] *Delfi AS v. Estonia* [GC], no. 64569/09, § 147, ECHR 2015-II, 16 June 2015.

[1244] Perry & Zarsky, 'Who Should Be Liable for Online Anonymous Defamation?', *University of Chicago Law Review Dialogue*, 2015, p. 174.

[1245] *Delfi AS v. Estonia* [GC], no. 64569/09, § 147-151, ECHR 2015-II, 16 June 2015.

[1246] *Delfi AS v. Estonia* [GC], no. 64569/09, § 137-139, ECHR 2015-II, 16 June 2015.

[1247] Goldman, 'The Complicated Story of FOSTA and Section 230', *First Amendment Law Review*, 2019, pp. 288-289.

[1248] Citron defines 'censorship creep' as "the expansion of speech policies beyond their original goals.", see Citron, 'Extremist Speech, Compelled Conformity, and Censorship Creep', *Notre Dame Law Review*, 2018, p. 1051.

[1249] Citron, 'Extremist Speech, Compelled Conformity, and Censorship Creep', *Notre Dame Law Review*, 2018, pp. 1049-1061.

content when a provider is notified of the existence of this content – even when the legal status of the content is not sufficiently clear.

However, knowledge-based liability regimes may cause providers to restrict their own (voluntary) efforts to moderate user-provided information. The liability regime offered by the Directive and as proposed by the DSA is based on knowledge or awareness. The provider is shielded from liability for the content of user-provided information as long as the hosting service provider does not have any knowledge or awareness of its illegal content. While moderating does not lead the provider to become liable for all illegal content of user-provided information because the provider moderates,[1250] it may expose the provider to liability for the content it encountered but failed to remove as illegal.[1251] The provider may be presumed to have gained knowledge of its illegal content. The provider may become liable because the provider did not remove the information with illegal content.

Users may thus be confronted with erroneous removal or wrongful moderation of user-provided information. Removing content that is not illegal or explicitly prohibited by the provider raises freedom of expression concerns. Erroneous removal, however, is not easy to address successfully.[1252] For example, the e-Commerce Directive, while stating that freedom of expression rights put weight on the scale, does not offer any accessible mechanisms for redress. While EU member states can establish other procedures, the only possibility is to sue the provider – with unpredictable chances of success in most EU countries.[1253] The e-Commerce Directive expresses that "the removal or disabling of access has to be undertaken in the observance of the principle of freedom of expression".[1254] The safeguards under the DSA are much more substantial. The question is, however, how these safeguards will function.

As noted, the residual indirect liability approach offered by Perry and Zarsky as an alternative for anonymous defamatory content does not pose risks for over-removal.[1255] The service provider, in this model, only becomes liable when the user that provided the information with illegal content is not amenable to litigation.[1256] In other words, when the user who provided the allegedly defamatory content takes responsibility or the provider can shift the responsibility to the user, the provider no longer bears this legal responsibility. Such a regime prevents a provider from removing allegedly defamatory content by default out of fear of liability.

[1250] Proactive measures do not render the provider liable for all content, see Communication COM(2017)555 final, pp. 10-12. In the DSA this rule is codified in Article 6 of Commission Proposal COM(2020) 825 final (*Digital Services Act*), p. 47.

[1251] Kuczerawy, 2018, 'The EU Commission on voluntary monitoring: Good Samaritan 2.0 or Good Samaritan 0.5?'.

[1252] Keller, 2019, 'Who Do You Sue? State and Platform Hybrid Power over Online Speech', p. 2.

[1253] For example, while not illegal or unlawful, disinformation can be removed by providers in the Netherlands, see Rb. Amsterdam (vzr.), 9 September 2020, ECLI:NL:RBAMS:2020:4435, *Jurisprudentie geneesmiddelenrecht* 2020/36, m.nt. M.D.B. Schutjens; Rb. Amsterdam (vzr.), 13 October 2020, ECLI:NL:RBAMS:2020:4966, *Computerrecht* 2021/66, m.nt. M. Klos (*Facebook*); Rb. Amsterdam (vzr.), 18 August 2021, ECLI:NL:RBAMS:2021:4308 (*BLCKBX/Google*). In Germany diverging case law is developing, see for example, Goujard, 2021, 'German Facebook ruling boosts EU push for stricter content moderation'.

[1254] Recital 46 of Directive 2000/31/EC (*Directive on electronic commerce*).

[1255] Perry & Zarsky, 'Who Should Be Liable for Online Anonymous Defamation?', *University of Chicago Law Review Dialogue*, 2015.

[1256] Perry & Zarsky, 'Who Should Be Liable for Online Anonymous Defamation?', *University of Chicago Law Review Dialogue*, 2015, p. 172.

The counter-notice procedure of the DMCA has similar rationality.[1257] In the counter-notice procedure, the provider is exempted from liability when the user who provided the information with (possible) infringing content assumes responsibility for it by providing its personalia.[1258] The counter-notice seeks to protect the user from over-removal. At the same time, the provider is exempted from liability for leaving (allegedly) infringing content up when the user submits a counter-notice. In terms of liability, the user replaces the service provided for the (possibly) infringing content of the information the user provided to the service.[1259]

The DMCA approach has two downsides alien to the residual liability approach offered by Perry and Zarsky. While the residual liability approach is argued to reduce the cost of compliance,[1260] the DMCA approach has the opposite effect. The cost of compliance is very high because of the administrative burden involved and may even function as a barrier for new providers.[1261] Besides, de DMCA may still lead to removal by default because the counter-notice procedure is rarely used.[1262] The counter-notice procedure does not remove the incentive to remove user-provided information.[1263] For the user, submitting a counter-notice raises a significant risk because by submitting a notice, the user is exposed to liability for possible infringements. Noteworthy to mention is that the DSA, unlike the DMCA, does not know such risks.

Not all liability regimes work as well for all categories of user-provided information. Applying residual liability approaches to providers that encounter, for example, sexual child abuse material or user-provided information with terrorist content would have strange consequences. Under a residual liability regime, a provider would only be required to provide the personalia of the provider of the information to the authorities or a plaintiff. The provider does not become liable for the (clearly) illegal content because it fulfilled its duty while the illegal content could remain up. Such an exemption for liability from content with such severe consequences would be unacceptable in most (if not all) jurisdictions. A residual liability approach, in other words, would only work for illegal content with a clear, individualised harm. In such a case, Perry and Zarsky view removal or providing the notifier with the user's contact information responsible for the content as fitting responses. The provider is, in such a case, shielded from liability.[1264]

A residual liability approach thus would not shield providers from liability from user-provided content that is clearly illegal. The provider cannot limit itself to merely reviewing the

---

[1257] Perry and Zarsky mention the notice and takedown procedure of the DMCA under "exclusive indirect liability", see Note 37 of Perry & Zarsky, 'Who Should Be Liable for Online Anonymous Defamation?', *University of Chicago Law Review Dialogue*, 2015, p. 169.

[1258] 17 USCA § 512(g)(2)(B) (West 2010, Westlaw Next through PL 116-179).

[1259] Par. 5.36 of Wilman, 2020, *The Responsibility of Online Intermediaries for Illegal User Content in the EU and the US*, pp. 148-149.

[1260] Perry & Zarsky, 'Who Should Be Liable for Online Anonymous Defamation?', *University of Chicago Law Review Dialogue*, 2015, p. 173.

[1261] Par. 5.53 of Wilman, 2020, *The Responsibility of Online Intermediaries for Illegal User Content in the EU and the US*, p. 159.

[1262] Par. 5.54 of Wilman, 2020, *The Responsibility of Online Intermediaries for Illegal User Content in the EU and the US*, pp. 159-160.

[1263] In contrast, the provider is even shielded for liability arising from erroneous removal, see 17 USCA § 512(g)(1) (West 2010, Westlaw Next through PL 116-179).

[1264] Perry & Zarsky, 'Who Should Be Liable for Online Anonymous Defamation?', *University of Chicago Law Review Dialogue*, 2015, p. 173.

information provided by users that did not provide their contact information.[1265] A residual liability approach would be fitting for unlawful content that requires awareness of the facts or circumstances that are not known to the provider that can be provided by the subject of the harm (in the case of defamation) or the rightsholder (in the case of violations of intellectual property law). The rights infringed by defamatory statements can (generally) be well remedied by financial compensation (for example, by compensating for lost business opportunities due to reputational harms). Violations of intellectual property law are also resolved by paying damages. The opposite is true for hate speech directed at a group or user-provided information with terrorist content. Even when these types of content hurt an individual user or a small group, they are hard to remedy by financial compensation. Hate speech and terrorist content are primarily regulated for their (perceived) harm to society. Sedition and incitement to violence may pose an imminent threat. Sexual child abuse material or revenge pornography constitutes a severe violation of the (bodily) integrity of the persons depicted in this content. These content categories are of such gravity that they are considered violations of the legal order – an offence against society.[1266]

The severity of the violation and whether it is clear for the provider that user-provided information contains illegal content was also why the ECtHR was more strict in *Delfi*.[1267] When content is involved that is not clearly illegal, a notice and takedown regime is regarded as suitable.[1268] When a provider has to deal with clearly illegal or unlawful content, this increases the responsibility of providers.[1269]

The role and responsibility of the provider, however, are pivotal to overregulation and underregulation. Therefore, Laidlaw proposes a "notice-and-takedown plus"-regime in which the provider is not required to decide over the defamatory character of the content of user-provided information. Besides, the provider is not required to provide the personalia of the (potentially) (pseudo)anonymous user. As a general rule, the provider must take down user-provided information when a notice is not challenged by the user who provided the information. The notice and takedown plus procedure seek to prevent abuse by establishing safeguards requiring a good faith declaration. Besides, it is necessary to codify the requirements for the notice to prevent providers take down content as defamatory for which insufficient proof is provided. The provider should not be put in the position of discretionary power and thus (in normal circumstances) cannot decide whether the notice must be followed up. Only in the case of (repeated) abuse of the notice procedure by a notifier is the provider offered some discretion. Because of the administrative burden for providers, Laidlaw proposes that providers charge a small fee. The notice and takedown

---

[1265] In contrast to the argument of Perry & Zarsky who argue that the provider only has to monitor for user-provided information with a defamatory content published by non-verified users, see Perry & Zarsky, 'Who Should Be Liable for Online Anonymous Defamation?', *University of Chicago Law Review Dialogue*, 2015, p. 173.

[1266] See also, E. Laidlaw, 'Notice-and-Notice-Plus: A Canadian Perspective Beyond the Liability and Immunity Divide', *in* G. Frosio (Ed.) *Oxford Handbook of Online Intermediary Liability*, Oxford, Oxford University Press, 2020, doi:10.1093/oxfordhb/9780198837138.013.23, pp. 456-465.

[1267] *Delfi AS v. Estonia* [GC], no. 64569/09, § 159, ECHR 2015-II, 16 June 2015.

[1268] *Magyar Tartalomszolgáltatók Egyesülete and Index.hu Zrt v. Hungary*, no. 22947/13, § 91, 2 February 2016.

[1269] McGonagle, 2020, 'Free Expression and Internet Intermediaries: The Changing Geometry of European Regulation', p. 473.

plus-procedure is required to have the necessary due process features, including a dispute procedure for the user who provided the information with the alleged defamatory content.[1270]

Conditional liability regimes provide a powerful incentive for providers to exploit their services so that they fulfil the requirements necessary to rely on the safe harbours while not fulfilling the conditions that cause them to forfeit them. Consequently, conditional liability regimes may lead to overregulation because providers gaining knowledge or awareness may choose to be safe than sorry. Besides, conditional liability regimes may lead to underregulation because providers refrain from voluntary monitoring their services. Perry and Zarsky, therefore, propose to 1) offer more clarity on what providers and activities can profit from the safe harbours offered by the Directive, 2) a 'Good Samaritan' clause similar to Section 230, 3) more clarity on the liability of internet intermediaries, 4) clarity about what amounts to knowledge under the Directive, and 5) clarity about when content is 'apparent' illegal.[1271]

The unclarity of the conditions of a conditional immunity approach can be contrasted with the clarity that immunity regimes offer, which is discussed in the next paragraph.

5.2.3 Content regulation under immunity regimes

Unlike strict liability and conditional liability regimes, immunity regimes are less common. As noted, the most notable immunity regime is the regime provided by Section 230.[1272] As mentioned before, Section 230 offers better protections for providers against liability for the content of user-provided information than Article 14 of the Directive.[1273]

Immunity regimes

Granting providers unconditional immunity for the content of user-provided information does not encourage providers to do anything. Offering immunity for the content of user-provided information and moderation efforts does not encourage the service to alter any moderation practices the provider might or might not have. As noted in Chapter 3, Section 230(c)(2) does also not provide a remedy against potential overregulation that qualifies as wrongful moderation. Section 230(c)(2)(A) even protects "Good faith moderation" of user-provided information "whether or not such material is constitutionally protected".[1274]

However, repealing or amending Section 230 is tied to harmful effects on providers. For example, Fishback argues that Section 230 was necessary for providers to allow user-provided information. Fishback: "[w]ithout Section 230, it is arguable that we would have little to no platforms that allow user content."[1275] Nowadays, very large service providers would not necessarily cease to offer such functionalities. In contrast, smaller providers would have to close

---

[1270] Laidlaw, 2020, 'Notice-and-Notice-Plus: A Canadian Perspective Beyond the Liability and Immunity Divide', pp. 453-455.

[1271] Perry & Zarsky, 'Who Should Be Liable for Online Anonymous Defamation?', *University of Chicago Law Review Dialogue*, 2015, pp. 878-879.

[1272] 47 USCA § 230 (West 2018, Westlaw Next through PL 116-91).

[1273] Goldman, 2020, 'An Overview of the United States' Section 230 Internet Immunity', p. 168; Kosseff, 2019, *The Twenty-Six Words That Created the Internet*, pp. 152-153; Balkin, 'Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation', *U.C. Davis Law Review*, 2018, pp. 152-153.

[1274] 47 USCA § 230 (West 2018, Westlaw Next through PL 116-91).

[1275] Fishback, 'How the Wolf of Wall Street Shaped the Internet: A Review of Section 230 of the Communications Decency Act', *Texas Intellectual Property Law Journal*, 2020, p. 283.

down services due to the risk of liability or the cost tied to moderation. Arguably, very large service providers would impose much stricter content moderation on their services to prevent liability.[1276]

Citron and Wittes propose to amend Section 230, limiting the protection only to intermediaries "that have taken reasonable steps to prevent or address the illegality of which plaintiffs are complaining".[1277] In doing so, Citron and Wittes propose to enact a conditional liability (or conditional immunity) approach based on fulfilling a duty of care. Such a conditional liability approach could consider the service provider's size, role, and functionalities, while a carve-out of Section 230 immunity equally exposes providers to liability. With respect to anonymous speech, Omer suggests that Section 230 could be amended to exclude anonymous speech for immunity.[1278] This proposal aligns with Perry and Zarsky's residual indirect liability approach, discussed in Paragraph 5.2.2.[1279] In my view, these proposals are all worthwhile to think through further. However, as the proposal for the DSA in the EU shows, it is hard to balance these different interests with minor adjustments to Section 230.

Regarding moderation decisions, Section 230 (combined with the First Amendment) offers a safe harbour or immunity. However, it is doubtful whether Section 230 is decisive for moderation cases. As Keller argues, "speakers in the United States have few or no legal rights when platforms take down their posts."[1280] Imposing providers that offer platform functionalities with a duty to carry user-provided information is unpopular. Balkin criticises such efforts to treat internet intermediaries like public squares. Let alone the legal objections and whether it is possible to impose such a requirement – applying the First Amendment to providers would lead to undesirable outcomes. Applying the First Amendment (as interpreted by the Supreme Court of the United States) to the relationship between providers and the users would make it impossible to moderate content that would be considered intolerable by most of their users.[1281] Such an obligation would thus be contrary to user interests.

The incentive that comes from immunity regimes

An advantage provided by immunity regimes compared to strict and conditional liability regimes is that the provider has a toolbox of remedies available to address illegal and harmful information. Where strict and conditional liability regimes lead to removing the user-provided content that is borderline illegal, under immunity regimes, the provider may choose to remedy borderline illegal content with less far-reaching measures. Such remedies may hamper the harmful effects (for example, by limiting the reach of information with illegal content).

Similar to immunity regimes, non-binding instruments that are not backed by liability could set a standard but offer little legal incentives to providers to adjust their moderation. These non-liability regimes are discussed in the next paragraph.

[1276] As was the case with FOSTA, see Goldman, 'The Complicated Story of FOSTA and Section 230', *First Amendment Law Review*, 2019, pp. 288-289.

[1277] Citron & Wittes, 'The Internet Will Not Break: Denying Bad Samaritans § 230 Immunity', *Fordham Law Review*, 2017, p. 420.

[1278] Omer, 'Intermediary Liability for Harmful Speech: Lessons from Abroad', *Harvard Journal of Law & Technology*, 2014, pp. 319-320.

[1279] Perry & Zarsky, 'Who Should Be Liable for Online Anonymous Defamation?', *University of Chicago Law Review Dialogue*, 2015, p. 172.

[1280] Keller, 2019, 'Who Do You Sue? State and Platform Hybrid Power over Online Speech', p. 2.

[1281] Balkin, 'Free Speech is a Triangle', *Columbia Law Review*, 2018, pp. 2026-2027.

5.2.4 Content regulation based on other regulatory instruments

As noted in Chapter 2, internet content regulation is not necessarily the result of state legislation nor the liability of providers. Regulation may also originate from an autonomous decision by providers. Besides, providers may decide to enact content regulation in consultation or cooperation with other actors – including state actors. Codes of conduct are an example of such cooperation.

Codes of conduct can address both illegal and lawful (but considered harmful) content of user-provided information. Of course, illegal content would generally be regarded as harmful within a given jurisdiction (why would the state otherwise qualify content as illegal in its legislation?). Not all content that is considered harmful is also declared illegal. There may be categories of content considered harmful by state actors, but for which regulation would conflict with freedom of expression rights, making it hard or undesirable to regulate these categories in legislation. There is a clear desire for regulation in other cases, but the category is a fast-moving target or a multi-headed beast that is not easily captured in legislation. For example, regulating online content by drafting criminal codes can lead to underregulation or overregulation because criminal codes must be precisely formulated and foreseeable in their application.

State actors in such a case thus turn to other instruments than legislation. While legislation can be referred to as "hard law", non-legislative instruments are often called "soft law". Soft law, according to Senden, can be best defined as:

> Rules of conduct that are laid down in instruments which have not been attributed legally binding force as such, but nevertheless may have (certain) indirect legal effects, and that are aimed at and may procedure practical effects.[1282]

Soft law can be either the result of cooperation between non-governmental actors, including the providers[1283] or concluded under the auspices of state actors. State actors typically fulfil a coordinating role in concluding such agreements. As noted, the e-Commerce Directive explicitly charges the EC and the Member States of the EU with the obligation to stimulate the drafting of codes of conduct at the EU level by "by trade, professional and consumer associations or organisations" for the implementation of (amongst others) Articles 14 and 15. Besides, the e-Commerce Directive stimulates drafting codes "regarding the protection of minors and human dignity."[1284]

The EC stimulates the adaptation of self-regulation in the form of voluntary agreements. In these codes of conduct, the providers agree to (legally) non-binding commitments.[1285] According to the Directive, the codes of conduct are voluntary to join or follow.[1286] One of the codes concluded at the EU level is the *Code of Conduct on Countering Illegal Hate Speech Online,* which includes obligations for providers to handle hate speech.[1287] On the Member State level, providers

---

[1282] L. Senden, *Soft Law in European Community Law*, Portland, Hart Publishing, 2004, p. 112.

[1283] See, for example, Global Internet Forum to Counter Terrorism, 'Membership', *Global Internet Forum to Counter Terrorism*, available at gifct.org/membership (retrieved on 14 February 2022).

[1284] Article 16(1)(a) and (d) of Directive 2000/31/EC (*Directive on electronic commerce*).

[1285] M.L. Montagnani, 'A New Liability Regime for Illegal Content in the Digital Single Market Strategy', *in* G. Frosio (Ed.) *Oxford Handbook of Online Intermediary Liability*, Oxford, Oxford University Press, 2020, doi:10.1093/oxfordhb/9780198837138.013.15, p. 296.

[1286] Recital 49 of Directive 2000/31/EC (*Directive on electronic commerce*).

[1287] European Commission, 2016, 'Code of Conduct on Countering Illegal Hate Speech Online'.

in the Netherlands concluded a code of conduct regarding the takedown of illegal or unlawful content such as child sexual abuse content.[1288]

Soft law instruments do not aim to replace but to complement legislation. Soft law has some advantages and disadvantages in comparison to legislation. While it takes a while to change or adopt new legislation, soft law instruments could often be more quickly adopted and revised.[1289] Soft law instruments can complement private law. For example, it could increase the accountability of private organisations and the legitimacy of the norms they carry out. Soft law instruments even can be used to incorporate public law values in private law relationships.[1290]

The involvement of state actors in drafting a code of conduct is troublesome – especially when this drafting occurs outside a legislative framework.[1291] State actors are not equal partners for providers. As shown, state actors can set out the general liability rules and enact binding legislation. When the state requires providers to regulate the content of user-provided information by legally obligating providers to do so, such obligations usually are laid down in legislation. When non-binding codes do not lead to the desired results, state actors can propose and possibly enact legislation backed by a fine. Precisely this possibility makes that Citron views soft law instruments as deployed by the EC for content moderation as "government coercion occurring outside the rule of law."[1292]

In the DSA, the character of codes of conduct changes from self-regulation to co-regulation.[1293] These codes of conduct may be concluded to address illegal content or systemic risks.[1294] Systemic risks may render participating in codes of conduct for very large online platforms "reasonable, proportionate and effective mitigation measures".[1295] When significant systemic risks involving very large online platforms are involved, the EC may invite the concerned online platforms and other intermediary providers, civil society organisations, and other parties interested in addressing a specific systemic risk.[1296] When a very large service provider refuses to partake in the drafting of such a code of conduct, this "could be taken into account, where relevant, when determining whether the online platform has infringed the obligations laid down by this Regulation."[1297]

Besides, the codes of conduct may gain binding force to end a violation of the obligations laid down in Section 4 of Chapter III of the DSA. This section imposes specific obligations on very large online platforms. The Digital Service Coordinator of a Member State in which a very large online platform is established may require the provider to draw up an action plan to end this infringement which can be given binding force backed by an independent audit.[1298] According to

---

[1288] ECP, 2018, 'NTD code'.

[1289] Mak, 2020, *Legal Pluralism in European Contract Law*, p. 125.

[1290] Bloch-Wehba, 'Global Platform Governance: Private Power in the Shadow of the State', *SMU law review*, 2019, pp. 64-65.

[1291] The DSA is remedying this by establishing a co-regualtory regime, see Article 35 of Commission Proposal COM(2020) 825 final (*Digital Services Act*), p. 65.

[1292] Citron, 'Extremist Speech, Compelled Conformity, and Censorship Creep', *Notre Dame Law Review*, 2018, p. 1070.

[1293] Recital 68 of Commission Proposal COM(2020) 825 final (*Digital Services Act*), p. 35.

[1294] Article 35(1) of Commission Proposal COM(2020) 825 final (*Digital Services Act*), p. 65.

[1295] Article 27(1) of Commission Proposal COM(2020) 825 final (*Digital Services Act*), p. 60.

[1296] Article 35(2) of Commission Proposal COM(2020) 825 final (*Digital Services Act*), p. 65.

[1297] Recital 68 of Commission Proposal COM(2020) 825 final (*Digital Services Act*), p. 35.

[1298] Article 50(2) of Commission Proposal COM(2020) 825 final (*Digital Services Act*), p. 75.

Article 50(2), this action plan may include a proposal by the provider to participate in a code of conduct.[1299]

In addition to voluntary codes of conduct, the DSA includes so-called "crisis protocols". The EC can invite very large online platforms and other online platforms to draft such protocols[1300] "for addressing crisis situations strictly limited to extraordinary circumstances affecting public security or public health."[1301] The crisis protocols mainly see to pushing authoritative information about the crisis supplied by the different authorities of the Member States, but also to DSA obligations that may need sharpening or adjustment to the crisis.[1302] The crisis protocols must contain the necessary safeguards, for example, "safeguards to address any negative effects on the exercise of the fundamental rights enshrined in the Charter"[1303] and procedures to activate and deactivate the protocols.[1304] The crisis protocols are required to have in place transparency mechanisms "to publicly report on any measures taken, their duration and their outcomes, upon the termination of the crisis situation."[1305] According to the DSA, crisis protocols are suitable for "[e]xtraordinary circumstances may entail any unforeseeable event, such as earthquakes, hurricanes, pandemics and other serious cross-border threats to public health, war and acts of terrorism" since in these events "online platforms may be misused for the rapid spread of illegal content or disinformation or where the need arises for rapid dissemination of reliable information."[1306] It is clear that the EC in drafting the DSA had disinformation in view as one of the regulatory issues.

When providers of very large online platforms participate in codes of conduct or crisis protocols, the obligations laid down in these instruments are part of an independent audit. Very large online platforms are required, according to the DSA, to pay for the audit to monitor their compliance with the codes of conduct.[1307] Non-compliance with the obligations laid down in the protocols and codes the providers are bound to may lead to a negative auditor opinion.[1308] Very large online platforms that receive a negative opinion from the auditor are required to implement the necessary measures. When the provider thinks there are reasons not to implement the recommendations, they must set out the reasons for this decision and whether they implemented other measures to remedy non-compliance.[1309] While adopting the code of conduct and the crisis protocols are voluntary, "the implementation of codes of conduct should be measurable and subject to public oversight".[1310] However, according to the recitals, very large online platforms can be required to "cooperate in the drawing-up and adhere to specific codes of conduct."[1311]

Content regulation based on other soft law instruments allows multiple remedies to be deployed. For example, the Code of Conduct, building upon already criminalised categories of

---

[1299] Article 50(2) of Commission Proposal COM(2020) 825 final (*Digital Services Act*), p. 75.
[1300] Article 37(1) and (2) of Commission Proposal COM(2020) 825 final (*Digital Services Act*), p. 66.
[1301] Article 37(1) of Commission Proposal COM(2020) 825 final (*Digital Services Act*), p. 66.
[1302] Article 37(2) of Commission Proposal COM(2020) 825 final (*Digital Services Act*), p. 66.
[1303] Article 37(4)(e) of Commission Proposal COM(2020) 825 final (*Digital Services Act*), p. 67.
[1304] Article 37(4)(c) and (d) of Commission Proposal COM(2020) 825 final (*Digital Services Act*), p. 67.
[1305] Article 37(4)(f) of Commission Proposal COM(2020) 825 final (*Digital Services Act*), p. 67.
[1306] Recital 71 of Commission Proposal COM(2020) 825 final (*Digital Services Act*), p. 35.
[1307] Article 28(1)(b) of Commission Proposal COM(2020) 825 final (*Digital Services Act*), p. 61.
[1308] Recital 61 of Commission Proposal COM(2020) 825 final (*Digital Services Act*), p. 33.
[1309] Article 28(4) of Commission Proposal COM(2020) 825 final (*Digital Services Act*), p. 61.
[1310] Recital 67 of Commission Proposal COM(2020) 825 final (*Digital Services Act*), p. 34.
[1311] Recital 67 of Commission Proposal COM(2020) 825 final (*Digital Services Act*), p. 34.

content, requires illegal content to be removed. The Code of Practice takes on another approach by countering disinformation with, for example, fact-checking instead of content removal.[1312] However, it is not a given that the providers implement the remedies laid down in soft law instruments as they are meant. The obligations laid down in the Code of Practice, for example, may result in the unsanctioned removal of content as disinformation.[1313]

5.3 What is the right liability regime?

After discussing the different models for liability regimes and how the US and European regimes fit in these models, the discussion now turns to what liability regime is most suitable to remedy (potential) legal incentives to service providers to overregulate or underregulate user-provided information. Strictly speaking, outside the scope of this dissertation, the overview provided in this chapter gives all the reasons to make a few remarks about these regimes.

The poorest candidates in terms of preventing underregulation are immunity regimes. While immunity regimes do not provide a legal incentive to overregulate or underregulate user-provided information, immunity regimes do not provide any legal incentive to regulate illegal content for which the service provider is immunised. However, immunity regimes take away legal incentives that may cause providers to refrain from moderation by shielding providers from liability that may arise from moderation.

Immunity regimes, however, put (too) much trust in the service provider to take on the responsibility to regulate illegal content. Not only "Good Samaritans" that indeed take action against illegal content profit from such an exemption, but also "Bad Samaritans"[1314] that (purposely) do not moderate. Under an immunity regime, it is hard to pressure service providers to regulate content legally without carving out the immunities. Immunity regimes are, therefore, the best shield against governmental pressure to regulate harmful content that is not strictly or necessary illegal. Immunity regimes are, due to their nature, not very sensitive to ambiguous legislation concerning (for example) hate speech or other categories of content because the service provider is likely to become in the position that they have to interpret and apply such legislation. Necessary to stipulate is that immunity regimes do not safeguard the freedom of expression rights of the users but merely the service providers' freedom of expression rights. Especially when the service provider is shielded from liability for moderation decisions, as is the case under US Section 230, the user lacks protection against overregulation. While the legal incentive to overregulate is taken away, there is no legal incentive to counter overregulation when it does occur. Immunity regimes – while taking away legal incentives to overregulate – are not helpful in remedying overregulation.

Strict liability regimes provide legal incentives that may lead to overregulation and underregulation. Because strict liability is tied to some level of editorial responsibility, service providers that exercise editorial discretion are given a powerful legal incentive to prevent illegal content from appearing on their service. The outcome of a strict liability approach can also be the complete opposite. Because of this legal burden, service providers may refrain from any editorial control (including moderation) because the service provider in such a case cannot be argued to

---

[1312] European Commission, 2021, 'Code of Practice on Disinformation'.

[1313] Rb. Amsterdam (vzr.), 9 September 2020, ECLI:NL:RBAMS:2020:4435, Rec. 4.11 (*YouTube*); Rb. Amsterdam (vzr.), 13 October 2020, ECLI:NL:RBAMS:2020:4966, Rec. 4.24, *Computerrecht* 2021/66, m.nt. M. Klos (*Facebook*).

[1314] Citron & Wittes, 'The Internet Will Not Break: Denying Bad Samaritans § 230 Immunity', *Fordham Law Review*, 2017.

have exercised any control over the content of user-provided information.[1315] Besides a strong legal incentive to overregulate, strict liability can also provide a powerful incentive not to regulate when such a regime is tied to editorial control. Such a legal incentive leads thus to underregulation of user-provided information with illegal content because the service provider takes a hand-off approach. Strict liability regimes are unlikely to be suitable because of the volume of user-provided information the service providers may have to moderate. Such a liability regime may easily overestimate the ability to review all user-provided information.

Conditional liability regimes do not require service providers to review all user-provided information for illegal content but only those instances of user-provided information that fulfil a set of criteria. For example, conditional liability approaches (such as in the EU) may require the service providers to review user-provided information pointed out to the provider. Conditional liability regimes, in sum, keeps the pool of suspicious information (relatively speaking) small and thus limiting the legal risk of the service provider to the user-provided information that is pointed out to the service providers. Only the user-provided information that exposes the service provider to liability (legally speaking) must be reviewed against applicable rules and regulations. A conditional liability approach does not leave third parties affected by user-provided information with illegal content empty-handed. Because of the conditional nature, a conditional liability approach, for example, requires reviewing user-provided information when others point out that this information contains illegal content.

However, a conditional liability approach is very vulnerable to poorly drafted legislation and misuse. Ambiguous legal definitions, requirements in how the service provider should evaluate potential illegal content, and the required remedy can cause overregulation. As discussed, a notice and takedown mechanism can be abused to pressure service providers into removing user-provided information with content that is hard to evaluate as illegal when 1) the service provider must interpret vague legal definitions, 2) within a short timeframe, and 3) with removal as the only viable remedy.

A conditional liability approach balances the risk of overregulation and underregulation. The risk of overregulation in such a regime is, for example, limited to user-provided information that is explicitly brought to the service provider's attention. Because of such a condition, others can point out user-provided information with illegal content harmful to them or others, limiting the risks of underregulation. One of the issues tied to a conditional liability regime is that moderation is entirely leftover to the service provider. The service provider must make (complex) judgements about the illegality of the content of user-provided information that is brought to its attention. The service provider may choose to remove content by default to prevent liability – especially when the removal of legal content is legally hard to contest.

The solutions proposed to remedy such a legal incentive include increasing the responsibilities of service providers for such erroneous removal, deanonymizing users and making them responsible themselves and establishing or strengthening independent oversight on content moderation. Most of the remedies for overregulation and underregulation see an *ex post*-review of the decision in question. That leaves us with the question of what must happen in between: must suspicious user-provided information remain up or taken down when the service provider is in doubt? Every intervention will result in 1) either illegal content remaining (longer) and thus causing

[1315] As discussed in Chapter 3, this was the problem at hand with *Prodigy*.

damage, or 2) in the (temporary) curtailment of the freedom of expression of individuals (partly due to state legislation).

The question, in essence, is what is acceptable in terms of underregulation or overregulation. Analogous to the *Blackstone ratio* in the case of criminal convictions,[1316] is it better to leave ten posts with illegal content up than take down one post with legal content? A conditional liability approach at least allows for an answer to this question. Balancing between the risk of overregulating user-provided information with legal content or underregulation of user-provided information with illegal content is highly context-dependent. One could imagine that in the case of child sexual abuse content, a higher error rate is acceptable in terms of overregulation because underregulation means that one of the most harmful categories of illegal content remains up. In the case of defamatory content, underregulation may be preferable because overregulation would mean that, for example, legitimate criticism or satire would be stifled. In such a case, the social costs may be higher than in the case of underregulation (leaving some defamatory content up). The theoretical framework provided by this dissertation does not provide normative answers to how such balancing should occur, nor does it answer the more normative question of what rights or costs are worth balancing. This dissertation, however, does provide the tools to address these questions in terms of internet intermediary service provider liability and how these tools provide a legal incentive to overregulate or underregulate user-provided information.

[1316] Oxford Reference, 'Blackstone ratio', *Oxford Reference*, available at oxfordreference.com/view/10.1093/oi/authority.20110803095510389 (retrieved on 23 March 2022).

Summary and conclusion

This dissertation deals with overregulation and underregulation of illegal and harmful content of user-provided information. As discussed, providers that offer functionalities for user-provided information are criticised from two perspectives. One perspective is that the providers fail in their effort to counter illegal and remedy harmful content of user-provided information. Another perspective is that providers are too actively involved in the content of user-provided information, which leads to a biased treatment of certain political viewpoints. New regulation, mainly in legislation, is proposed to counter (perceived) overregulation and underregulation. As noted, these proposals relate to already established liability regimes.

The central question of this dissertation is to what extent (proposals for) regulation applicable to providers laid down in the e-Commerce Directive in the EU and Section 230 in the US provide a legal incentive to overregulate or underregulate user-provided information. The focus lies on so-called hosting service providers, service providers that offer online platforms for user-provided information. These providers, as argued, can moderate and curate user-provided information based on its content. Content regulation may target content that can be considered illegal, unlawful, or harmful. Both underregulation and overregulation may raise concerns regarding freedom of expression rights laid down in the First Amendment (US) or Article 10 of the ECHR (the European context) in the case they would apply to the provider-user relationship.

In the first chapter, I discussed whether and to what extent providers are regulated differently from offline information intermediaries based on their characteristics. This chapter sets out that regulation is (and must) be differentiated between offline information intermediaries and providers to prevent overregulation and underregulation based on the content of the information. While providers were subjected to the same legal principles and obligations as traditional information intermediaries, legislators enacted legislation differentiating between offline information intermediaries and online providers. This legislation, however, also distinguishes between different providers. In this first chapter, I argued that it is necessary to distinguish between providers that should refrain from content-based interventions and providers that could (and perhaps even should) intervene in the content of user-provided information. As a rule of thumb, providers that offer platform functionalities for user-provided information can impose content-based restrictions (and can also be regulated by the state to do so). Service providers that merely provide the infrastructure to users and/or providers to make use of the possibilities of the internet should not engage in content-based restrictions and thus should not be required by the state to carry out such restrictions. This last category of providers, when required to regulate user-provided information, poses a significant risk of overregulation because of their technological and functional relationship with the content of information.

In Chapter 2, I explored which actors can be targeted by internet content regulation. In addition, this chapter discusses the instruments and remedies the targets of internet content regulation can deploy in regulating user-provided information. The fourth element of internet content regulation is the scope of regulation: do the instruments and remedies have extraterritorial effects outside the jurisdiction that imposes the measures? The proposals to increase the liability of internet intermediary services are numerous. As noted, this does not mean that the provider is exclusively liable for the content of the user-provided information. In principle: the user is legally responsible for the content of the information provided to the service. However, the provider is a preferred target of state regulation. The user responsible for information with illegal content is not

necessarily from the same jurisdiction as the third party that is hurt by illegal content. Besides, the responsible user may be (pseudo)anonymous and therefore hard to identify. The service provider, in contrast, can swiftly remedy the harmful effects of the illegal content of user-provided information.

As noted, providers can moderate and curate user-provided information. While moderation is discovering, evaluating, and remedying illegal content, curation encompasses how user-provided information is shown to other users. Providers may moderate or curate out of their initiative or because of a legal obligation. Moderation leads to a remedy, while curation has nothing to do with violations. However, instruments that are typically viewed as curation could be deployed as a content moderation remedy. Decisive, in this respect, is why the provider intervenes in the user-provided information. Concerning the remedies, providers have an extensive toolkit. Service providers can carry out all remedies that they can code. Service providers, however, cannot deploy coercive measures that are reserved for the State. Moderation often results in the decision to leave the content up when there is no violation or to take content down when there is a violation. As argued, a more refined approach to content moderation remedies would be less impactful to users' freedom of expression rights. The second chapter concluded with the scope of content moderation remedies. Providers operate on a global network and often provide their service to users from all over the world. Whether enacted by the state or out of the provider its initiative, internet content regulation may impact freedom of expression rights in other jurisdictions because the remedies (removal) have an extraterritorial effect.

In Part 2, the internet intermediary liability regimes in the US and the EU are discussed. Providers were granted broad discretion to moderate as they see fit in the US. Providers are excluded from civil liability for moderation decisions taken in "good faith". At the same time, providers are also exempted from liability for user-provided information as a publisher or speaker. While providers are increasingly criticised for not doing enough (regarding disinformation) or too much, this liability exemption is largely left in place. For the EU, this is different. Service providers are and were not to this extent immunised under EU law. Service providers are merely exempted from liability for user-provided information with illegal content as long as the provider had no knowledge or awareness of its existence. Besides, the EU liability approach does not offer any safe harbour to hosting service providers that are active with respect to user-provided information when this gives knowledge or control over the illegal content, which provides an incentive to refrain from voluntary moderation. Providers must remove or disable access to user-provided information with illegal content when they encounter such content. The e-Commerce Directive, however, prohibits member states of the EU from imposing a general obligation to monitor illegal content. New proposals for regulation leave the safe harbours of the e-Commerce Directive largely unaffected. New legislation (that builds upon the e-Commerce Directive) is proposed to regulate new categories of content or to impose duties of care to (mainly) providers that offer very large online platforms.

After discussing the characteristics of the US and EU intermediary regulation and the theory behind internet content regulation, the EU approach stands out as more responsive to novel threats due to the less absolute nature of the safe harbours. In the US, requiring providers to impose content-based restrictions on their users would encounter obstacles in the form of the broad and almost unconditional immunity provided by Section 230 and, of course, the First Amendment. Because the e-Commerce Directive has a conditional nature, content-based

restrictions can easily be added in the European context by imposing new responsibilities to providers.

The legal incentives from different regulatory regimes in the US and the European context are discussed in the fifth and final chapter. Do these liability regimes incentivise providers to underregulate or overregulate user-provided information? Based on how providers regulate user-provided content as set out in Chapter 2, three factors are discussed to contribute to overregulation or underregulation. The first factor is the ambiguity of unclear and unpredictable legislation that applies to providers. Ambiguous definitions in legislation especially force providers to overregulate user-provided information based on its content. The second factor is the instruments that are required by legislation. Especially automatic content moderation systems deployed to evaluate content are tied to overregulation. Besides, requiring providers to proactively moderate content may lead to overregulation. Ambiguity and these means are primarily tied to overregulation when combined with the obligation to remove user-provided information when it contains violating content. Ambiguity, for example, mainly raises direct freedom of expression concerns when providers turn to removal as a default option – even when the provider is not sure that the content in question indeed violates the rules. Special attention, in this respect, must be paid to requiring providers to include a prohibition for specific content in their terms and conditions. These terms and conditions directly influence the relationship between the provider and the user. Because the terms and conditions are often applied globally, the terms and conditions have a global effect. Such passive obligations have little direct influence on the freedom of expression rights of the users. However, because the terms and conditions form the foundation for the means and the remedies, such passive instruments may lead to extraterritorial overregulation.

So far, overregulation seems a far more significant risk than underregulation. Such a conclusion, however, would be wrong. Overregulation is clearly based on the fear of being held liable. Overregulation – and regulation in itself – has a strong dependency on the liability regime, or in the absence of such a regime, at least the possibility of introducing legislation imposing such liability. Without obligations backed by some legal liability, it is unlikely that all providers would moderate user-provided information as consequently as if it were backed by such a liability. So much is shown by immunity regimes, for example, as provided by Section 230.

Strict liability regimes hold the most risks for both overregulation and underregulation. Overregulation under strict liability regimes occurs when the provider is held liable for all user-provided information with illegal content. Then, the provider is regarded as a publisher for the content and thus expected to filter illegal content proactively. Both editorial responsibilities as proactive filtering are expected to result in overregulation because the provider is expected to turn to removal when in doubt. In addition, the provider may also cease to offer services out of fear of legal liability. Strict liability is only possible when providers perform editorial activities in some jurisdictions. In these jurisdictions, providers may cease all moderation efforts because the provider is then no longer an editor and thus not exposed to strict liability. Strict liability approaches that function this way may thus foster underregulation: both the state and the provider would wish to regulate more than the liability regime incentivises to moderate. In addition, strict liability regimes leave little room to decide on the remedies: only removing the content in question exempts the provider from being liable.

Conditional liability regimes have little risk for underregulation and a decreased risk for overregulation than strict liability regimes. Conditional liability regimes require service providers to take action against user-provided information when fulfilling a set of conditions. These

conditions mainly see to some knowledge of the provider regarding the illegal content. Conditional liability regimes normally do not incentivise providers to use proactive moderation. However, when the provider gains knowledge of the illegal content, the provider has a strong incentive to remove the content in question to prevent becoming liable. As discussed, the requirement for providers to act "expeditiously" and remove content that violates ambiguous legislation within limited timeframes may lead to overregulation. At the same time, providers may be cautious in voluntary monitoring their services for illegal content because knowledge of illegal content may render them liable. As noted, especially the liability regime of the EU raises this risk.

Immunity from legal liability, as noted, has the most significant risk for underregulation because providers cannot be held liable for illegal content on their services. While exemptions from liability for overregulation may provide a climate in which providers can moderate user-provided information as they see fit, such an exemption from liability does not provide such an incentive. Immunity, however, most likely does not lead to overregulation now that the provider has discretion in what the provider deems. Because immunity regimes do not require providers to deploy instruments or a specific set of sanctions, immunity regimes allow a more refined set of remedies that may follow a rule violation.

As noted, there are also non-liability regimes in the form of, for example, voluntary codes of conduct. This fourth possibility is characterised as non-liability regulation. However, these regulatory instruments are the most effective in regimes in which a legal liability regime does exist. Voluntary codes are then easily translated to legislation, exposing the provider to liability. Non-liability regulation, thus, may also lead to overregulation because the provider may fear that non-compliance may lead to new legislation. In addition, the provider may similarly adhere to such voluntary codes as to legal obligations under the liability regime. Because of how providers implement these codes, less harmful content (such as disinformation) could be treated similar to illegal content. While a more nuanced remedy may be more desirable, providers have an incentive to remove such content.

Because liability regimes operate in a broader legal environment, not all liability regimes are possible in all jurisdictions. For example, Section 230-like immunity would be incompatible with the ECHR, which requires a more nuanced balancing of speech and (for example) privacy rights.[1317] The liability regime, thus, has to be tailored to the human rights standards within a given jurisdiction. The EU approach offers more of a balance between freedom of expression rights and the rights of others. In addition, a conditional liability approach knows considerable advantages in countering new threats. However, a too light-hearted approach in making providers liable for the content of user-provided information may lead to adverse effects regarding freedom of expression rights. Especially when the conditions that the provider has to fulfil are unclear or increasingly complicated, a conditional liability approach may lose its competitive value. Unclear conditions may lead to very similar effects as strict liability approaches. Because the provider is unsure whether the provider could rely on the exemptions from liability, the provider may engage in extensive overregulation, cease (parts) of the service, or cease all moderation efforts to prevent gaining knowledge.

Liability regimes should clarify when the provider becomes liable, leave room for some error, and not 'hard-code' all potential reactions of the service provider. The threat to freedom of

[1317] Husovec, 'General monitoring of third-party content: compatible with freedom of expression?', *Journal of Intellectual Property Law & Practice*, 2016, p. 120.

expression rights does not come from wrongful moderation from providers engaging in moderating out of their initiative. In contrast: the real threat is state pressure backed by state legislation that is not sufficiently clear and signals that the provider should do better or else.

References

Bibliography

'Draft Online Safety Bill', *Department for Digital, Culture, Media & Sport,* 12 May 2021, available at gov.uk/government/publications/draft-online-safety-bill (retrieved on 15 February 2022).

Adviesraad Internationale Vraagstukken, 'Regulering van online content: Naar een herijking van het Nederlandse internetbeleid (AIV-advies 113)', *Adviesraad Internationale Vraagstukken,* 2020, available at adviesraadinternationalevraagstukken.nl/documenten/publicaties/2020/06/24/regulering-van-online-content (retrieved on 14 February 2022).

Angelopoulos, C., et al., 'Study of fundamental rights limitations for online enforcement through self-regulation Institute', *Institute for Information Law (IViR),* 2015, available at hdl.handle.net/11245.1/7317bf21-e50c-4fea-b882-3d819e0da93a.

Anti-Defamation League, 'Stop Hate for Profit', *Anti-Defamation League,* 16 June 2021, available at stophateforprofit.org (retrieved on 14 February 2022).

Baker, M., T. Berners-Lee & V. Cerf, 'EU Terrorist Content regulation will damage the internet in Europe without meaningfully contributing to the fight against terrorism', *Politico,* 2 April 2019, available at politico.eu/wp-content/uploads/2019/04/TCO-letter-to-rapporteurs.pdf (retrieved on 14 February 2022).

Balkin, J., 'Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation', *U.C. Davis Law Review*, Vol. 51, No. 3, 2018, pp. 1149-1210.

Balkin, J., 'Free Speech is a Triangle', *Columbia Law Review*, Vol. 118, No. 7, 2018, pp. 2011-2055.

Barlow, J., 'A Declaration of the Independence of Cyberspace', *Electronic Frontier Foundation,* 8 February 1996, available at eff.org/nl/cyberspace-independence (retrieved on 14 February 2022).

BBC, 'Twitter fears for freedom of expression in India', *BBC,* 27 May 2021, available at bbc.com/news/world-asia-india-57265331 (retrieved on 14 February 2022).

Benedek, W. & M.C. Kettemann, *Freedom of Expression and the Internet*, Strasbourg, Council of Europe Publishing, 2020.

Bloch-Wehba, H., 'Global Platform Governance: Private Power in the Shadow of the State', *SMU law review*, Vol. 72, No. 1, 2019, pp. 27-80.

Blommestijn, R. & M. Klos, 'Een giftige paddenstoel voor de vrijheid van meningsuiting: Bol.com en het verbieden van 'foute' boeken', *Nederlands Juristenblad*, 2020/1209, pp. 1388-1394.

Body of European Regulators for Electronic Communications, 'What is zero-rating?', *Body of European Regulators for Electronic Communications*, available at berec.europa.eu/eng/netneutrality/zero_rating/ (retrieved on 14 February 2022).

Bradford, A., 'The Brussels Effect', *Northwestern University Law Review*, Vol. 107, No. 1, 2012, pp. 1-68.

Bradford, A., *The Brussels Effect: How the European Union Rules the World*, New York, Oxford University Press, 2020.

Brandom, R., 'Jessica Rosenworcel confirmed by Senate to lead the FCC', *The Verge,* 7 December 2021, available at theverge.com/2021/12/7/22820873/jessica-rosenworcel-fcc-chair-confirmed-biden-net-neutrality (retrieved on 28 January 2022).

Bridy, A., 'Remediating Social Media: A Layer-Conscious Approach', *Boston University Journal of Science and Technology Law*, Vol. 24, No. 2, 2018, pp. 193-228.

Britannica, 'Internet service provider', *Encyclopedia Britannica,* 13 March 2018, available at britannica.com/technology/Internet-service-provider (retrieved on 14 February 2022).

Brunner, L., 'The Liability of an Online Intermediary for Third Party Content: The Watchdog Becomes the Monitor: Intermediary Liability after Delfi v Estonia', *Human Rights Law Review*, Vol. 16, No. 1, 2016, doi:10.1093/hrlr/ngv048, pp. 163-174.

Byford, S., 'Gab.com goes down after GoDaddy threatens to pull domain', *The Verge,* 28 October 2018, available at theverge.com/2018/10/28/18036520/gab-down-godaddy-domain-blocked (retrieved on 14 February 2022).

Caglayan, C., et al., 'YouTube says to appoint Turkey representative in line with new law', *Reuters,* 16 December 2020, available at reuters.com/article/us-turkey-socialmedia-youtube-idUSKBN28Q1T2 (retrieved on 14 February 2022).

Callamard, A., 'Are courts re-inventing Internet regulation?', *International Review of Law, Computers & Technology*, Vol. 31, No. 3, 2017, doi:10.1080/13600869.2017.1304603, pp. 323-339.

Chowdhury, R. & L. Belli, 'Examining algorithmic amplification of political content on Twitter', *Twitter Blog,* 21 October 2021, available at blog.twitter.com/en_us/topics/company/2021/rml-politicalcontent (retrieved on 14 February 2022).

Citron, D., 'Extremist Speech, Compelled Conformity, and Censorship Creep', *Notre Dame Law Review*, Vol. 93, No. 3, 2018, pp. 1035-1072.

Citron, D. & B. Wittes, 'The Internet Will Not Break: Denying Bad Samaritans § 230 Immunity', *Fordham Law Review*, Vol. 86, No. 2, 2017, pp. 401-424.

Citron, D.K., 'Fix Section 230 and hold tech companies to account', *Wired,* 6 May 2021, available at wired.co.uk/article/section-230-social-media (retrieved on 14 February 2022).

Cole, S., 'Sex Workers Say Porn on Google Drive Is Suddenly Disappearing', *Vice,* 21 March 2018, available at vice.com/en/article/9kgwnp/porn-on-google-drive-error (retrieved on 14 February 2022).

Committee of Ministers, 'Declaration of the Committee of Ministers on network neutrality (Adopted by the Committee of Ministers on 29 September 2010 at the 1094th meeting of the Ministers' Deputies)', *Council of Europe,* 29 September 2010, available at rm.coe.int/09000016805ce58f (retrieved on 14 February 2022).

Committee of Ministers, 'Recommendation CM/Rec(2016)1 of the Committee of Ministers to member States on protecting and promoting the right to freedom of expression and the right to private life with regard to network neutrality (Adopted by the Committee of Ministers on 13 January 2016, at the 1244th meeting of the Ministers' Deputies)', *Council of Europe,* 13 January 2016, available at rm.coe.int/09000016805c1e59 (retrieved on 14 February 2022).

Council of Europe, 'Comparative Study on Blocking, Filtering and Take Down of Illegal Internet Content', *Council of Europe,* 2017, available at edoc.coe.int/en/internet/7289-pdf-comparative-study-on-blocking-filtering-and-take-down-of-illegal-internet-content-.html.

Council of Europe, 'Content moderation: best practices towards effective legal and procedural frameworks for self-regulatory and co-regulatory mechanisms of content moderation' (Guidance Note adopted by the Steering Committee for Media and Information Society (CDMSI) at its 19th plenary meeting, 19-21 May 2021), Strasbourg, Council of Europe, 2021, available at rm.coe.int/0900001680a2cc18.

Craigslist, 'FOSTA', *Craigslist*, available at craigslist.org/about/FOSTA (retrieved on 14 February 2022).

D'anastasio, C., 'Twitch Sues Users Over Alleged 'Hate Raids' Against Streamers', *Wired,* 10 September 2021, available at wired.com/story/twitch-sues-users-over-alleged-hate-raids (retrieved on 14 February 2022).

De Gregorio, G., 'Democratising online content moderation: A constitutional framework', *Computer Law & Security Review*, Vol. 36, No. 105374, 2020, doi:10.1016/j.clsr.2019.105374.

De La Chapelle, B. & P. Fehlinger, 'From Legal Arms Race to Transnational Cooperation', *in* G. Frosio (Ed.) *The Oxford Handbook of Online Intermediary Liability*, Oxford, Oxford University Press, 2020, doi:10.1093/oxfordhb/9780198837138.013.38, pp. 727-748.

Van Dijck, J., T. De Winkel & M.T. Schäfer, 'Deplatformization and the governance of the platform ecosystem', *New Media & Society*, 2021 (available at journals.sagepub.com/doi/full/10.1177/14614448211045662), doi:10.1177/14614448211045662.

Dinwoodie, G., 'Who are Internet Intermediaries?', *in* G. Frosio (Ed.) *Oxford Handbook of Online Intermediary Liability*, Oxford, Oxford University Press, 2020, doi:10.1093/oxfordhb/9780198837138.013.2, pp. 37-56.

douek, e., 'The Rise of Content Cartels', *Knight Columbia,* 11 February 2020, available at knightcolumbia.org/content/the-rise-of-content-cartels (retrieved on 14 February 2022).

ECP, 'NTD code', *ECP,* 2018, available at noticeandtakedowncode.nl/ntd-code (retrieved on 14 February 2022).

Van Eecke, P., 'Online service providers and liability: A plea for a balanced approach', *Common Market Law Review*, Vol. 48, No. 5, 2011, pp. 1455–1502.

Electronic Frontier Foundation, 'CDA 230: Key Legal Cases', *Electronic Frontier Foundation*, available at eff.org/issues/cda230/legal (retrieved on 14 February 2022).

Facebook, 'COVID-19 policy updates and protections', *Facebook Help Center*, available at facebook.com/help/230764881494641 (retrieved on 14 February 2022).

Fishback, G., 'How the Wolf of Wall Street Shaped the Internet: A Review of Section 230 of the Communications Decency Act', *Texas Intellectual Property Law Journal*, Vol. 28, No. 2, 2020, pp. 275-296.

FT Reporters, 'Tying up the internet', *Financial Times,* 16 September 2014, available at ft.com/content/2f2f7274-3a5e-11e4-bd08-00144feabdc0 (retrieved on 18 June 2021).

Fukuyama, F., et al., 'Middleware for Dominant Digital Platforms: A Technological Solution to a Threat to Democracy', *Stanford Cyber Policy Center,* 2021, available at cyber.fsi.stanford.edu/content/biden-recommendations-cyber-policy-center (retrieved on 14 February 2022).

Geiger, A.W., 'Key findings about the online news landscape in America', *Pew Research Center,* 11 September 2019, available at pewresearch.org/fact-tank/2019/09/11/key-findings-about-the-online-news-landscape-in-america (retrieved on 14 February 2022).

Geist, M., 'The Equustek Effect: A Canadian Perspective', *in* G. Frosio (Ed.) *The Oxford Handbook of Online Intermediary Liability*, Oxford, Oxford University Press, 2020, doi:10.1093/oxfordhb/9780198837138.013.37, pp. 709-726.

Geist, M., 'Picking Up Where Bill C-10 Left Off: The Canadian Government's Non-Consultation on Online Harms Legislation', *Michael Geist,* 30 July 2021, available at michaelgeist.ca/2021/07/onlineharmsnonconsult (retrieved on 14 February 2022).

Gillespie, T., *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions that Shape Social Media*, New Haven, Yale University Press, 2018.

Global Internet Forum to Counter Terrorism, 'Membership', *Global Internet Forum to Counter Terrorism*, available at gifct.org/membership (retrieved on 14 February 2022).

Goldman, E., 'Unlawful Internet Gambling Enforcement Act of 2006', *Technology & Marketing Law Blog,* 13 December 2006, available at blog.ericgoldman.org/archives/2006/12/unlawful_intern.htm (retrieved on 14 February 2022).

Goldman, E., 'The Third Wave of Internet Exceptionalism', *in* B. Szoka & A. Marcus (Eds.), *The Next Digital Decade: Essays On The Future Of The Internet*, Washington, D.C., TechFreedom, 2010, pp. 165-168.

Goldman, E., 'Online User Account Termination and 47 U.S.C. §230(c)(2)', *UC Irvine Law Review*, Vol. 2, No. 2, 2012 (available at escholarship.org/uc/item/1hh0t3w6), pp. 659-673.

Goldman, E., 'First Amendment Protects Google's De-Indexing of "Pure Spam" Websites–e-ventures v. Google', *Technology & Marketing Law Blog,* 9 February 2017, available at blog.ericgoldman.org/archives/2017/02/first-amendment-protects-googles-de-indexing-of-pure-spam-websites-e-ventures-v-google.htm (retrieved on 14 February 2022).

Goldman, E., 'Failure-to-Warn Claim Against Match.com Fails–Beckman v. Match.com', *Technology & Marketing Law Blog,* 27 November 2018, available at blog.ericgoldman.org/archives/2018/11/failure-to-warn-claim-against-match-com-fails-beckman-v-match-com.htm (retrieved on 14 February 2022).

Goldman, E., 'The Complicated Story of FOSTA and Section 230', *First Amendment Law Review*, Vol. 17, 2019 (available at firstamendmentlawreview.org/volume-17), pp. 279-293.

Goldman, E., 'Why Section 230 Is Better Than the First Amendment', *Notre Dame Law Review*, Vol. 95, No. 1, 2019 (available at scholarship.law.nd.edu/ndlr_online/vol95/iss1/3), pp. 33-46.

Goldman, E., 'An Overview of the United States' Section 230 Internet Immunity', *in* G. Frosio (Ed.) *The Oxford Handbook of Online Intermediary Liability*, Oxford, Oxford University Press, 2020, doi:10.1093/oxfordhb/9780198837138.013.8, pp. 155-171.

Goldman, E., 'Content Moderation Remedies', *Michigan Technology Law Review*, Vol. 28, No. 1, 2021, doi:10.36645/mtlr.28.1.content.

Goldman, E., 'The Ninth Circuit's Confusing Ruling Over Snapchat's Speed Filter–Lemmon v. Snap', *Technology & Marketing Law Blog,* 21 May 2021, available at blog.ericgoldman.org/archives/2021/05/the-ninth-circuits-confusing-ruling-over-snapchats-speed-filter-lemmon-v-snap.htm (retrieved on 14 February 2022).

Goldman, E. & J. Miers, 'Regulating Internet Services by Size', *CPI Antitrust Chronicle*, 2021 (available at ssrn.com/abstract=3863015).

Goldsmith, J. & T. Wu, *Who Controls the Internet: Illusions of a Borderless World*, New York, Oxford University Press, 2008.

Google, 'COVID-19 medical misinformation policy', *YouTube Help,* 20 May 2020, available at support.google.com/youtube/answer/9891785 (retrieved on 15 February 2022).

Gorwa, R., R. Binns & C. Katzenbach, 'Algorithmic content moderation: Technical and political challenges in the automation of platform governance', *Big data & society*, Vol. 7, No. 1, 2020, doi:10.1177/2053951719897945, p. 1.

Goujard, C., 'German Facebook ruling boosts EU push for stricter content moderation', *Politico,* 29 July 2021, available at politico.eu/article/german-court-tells-facebook-to-reinstate-removed-posts (retrieved on 15 February 2022).

Government of Canada, 'Consultation closed: The Government's proposed approach to address harmful content online - Discussion guide', *Government of Canada,* 29 July 2021, available at canada.ca/en/canadian-heritage/campaigns/harmful-online-content/discussion-guide.html (retrieved on 15 February 2022).

Government of Canada, 'Consultation closed: The Government's proposed approach to address harmful content online - Discussion guide - Technical paper', *Government of Canada,* 29 July 2021, available at canada.ca/en/canadian-heritage/campaigns/harmful-online-content/technical-paper.html (retrieved on 15 February 2022).

Grimmelmann, J., 'Speech Engines', *Minnesota Law Review*, Vol. 98, No. 3, 2014 (available at scholarship.law.umn.edu/mlr/299), pp. 868-952.

Guiora, A. & E. Park, 'Hate Speech on Social Media', *Philosophia*, Vol. 45, No. 3, 2017, doi:10.1007/s11406-017-9858-4, pp. 957-971.

Harmon, E., 'No, Section 230 Does Not Require Platforms to Be "Neutral"', *Electronic Frontier Foundation,* 12 April 2018, available at eff.org/deeplinks/2018/04/no-section-230-does-not-require-platforms-be-neutral (retrieved on 15 February 2022).

Hautala, L., 'PayPal and Shopify remove Trump-related accounts, citing policies against supporting violence', *cnet,* 7 April 2021, available at cnet.com/news/paypal-and-shopify-remove-trump-related-accounts-citing-policies-against-supporting-violence (retrieved on 15 February 2022).

High Level Group on Fake News and Online Disinformation, *A multi-dimensional approach to disinformation: Report of the independent High level Group on fake news and online disinformation*, Luxembourg, Publications Office of the European Union, 2018, doi:10.2759/739290.

Van Hoboken, J., et al., 'Het juridisch kader voor de verspreiding van desinformatie via internetdiensten en de regulering van politieke advertenties', Amsterdam, IVIR, 2019, available at ivir.nl/publicaties/download/Rapport_desinformatie_december2019.pdf.

Van Hoboken, J. & D. Keller, 'Design Principles for Intermediary Liability Laws', *Transatlantic High Level Working Group on Content Moderation Online and Freedom of Expression*, 2019, available at ivir.nl/nl/twg/publications-transatlantic-working-group.

Van Hoboken, J., et al., *Hosting intermediary services and illegal content online: An analysis of the scope of article 14 ECD in light of developments in the online service landscape*, Luxembourg, Publications Office of the EU, 2018, doi:10.2759/284542.

Hochman, N., 'Conservatives Should Support Section 230 Reform', *National Review,* 16 October 2021, available at nationalreview.com/2021/10/conservatives-should-support-section-230-reform (retrieved on 15 February 2022).

Holland, H.B., 'In Defense of Online Intermediary Immunity: Facilitating Communities of Modified Exceptionalism', *University of Kansas Law Review*, Vol. 56, No. 2, 2008, doi:10.17161/1808.19996, pp. 369-404.

Van Huijstee, M., et al., 'Online ontspoord: Een verkenning van schadelijk en immoreel gedrag op het internet in Nederland', *Rathenau Instituut,* 7 July 2021, available at rathenau.nl/nl/digitaal-samenleven/online-ontspoord (retrieved on 15 February 2022).

Hull, J., 'Google Hummingbird: Where No Search Has Gone Before', *Wired,* 15 October 2013, available at wired.com/insights/2013/10/google-hummingbird-where-no-search-has-gone-before (retrieved on 15 February 2022).

Hulsen, S., 'Het zwartepietbeleid van sociale media: 'YouTube trekt rookgordijn op", *nu.nl,* 30 November 2019, available at nu.nl/tech/6014528/het-zwartepietbeleid-van-sociale-media-youtube-trekt-rookgordijn-op.html (retrieved on 15 February 2022).

Husovec, M., 'General monitoring of third-party content: compatible with freedom of expression?', *Journal of Intellectual Property Law & Practice*, Vol. 11, No. 1, 2016, doi:10.1093/jiplp/jpv200, pp. 17-20.

Jee, C., 'Facebook needs 30,000 of its own content moderators, says a new report', *Technology Review,* 8 June 2020, available at technologyreview.com/2020/06/08/1002894/facebook-needs-30000-of-its-own-content-moderators-says-a-new-report (retrieved on 15 February 2022).

Jones, M.L., 'Silencing Bad Bots: Global, Legal and Political Questions for Mean Machine Communication', *Communication Law and Policy*, Vol. 23, No. 2, 2018, doi:10.1080/10811680.2018.1430418, pp. 159-195.

Judd, D., M. Vazquez & D. O'Sullivan, 'Biden says platforms like Facebook are 'killing people' with Covid misinformation', *CNN,* 17 July 2021, available at edition.cnn.com/2021/07/16/politics/biden-facebook-covid-19/index.html (retrieved on 15 February 2022).

Kastrenakes, J., 'The FCC just killed net neutrality', *The Verge,* 14 December 2017, available at theverge.com/2017/12/14/16776154/fcc-net-neutrality-vote-results-rules-repealed (retrieved on 15 February 2022).

Kastrenakes, J., 'Trump's new FCC chief is Ajit Pai, and he wants to destroy net neutrality', *The Verge,* 23 July 2017, available at theverge.com/2017/1/23/14338522/fcc-chairman-ajit-pai-donald-trump-appointment (retrieved on 15 February 2022).

Kayali, L., 'Europe's struggle against viral terrorist content', *Politico,* 21 May 2019, available at politico.eu/article/how-europe-plans-to-fight-christchurch-style-viral-content-its-complicated-fake-news-social-media-facebook-twitter-eu-terrorism (retrieved on 15 February 2022).

Keller, D., 'Who Do You Sue? State and Platform Hybrid Power over Online Speech', *Aegis Paper Series* No. 1902, Hoover Working Group on National Security, Technology, and Law, 2019, available at lawfareblog.com/who-do-you-sue-state-and-platform-hybrid-power-over-online-speech.

Keller, D., 'Empirical Evidence of "Over-Removal" by Internet Companies Under Intermediary Liability Laws', *The Center for Internet and Society,* 8 May 2020, available at cyberlaw.stanford.edu/blog/2015/10/empirical-evidence-over-removal-internet-companies-under-intermediary-liability-laws (retrieved on 15 February 2022).

Keller, D., 'The Future of Platform Power: Making Middleware Work', *Journal of Democracy*, Vol. 32, No. 3, 2021 (available at muse.jhu.edu/article/797795), doi:10.1353/jod.2021.0043, pp. 168-172.

Keller, D., 'Six Constitutional Hurdles for Platform Speech Regulation', *The Center for Internet and Society,* 22 January 2021, available at cyberlaw.stanford.edu/blog/2021/01/six-constitutional-hurdles-platform-speech-regulation-0 (retrieved on 15 February 2022).

Kimball, W., 'Biden Nominates Net Neutrality Champion Jessica Rosenworcel to Head the FCC', *Gizmodo,* 26 October 2021, available at gizmodo.com/biden-nominates-net-neutrality-champion-jessica-rosenwo-1847938641 (retrieved on 15 February 2022).

Klein, B., M. Vazquez & K. Collins, 'Biden backs away from his claim that Facebook is 'killing people' by allowing Covid misinformation', *CNN,* 20 July 2021, available at edition.cnn.com/2021/07/19/politics/joe-biden-facebook/index.html (retrieved on 15 February 2022).

Klonick, K., 'Facebook Under Pressure', *Slate,* 13 September 2016, available at slate.com/technology/2016/09/facebook-erred-by-taking-down-the-napalm-girl-photo-what-happens-next.html (retrieved on 15 February 2022).

Klonick, K., 'The New Governors: The People, Rules, and Processes Governing Online Speech', *Harvard Law Review*, Vol. 131, No. 6, 2018, pp. 1598-1670.

Klos, M., 'Tackling Online Freedom of Expression: the European Approach', *in* A. Ellian & R. Blommestijn (Eds.), *Reflections on democracy in the European Union*, The Hague, Eleven International Publishing, 2020, pp. 27-56.

Klos, M., "Wrongful moderation': Aansprakelijkheid van internetplatforms voor het beperken van de vrijheid van meningsuiting van gebruikers', *Nederlands Juristenblad*, 2020/2976.

Klos, M., 'Closed Online Communities and Freedom of Speech', *in* A. Ellian & P. Cliteur (Eds.), *The Open Society and its Closed Communities*, The Hague, Eleven, 2021, pp. 173-215.

Klos, M., 'De Digital Services Act: implicaties voor het recht op vrijheid van meningsuiting van gebruikers van onlineplatforms', *NTM/NJCM-bull.*, 2021/13.

Klos, M., 'Westphalian Sovereignty and the 4th Industrial Revolution: In Search of Legitimate Governmental Control over Online Content', *in* C. Sieber-Gasser & A. Ghibellini (Eds.), *Democracy and Globalization: Legal and Political Analysis on the Eve of the 4th Industrial Revolution*, Cham, Springer International Publishing, 2021, doi:10.1007/978-3-030-69154-7, pp. 81-121.

Kolodyazhnyy, A., A. Marrow & A. Osborn, 'Russia says Twitter complying with demand to remove 'banned content'', *Reuters,* 30 April 2021, available at reuters.com/technology/russia-says-twitter-is-complying-with-demand-remove-banned-content-2021-04-30 (retrieved on 15 February 2022).

Kosseff, J., 'Resources', *Jeff Kosseff*, available at jeffkosseff.com/resources (retrieved on 15 February 2022).

Kosseff, J., 'First Amendment Protection for Online Platforms', *Computer Law & Security Review*, Vol. 35, No. 5, 2019, doi:10.1016/j.clsr.2019.105340.

Kosseff, J., *The Twenty-Six Words That Created the Internet*, Ithaca, Cornell University Press, 2019.

Kosti, N., D. Levi-Faur & G. Mor, 'Legislation and regulation: three analytical distinctions', *The Theory and Practice of Legislation*, Vol. 7, No. 3, 2019, doi:10.1080/20508840.2019.1736369, pp. 169-178.

Krishan, N., 'FCC nominee's record is at odds with Biden censorship goals', *The Washington Examiner,* 30 November 2021, available at washingtonexaminer.com/policy/fcc-nominees-record-is-at-odds-with-biden-censorship-goals (retrieved on 15 February 2022).

Kruse, M., 'Can Donald Trump Survive 'Virtual Impeachment'?', *Politico,* 8 January 2021, available at politico.com/news/magazine/2021/01/08/donald-trump-capitol-insurrection-riot-impeachment-456352 (retrieved on 15 February 2022).

Kuczerawy, A., 'The EU Commission on voluntary monitoring: Good Samaritan 2.0 or Good Samaritan 0.5?', *CiTiP Blog,* 14 April 2018, available at law.kuleuven.be/citip/blog/the-eu-commission-on-voluntary-monitoring-good-samaritan-2-0-or-good-samaritan-0-5 (retrieved on 15 February 2022).

Kuczerawy, A., 'From 'Notice and Takedown' to 'Notice and Stay Down': Risks and Safeguards for Freedom of Expression', *in* G. Frosio (Ed.) *The Oxford Handbook of Online Intermediary Liability*, Oxford, Oxford University Press, 2020, doi:10.1093/oxfordhb/9780198837138.013.27, pp. 525-543.

Kuczerawy, A. & P.-J. Ombelet, 'Not so different after all? Reconciling Delfi vs. Estonia with EU rules on intermediary liability', *CiTiP Blog,* 2 July 2015, available at law.kuleuven.be/citip/blog/not-so-different-after-all-reconciling-delfi-vs-estonia-with-eu-rules-on-intermediary-liability (retrieved on 15 February 2022).

Laidlaw, E., 'Notice-and-Notice-Plus: A Canadian Perspective Beyond the Liability and Immunity Divide', *in* G. Frosio (Ed.) *Oxford Handbook of Online Intermediary Liability*, Oxford, Oxford University Press, 2020, doi:10.1093/oxfordhb/9780198837138.013.23, pp. 444-466.

Laidlaw, E.B., *Regulating Speech in Cyberspace: Gatekeepers, Human Rights and Corporate Responsibility*, Cambridge, Cambridge University Press, 2015, doi:10.1017/CBO9781107278721.

Lakier, G., 'The Non-First Amendment Law of Freedom of Speech', *Harvard Law Review*, Vol. 134, No. 7, 2021 (available at harvardlawreview.org/2021/05/the-non-first-amendment-law-of-freedom-of-speech), pp. 2299-.

Lakier, G., 'The Trump Lawsuits, the Biden Administration's Misinformation Advisory and the Thorny First Amendment Problem of Jawboning', *Lawfare,* 26 July 2021, available at lawfareblog.com/trump-lawsuits-biden-administrations-misinformation-advisory-and-thorny-first-amendment-problem (retrieved on 15 February 2022).

Land, M.K., 'Against Privatized Censorship: Proposals for Responsible Delegation', *Virginia Journal of International Law*, Vol. 60, No. 2, 2020, pp. 363-432.

Lawler, R. & A. Robertson, 'Biden signs executive order targeting right to repair, ISPs, net neutrality, and more', *The Verge,* 9 July 2021, available at theverge.com/2021/7/9/22569869/biden-executive-order-right-to-repair-isps-net-neutrality (retrieved on 15 February 2022).

Leiser, M.R., 'Regulating computational propaganda: lessons from international law', *Cambridge International Law Journal*, Vol. 8, No. 2, 2019, doi:10.4337/cilj.2019.02.03, pp. 218–240.

Lerman, R., 'Social media liability law is likely to be reviewed under Biden', *The Washington Post,* 18 January 2021, available at washingtonpost.com/politics/2021/01/18/biden-section-230 (retrieved on 28 March 2022).

Lessig, L., *Code Version 2.0*, New York, Basic Books, 2006.

Lexico, 'Meaning of chilling effect in English', *Lexico*, available at lexico.com/definition/chilling_effect (retrieved on 15 February 2022).

Lexico, 'Meaning of curate in English', *Lexico*, available at lexico.com/definition/curate (retrieved on 15 February 2022).

Lexico, 'Meaning of intermediary in English', *Lexico*, available at lexico.com/definition/intermediary (retrieved on 15 February 2022).

Lexico, 'Meaning of intermediate in English', *Lexico*, available at lexico.com/definition/intermediate (retrieved on 15 February 2022).

Lexico, 'Meaning of moderation in English', *Lexico*, available at lexico.com/definition/moderation (retrieved on 15 February 2022).

Lexico, 'Meaning of regulate in English', *Lexico*, available at lexico.com/definition/regulate (retrieved on 15 February 2022).

Linkedin, 'Professional community policies', *Linkedin*, available at linkedin.com/legal/professional-community-policies (retrieved on 15 February 2022).

Llansó, E.J., 'No amount of "AI" in content moderation will solve filtering's prior-restraint problem', *Big Data & Society*, Vol. 7, No. 1, 2020, doi:10.1177/2053951720920686, pp. 1-6.

Lowrey, T., 'Social media companies could be forced to give out names and contact details, under new anti-troll laws', *ABC News,* 28 November 2021, available at abc.net.au/news/2021-11-28/social-media-laws-online-trolls/100657004 (retrieved on 15 February 2022).

MacKinnon, R., *Consent of the Networked: The Worldwide Struggle For Internet Freedom*, New York, Basic Books, 2013 [2012].

Mak, V., *Legal Pluralism in European Contract Law*, Oxford, Oxford University Press, 2020, doi:10.1093/oso/9780198854487.001.0001.

Malcolm, J., 'As Threats to Korean and European Web Hosts Rise, the Manila Principles Go on Tour', *Electronic Frontier Foundation,* 10 July 2015, available at eff.org/deeplinks/2015/07/threats-korean-and-european-web-hosts-rise-manila-principles-go-tour (retrieved on 15 February 2022).

Manila Principles on Intermediary Liability, 'Manila Principles on Intermediary Liability: Best Practices Guidelines for Limiting Intermediary Liability for Content to Promote Freedom of Expression and Innovation', *Manila Principles on Intermediary Liability,* 24 March 2015, available at eff.org/files/2015/10/31/manila_principles_1.0.pdf (retrieved on 15 February 2022).

Masnick, M., 'Huge Loss For Free Speech In Europe: Human Rights Court Says Sites Liable For User Comments', *techdirt,* 16 June 2015, available at techdirt.com/articles/20150616/11252831361/huge-loss-free-speech-europe-human-rights-court-says-sites-liable-user-comments.shtml (retrieved on 15 February 2022).

Masnick, M., 'Protocols, Not Platforms: A Technological Approach to Free Speech', *Knight Columbia,* 21 August 2019, available at knightcolumbia.org/content/protocols-not-platforms-a-technological-approach-to-free-speech (retrieved on 15 February 2022).

Masnick, M., 'Hello! You've Been Referred Here Because You're Wrong About Section 230 Of The Communications Decency Act', *techdirt,* 23 June 2020, available at techdirt.com/2020/06/23/hello-youve-been-referred-here-because-youre-wrong-about-section-230-communications-decency-act (retrieved on 11 July 2022).

Masnick, M., 'Florida Man Governor Wastes More Florida Taxpayer Money Appealing Ruling About His Unconstitutional Social Media Law', *techdirt,* 13 July 2021, available at techdirt.com/articles/20210713/09513247161/florida-man-governor-wastes-more-florida-taxpayer-money-appealing-ruling-about-his-unconstitutional-social-media-law.shtml (retrieved on 15 February 2022).

Mcgill, M.H. & D. Lippman, 'White House Drafting Executive Order to Tackle Silicon Valley's Alleged Anti-Conservative Bias', *Politico,* 7 August 2019, available at politico.com/story/2019/08/07/white-house-tech-censorship-1639051 (retrieved on 15 February 2022).

McGonagle, T., 'Free Expression and Internet Intermediaries: The Changing Geometry of European Regulation', *in* G. Frosio (Ed.) *Oxford Handbook of Online Intermediary Liability*, Oxford, Oxford University Press, 2020, doi:10.1093/oxfordhb/9780198837138.013.24, pp. 467-485.

Mchangama, J., N. Alkiviadou & R. Mendiratta, 'Rushing to Judgment: Are Short Mandatory Takedown Limits for Online Hate Speech Compatible with The Freedom of Expression?', *The Future of Free Speech Project,*January 2021, available at futurefreespeech.com/rushing-to-judgment-are-short-mandatory-takedown-limits-for-online-hate-speech-compatible-with-the-freedom-of-expression (retrieved on 15 February 2022).

Meta, 'Mark Zuckerberg Stands for Voice and Free Expression', *Meta Newsroom,* 17 October 2019, available at about.fb.com/news/2019/10/mark-zuckerberg-stands-for-voice-and-free-expression (retrieved on 14 February 2022).

Meta, 'Adult Nudity and Sexual Activity', *Facebook Transparency Center,* 23 December 2021, available at facebook.com/communitystandards/adult_nudity_sexual_activity (retrieved on 14 February 2022).

Meta, 'Hate Speech', *Facebook Transparency Center,* 24 November 2021, available at facebook.com/communitystandards/hate_speech (retrieved on 14 February 2022).

Microsoft, 'PhotoDNA', *Microsoft*, available at microsoft.com/en-us/PhotoDNA (retrieved on 15 February 2022).

Montagnani, M.L., 'A New Liability Regime for Illegal Content in the Digital Single Market Strategy', *in* G. Frosio (Ed.) *Oxford Handbook of Online Intermediary Liability*, Oxford, Oxford University Press, 2020, doi:10.1093/oxfordhb/9780198837138.013.15, pp. 295-311.

Morozov, E., *The Net Delusion: How Not to Liberate The World*, London, Penguin Books, 2012.

Morozov, E., *To Save Everything, Click Here*, London, Penguin, 2014.

Morrison, S., 'Florida's social media free speech law has been blocked for likely violating free speech laws', *Vox Recode,* 1 July 2021, available at vox.com/recode/2021/7/1/22558980/florida-social-media-law-injunction-desantis (retrieved on 15 February 2022).

Mueller, M., *Will the Internet Fragment?*, Cambridge, Polity Press, 2017.

Musiani, F., '*Infrastructuring* digital sovereignty: a research agenda for an infrastructure-based sociology of digital self-determination practices', *Information, Communication & Society*, Vol. 25, No. 6, 2022, doi:10.1080/1369118X.2022.2049850, pp. 785-800.

Nguyen, T. & M. Scott, "Hashtags come to life': How online extremists fueled Wednesday's Capitol Hill insurrection', *Politico,* 8 January 2021, available at politico.com/news/2021/01/07/right-wing-extremism-capitol-hill-insurrection-456184 (retrieved on 8 January 2021).

Nylen, L., B.W. Swan & C. Lima, 'DOJ proposes crackdown on tech industry's legal shield', *Politico,* 17 June 2020, available at politico.com/news/2020/06/17/doj-crackdown-tech-industry-legal-shield-325594 (retrieved on 15 February 2022).

OECD, 'Participative Web and User-Created Content', *OECD,* 2007, available at oecd-ilibrary.org/science-and-technology/participative-web-and-user-created-content_9789264037472-en, doi:10.1787/9789264037472-en.

Omer, C., 'Intermediary Liability for Harmful Speech: Lessons from Abroad', *Harvard Journal of Law & Technology*, Vol. 28, No. 1, 2014, pp. 289-324.

Ong, T., 'Neo-nazi site Daily Stormer threatened by hosting providers and possible hackers', *The Verge,* 14 August 2017, available at theverge.com/2017/8/14/16142384/daily-stormer-site-go-daddy-hosting-providers-hackers-anonymous (retrieved on 15 February 2022).

Oster, J., *Media Freedom as a Fundamental Right*, Cambridge, Cambridge University Press 2015, doi:10.1017/CBO9781316162736.

Oversight Board, 'Case decision 2021-001-FB-FBR', *Oversight Board,* 5 May 2021, available at oversightboard.com/decision/FB-691QAMHJ (retrieved on 15 February 2022).

Oversight Board, 'Case decision 2021-002-FB-UA', *Oversight Board,* 13 April 2021, available at oversightboard.com/decision/FB-S6NRTDAJ (retrieved on 15 February 2022).

Oversight Board, 'Case decision 2021-007-FB-UA', *Oversight Board,* 11 August 2021, available at oversightboard.com/decision/FB-ZWQUPZLZ (retrieved on 15 February 2022).

Oxford Reference, 'Blackstone ratio', *Oxford Reference*, available at oxfordreference.com/view/10.1093/oi/authority.20110803095510389 (retrieved on 23 March 2022).

Parler, 'Community Guidelines', *Parler,* 2 November 2021, available at parler.com/documents/guidelines.pdf (retrieved on 15 February 2022).

Pasquale, F., 'Platform Neutrality: Enhancing Freedom of Expression in Spheres of Private Power', *Theoretical Inquiries in Law*, Vol. 17, No. 2, 2016, doi:10.1515/til-2016-0018, pp. 487-514.

Passchier, R., *Artificiële intelligentie en de rechtsstaat: Over verschuivende overheidsmacht, Big Tech en de noodzaak van constitutioneel onderhoud*, Den Haag, Boom juridisch, 2021.

Perry, R. & T.Z. Zarsky, 'Who Should Be Liable for Online Anonymous Defamation?', *University of Chicago Law Review Dialogue*, Vol. 82, 2015, pp. 162-176.

Perset, K., 'The Economic and Social Role of Internet Intermediaries', *OECD Digital Economy Papers* No. 117, Paris, OECD, 2010, doi:10.1787/20716826.

Polański, P.P., 'Rethinking the notion of hosting in the aftermath of Delfi: Shifting from liability to responsibility?', *Computer Law & Security Review*, Vol. 34, No. 4, 2018, doi:10.1016/j.clsr.2018.05.034, pp. 870-880.

Pollet, M., 'French senator calls for creation of digital identity supervisory body', *Euractiv,* 21 October 2021, available at euractiv.com/section/digital/news/french-senator-calls-for-creation-of-digital-identity-supervisory-body (retrieved on 15 February 2022).

Pollicino, O. & E. Bietti, 'Truth and Deception across the Atlantic: A Roadmap of Disinformation in the US and Europe', *Italian Journal of Public Law*, Vol. 11, No. 1, 2019, pp. 43-85.

Prime Minister of Australia, 'Combatting online trolls and strengthening defamation laws', *Prime Minister of Australia,* 28 November 2021, available at pm.gov.au/media/combatting-online-trolls-and-strengthening-defamation-laws (retrieved on 15 February 2022).

Riordan, J., *The Liability of Internet Intermediaries*, Oxford, Oxford University Press, 2016.

Riordan, J., 'A Theoretical Taxonomy of Intermediary Liability', *in* G. Frosio (Ed.) *Oxford Handbook of Online Intermediary Liability*, Oxford, Oxford University Press, 2020, doi:10.1093/oxfordhb/9780198837138.013.3, pp. 57-89.

Romano, A., 'A new law intended to curb sex trafficking threatens the future of the internet as we know it', *Vox,* 2 July 2018, available at vox.com/culture/2018/4/13/17172762/fosta-sesta-backpage-230-internet-freedom (retrieved on 15 February 2022).

Roth, E., 'Australian PM proposes defamation laws forcing social platforms to unmask trolls', *The Verge,* 28 November 2021, available at theverge.com/2021/11/28/22806369/australia-proposes-defamation-laws-unmask-trolls (retrieved on 15 February 2022).

RTL Nieuws, 'Facebook verwijdert groepen met Zwarte Piet, oprichters zoeken alternatief', *RTL Nieuws,* 4 September 2020, available at rtlnieuws.nl/nieuws/nederland/artikel/5181622/facebook-zwarte-piet-blackface-beheerders-racisme (retrieved on 15 February 2022).

Ruane, K.A., 'Net Neutrality: Selected Legal Issues Raised by the FCC's 2015 Open Internet Order', *in* D. Lambert (Ed.) *Net Neutrality and the FCC: Legal Issues and Matters of Debate*, New York, Nova Science Publishers, 2015, pp. 1-47.

Rushe, D. & Associated Press, 'Mark Zuckerberg: advertisers' boycott of Facebook will end 'soon enough'", *The Guardian,* 2 July 2020, available at theguardian.com/technology/2020/jul/02/mark-zuckerberg-advertisers-boycott-facebook-back-soon-enough (retrieved on 15 February 2022).

Sabel, P., 'Toezichtsraad: Zwarte Piet is raciaal stereotype en wordt terecht geweerd van Facebook en Instagram ', *De Volkskrant,* 13 April 2021, available at volkskrant.nl/nieuws-achtergrond/toezichtsraad-zwarte-piet-is-raciaal-stereotype-en-wordt-terecht-geweerd-van-facebook-en-instagram~b03f7cee (retrieved on 15 February 2022).

Sander, B., 'Democratic Disruption in the Age of Social Media: Between Marketized and Structural Conceptions of Human Rights Law', *European Journal of International Law*, 2021, doi:10.1093/ejil/chab022.

Sanderson, B., *The Way of Kings*, New York, Tor Books, 2010.

Santora, M., 'Turkey Passes Law Extending Sweeping Powers Over Social Media', *The New York Times,* 29 July 2020, available at nytimes.com/2020/07/29/world/europe/turkey-social-media-control.html (retrieved on 15 February 2022).

Schauer, F., 'The Exceptional First Amendment', *in* M. Ignatieff (Ed.) *American Exceptionalism and Human Rights*, Princeton, Princeton University Press 2005, doi:10.1515/9781400826889.29, pp. 29-56.

Senden, L., *Soft Law in European Community Law*, Portland, Hart Publishing, 2004.

Smith, M.D. & M. Van Alstyne, 'It's Time to Update Section 230', *Harvard Business Review,* 12 August 2021, available at hbr.org/2021/08/its-time-to-update-section-230 (retrieved on 15 February 2022).

Solum, L.B., 'Legal Theory Lexicon: Path Dependency', *Legal Theory Blog,* 2 September 2018, available at lsolum.typepad.com/legaltheory/2018/09/legal-theory-lexicon-path-dependency.html (retrieved on 15 February 2022).

Spano, R., 'Intermediary Liability for Online User Comments under the European Convention on Human Rights', *Human Rights Law Review*, Vol. 17, No. 4, 2017, doi:10.1093/hrlr/ngx001, pp. 665–679.

Speta, J.B., 'A Common Carrier Approach to Internet Interconnection', *Federal Communications Law Journal*, Vol. 54, No. 2, 2002, pp. 225-280.

Stacey, K. & R. Waters, 'What can Silicon Valley expect from Joe Biden?', *Financial Times,* 8 November 2020, available at ft.com/content/44f738e8-eb6f-4394-b833-6b3207ce31bf (retrieved on 20 May 2021).

Stjernfelt, F. & A.M. Lauritzen, *Your Post has been Removed: Tech Giants and Freedom of Speech*, Cham, Springer, 2019, doi:10.1007/978-3-030-25968-6.

Sunstein, C.R., 'Falsehoods and the First Amendment', *Harvard Journal of Law & Technology*, Vol. 33, No. 2, 2020, pp. 387-426.

Susser, D., B. Roessler & H. Nissenbaum, 'Technology, autonomy, and manipulation', *Internet Policy Review*, Vol. 8, No. 2, 2019, doi:10.14763/2019.2.1410.

Svantesson, D.J.B., 'Internet Jurisdiction and Intermediary Liability', *in* G. Frosio (Ed.) *The Oxford Handbook of Online Intermediary Liability*, Oxford, Oxford University Press, 2020, doi:10.1093/oxfordhb/9780198837138.013.3, pp. 691-708.

Švedkauskas, Ž., C. Sirikupt & M. Salzer, 'Russia's disinformation campaigns are targeting African Americans', *The Washington Post,* 24 July 2020, available at washingtonpost.com/politics/2020/07/24/russias-disinformation-campaigns-are-targeting-african-americans/ (retrieved on 15 February 2022).

Sweney, M., 'Squid Game's success reopens who pays debate over rising internet traffic', *The Guardian,* 10 October 2021, available at theguardian.com/business/2021/oct/10/squid-games-success-reopens-debate-over-who-should-pay-for-rising-internet-traffic-netflix (retrieved on 15 February 2022).

The Editorial Board of the New York Times, 'Joe Biden: Former vice president of the United States', *The New York Times,* 17 January 2020, available at nytimes.com/interactive/2020/01/17/opinion/joe-biden-nytimes-interview.html (retrieved on 15 February 2022).

The Santa Clara Principles, 'Santa Clara Principles 1.0', *The Santa Clara Principles on Transparency and Accountability in Content Moderation,* 7 May 2018, available at santaclaraprinciples.org/scp1/ (retrieved on 15 February 2022).

The Seventy-Third World Health Assembly, 'Resolution WHA73.1: COVID-19 response', *World Health Organization,* 19 May 2020, available at apps.who.int/gb/ebwha/pdf_files/WHA73/A73_R1-en.pdf (retrieved on 15 February 2022).

The White House, 'Executive Order on Preventing Online Censorship', *The White House,* 28 May 2020, available at trumpwhitehouse.archives.gov/presidential-actions/executive-order-preventing-online-censorship (retrieved on 15 February 2022).

The White House, 'Executive Order on the Revocation of Certain Presidential Actions and Technical Amendment', *The White House,* 14 May 2021, available at whitehouse.gov/briefing-room/presidential-actions/2021/05/14/executive-order-on-the-revocation-of-certain-presidential-actions-and-technical-amendment (retrieved on 15 February 2022).

Thielman, S., 'Silk Road operator Ross Ulbricht sentenced to life in prison', *The Guardian,* 29 May 2015, available at theguardian.com/technology/2015/may/29/silk-road-ross-ulbricht-sentenced (retrieved on 15 February 2022).

Transatlantic High Level Working Group on Content Moderation Online and Freedom of Expression, 'Freedom and Accountability: A Transatlantic Framework for Moderating Speech Online', *The Annenberg Public Policy Center of the University of Pennsylvania,* 2020, available at annenbergpublicpolicycenter.org/feature/transatlantic-working-group-freedom-and-accountability.

Trendacosta, K., 'Busting Two Myths About Paid Prioritization', *Electronic Frontier Foundation,* 16 April 2018, available at eff.org/deeplinks/2018/04/busting-two-myths-about-paid-prioritization (retrieved on 15 February 2022).

Tsesis, A., 'Terrorist Speech on Social Media', *Vanderbilt Law Review*, Vol. 70, No. 2, 2017 (available at scholarship.law.vanderbilt.edu/vlr/vol70/iss2/4), pp. 651-708.

Twitter, 'COVID-19 misleading information policy', *Twitter Help Center*, available at help.twitter.com/en/rules-and-policies/medical-misinformation-policy (retrieved on 15 February 2022).

Twitter, 'Hateful conduct policy', *Twitter Help Center*, available at help.twitter.com/en/rules-and-policies/hateful-conduct-policy (retrieved on 15 February 2022).

Twitter, 'Twitter API', *Twitter Developer Platform*, available at developer.twitter.com/en/docs/twitter-api (retrieved on 15 February 2022).

u/Reddit-Policy, 'New addition to site-wide rules regarding the use of Reddit to conduct transactions', *reddit.com/r/announcements,* 21 March 2018, available at reddit.com/r/announcements/comments/863xcj/new_addition_to_sitewide_rules_regarding_the_use (retrieved on 15 February 2022).

US Department of Justice, 'Department of Justice's Review of Section 230 of the Communications Decency Act of 1996', *US Department of Justice,* 2020, available at justice.gov/archives/ag/department-justice-s-review-section-230-communications-decency-act-1996 (retrieved on 15 February 2022).

US Department of Justice, 'Department of Justice's Review of Section 230 of the Communications Decency Act of 1996 (Cover Letter)', *US Department of Justice,* 23 September 2020, available at justice.gov/file/1319346/download (retrieved on 15 February 2022).

US Department of Justice, 'Department of Justice's Review of Section 230 of the Communications Decency Act of 1996 (Redline)', *US Department of Justice,* 23 September 2020, available at justice.gov/file/1319331/download (retrieved on 15 February 2022).

Vermeulen, B.P., 'Artikel 7 - Vrijheid van meningsuiting', *NederlandRechtsstaat*, available at nederlandrechtsstaat.nl/grondwet/inleiding-bij-hoofdstuk-1-grondrechten/artikel-7-grondwet-vrijheid-van-meningsuiting (retrieved on 15 February 2022).

Volokh, E. & D.M. Falk, 'Google: First Amendment Protection for Search Engine Search Results', *Journal of Law, Economics & Policy*, Vol. 8, No. 4, 2012, pp. 883-900.

Voorhoof, D., 'De aansprakelijkheid van online nieuwsplatforms na Delfi', *MediaForum*, No. 6, 2015, pp. 194-204.

Wakabayashi, D., 'Legal Shield for Social Media Is Targeted by Lawmakers', *The New York Times,* 15 December 2020, available at nytimes.com/2020/05/28/business/section-230-internet-speech.html (retrieved on 2 February 2022).

Wex, 'Incentive', *Legal Information Institute*, available at law.cornell.edu/wex/incentive (retrieved on 18 January 2022).

Whittaker, J., et al., 'Recommender systems and the amplification of extremist content', *Internet Policy Review*, Vol. 10, No. 2, 2021, doi:10.14763/2021.2.1565

Wikipedia, 'OSI model', *Wikipedia,* 15 February 2022, available at en.wikipedia.org/wiki/OSI_model (retrieved on 15 February 2022).

Wilman, F., *The Responsibility of Online Intermediaries for Illegal User Content in the EU and the US*, Cheltenham, Edward Elgar Publishing, 2020, doi:10.4337/9781839104831.

Wilman, F., 'Het voorstel voor de Digital Services Act: Op zoek naar nieuw evenwicht in regulering van onlinediensten met betrekking tot informatie van gebruikers', *Nederlands tijdschrift voor Europees recht*, No. 1-2, 2021, doi:10.5553/NtER/138241202021027102002, pp. 27-36.

World Health Organization, 'Managing the COVID-19 infodemic: Promoting healthy behaviours and mitigating the harm from misinformation and disinformation', *World Health Organization,* 23 September 2020, available at who.int/news-room/detail/23-09-2020-managing-the-covid-19-infodemic-promoting-healthy-behaviours-and-mitigating-the-harm-from-misinformation-and-disinformation (retrieved on 15 February 2022).

Wu, F.T., 'Collateral Censorship and the Limits of Intermediary Immunity', *Notre Dame Law Review*, Vol. 87, No. 1, 2011, pp. 293-350.

Wu, T., 'Network Neutrality, Broadband Discrimination', *Journal on Telecommunications & High Technology Law*, Vol. 2, 2003, pp. 141-176.

Wu, T., *The Master Switch: The Rise and Fall of Information Empires*, New York, Vintage Books, 2011.

Wu, T., *The Curse of Bigness: Antitrust in the New Gilded Age*, New York, Columbia Global Reports, 2018.

Yemini, M., 'The New Irony of Free Speech', *Columbia Science and Technology Law Review*, Vol. 201, No. 1, 2018, pp. 119-194.

Zittrain, J., 'Internet Points of Control', *Boston College Law Review*, Vol. 44, No. 2, 2003, pp. 653-688.

Zuboff, S., *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, London, Profile Books, 2019.

Treaties

European Convention on Human Rights.

International Covenant on Civil and Political Rights, 16 December 1966, 999 U.N.T.S. 171.

United Nations

Office of the United Nations High Commissioner for Human Rights, *Guiding Principles on Business and Human Rights: Implementing the United Nations "Protect, Respect and Remedy" Framework*, UN Doc. HR/PUB/11/04 (2011).

European Union

Charter of Fundamental Rights of the European Union, *OJ C 326*, *26.10.2012* (data.europa.eu/eli/treaty/char_2012/oj), pp. 391-407.

Proposal COM(1998) 586 final of 23 December 1998 for a European Parliament and Council Directive on certain legal aspects of electronic commerce in the internal market, *OJ C 30*, *5.2.1999*, pp. 4-16.

Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market (*Directive on electronic commerce*), *OJ L 178*, *17.7.2000* (data.europa.eu/eli/dir/2000/31/oj), pp. 1-16.

Directive 2005/29/EC of the European Parliament and of the Council of 11 May 2005 concerning unfair business-to-consumer commercial practices in the internal market and amending Council Directive 84/450/EEC, Directives 97/7/EC, 98/27/EC and 2002/65/EC of the European Parliament and of the Council and Regulation (EC) No 2006/2004 of the European Parliament and of the Council (*Unfair Commercial Practices Directive*), *OJ L 149*, *11.6.2005* (data.europa.eu/eli/dir/2005/29/oj), pp. 22-39.

Council Framework Decision 2008/913/JHA of 28 November 2008 on combating certain forms and expressions of racism and xenophobia by means of criminal law, *OJ L 328*, *6.12.2008* (data.europa.eu/eli/dec_framw/2008/913/oj), pp. 55–58.

Directive 2010/13/EU of the European Parliament and of the Council of 10 March 2010 on the coordination of certain provisions laid down by law, regulation or administrative action in Member States concerning the provision of audiovisual media services (*Audiovisual Media Services Directive*), *OJ L 95*, *15.4.2010* (data.europa.eu/eli/dir/2010/13/oj), pp. 1-24.

Directive (EU) 2015/1535 of the European Parliament and of the Council of 9 September 2015 laying down a procedure for the provision of information in the field of technical regulations and of rules on Information Society services, *OJ L 241*, *17.9.2015* (data.europa.eu/eli/dir/2015/1535/oj), pp. 1-15.

Regulation (EU) 2015/2120 of the European Parliament and of the Council of 25 November 2015 laying down measures concerning open internet access and amending Directive 2002/22/EC on universal service and users' rights relating to electronic communications networks and services and Regulation (EU) No 531/2012 on roaming on public mobile communications networks within the Union, *OJ L 310*, *26.10.2015* (data.europa.eu/eli/reg/2015/2120/oj), pp. 1-18.

European Parliament resolution of 13 December 2016 on the situation of fundamental rights in the European Union in 2015 (2016/2009(INI)), *OJ C 238*, *6.7.2018*, pp. 2-27.

Communication COM(2017)555 final of the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions of 28 September 2017 Tackling Illegal Content Online Towards an enhanced responsibility of online platforms.

Directive (EU) 2018/1808 of the European Parliament and of the Council of 14 November 2018 amending Directive 2010/13/EU on the coordination of certain provisions laid down by law, regulation or administrative action in Member States concerning the provision of audiovisual media services (Audiovisual Media Services Directive) in view of changing market realities, *OJ L 303*, *28.11.2018* (data.europa.eu/eli/dir/2018/1808/oj), pp. 69-92.

Commission Recommendation (EU) 2018/334 of 1 March 2018 on measures to effectively tackle illegal content online, *OJ L 63*, *6.3.2018* (data.europa.eu/eli/reco/2018/334/oj), pp. 50-61.

Commission Proposal COM(2018) 640 final of 12 September 2018 Regulation of the European Parliament and of the Council on preventing the dissemination of terrorist content online.

Communication COM(2018)236 final of the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions of 26 April 2018 Tackling Online Disinformation: A European Approach.

Directive (EU) 2019/2161 of the European Parliament and of the Council of 27 November 2019 amending Council Directive 93/13/EEC and Directives 98/6/EC, 2005/29/EC and 2011/83/EU of the European Parliament and of the Council as regards the better enforcement and modernisation of Union consumer protection rules, *OJ L 328*, *18.12.2019* (data.europa.eu/eli/dir/2019/2161/oj), pp. 7-28.

Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC, *OJ L 130*, *17.5.2019* (data.europa.eu/eli/dir/2019/790/oj), pp. 92-125.

Commission Proposal COM(2020) 825 final of 15 December 2020 Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC (*Digital Services Act*).

Joint Communication JOIN(2020) 8 final of the European Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions of 6 June 2020 Tackling COVID-19 disinformation - Getting the facts right.

Communication COM(2021) 262 final of the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions of 26 May 2021 European Commission Guidance on Strengthening the Code of Practice on Disinformation.

Regulation (EU) 2021/784 of the European Parliament and of the Council of 29 April 2021 on addressing the dissemination of terrorist content online, *OJ L 172*, *17.5.2021* (data.europa.eu/eli/reg/2021/784/oj), pp. 79-109.

European Commission, 'The EU Code of conduct on countering illegal hate speech online', *European Commission*, available at ec.europa.eu/info/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/eu-code-conduct-countering-illegal-hate-speech-online_en (retrieved on 14 February 2022).

European Commission, 'Code of Conduct on Countering Illegal Hate Speech Online', *European Commission,* 30 June 2016, available at ec.europa.eu/info/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/eu-code-conduct-countering-illegal-hate-speech-online_en (retrieved on 14 February 2022).

European Commission, 'European Commission and IT Companies announce Code of Conduct on illegal online hate speech', *European Commission,* 31 May 2016, available at ec.europa.eu/commission/presscorner/detail/en/IP_16_1937 (retrieved on 14 February 2022).

European Commission, 'Disinformation: EU assesses the Code of Practice and publishes platform reports on coronavirus related disinformation', *European Commission,* 10 September 2020, available at ec.europa.eu/commission/presscorner/detail/en/ip_20_1568 (retrieved on 14 February 2022).

European Commission, 'Europe fit for the Digital Age: Commission proposes new rules for digital platforms', *European Commission,* 15 December 2020, available at ec.europa.eu/commission/presscorner/detail/en/ip_20_2347 (retrieved on 14 February 2022).

European Commission, 'Code of Practice on Disinformation', *European Commission,* 2 December 2021, available at digital-strategy.ec.europa.eu/en/policies/code-practice-disinformation (retrieved on 14 February 2022).

Legislation - Germany

Network Enforcement Act 2017 (*Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken*)

Legislation - The Netherlands

Wetboek van Strafrecht (Dutch Criminal Code).

Legislation - United States of America

H.R. Rep. No. 115–572, pt. 1 (2018).

Allow States and Victims to Fight Online Sex Trafficking Act of 2017 (FOSTA-SESTA), H.R. 1865, 115th Cong. (2018 through PL 115-164).

106.072. Social media deplatforming of political candidates, Fla. Stat. Ann § 106.072 (West 2021, Westlaw Next).

501.2041. Unlawful acts and practices by social media platforms, Fla. Stat. Ann § 501.2041 (West 2021, Westlaw Next).

§512. Limitations on liability relating to material online, 17 USCA § 512 (West 2010, Westlaw Next through PL 116-179).

§4102. Recognition of foreign defamation judgments, 28 USCA § 4102 (West 2010, Westlaw Next through PL 116-150).

§230. Protection for private blocking and screening of offensive material, 47 USCA § 230 (West 2018, Westlaw Next through PL 116-91).

§1591. Sex trafficking of children or by force, fraud, or coercion, 18 USCA § 1591 (West 2018, Westlaw Next through PL 116-193).

§1593. Mandatory restitution, 18 USCA § 1593 (West 2018, Westlaw Next through PL 116-193).

§2421A. Promotion or facilitation of prostitution and reckless disregard of sex trafficking, 18 USCA § 2421A (West 2018, Westlaw Next through PL 116-193).

A bill to amend the Communications Act of 1934 to provide that, under certain circumstances, an interactive computer service provider that allows for the proliferation of health misinformation through that service shall be treated as the publisher or speaker of that misinformation, and for other purposes (Health Misinformation Act of 2021), S. 2448, 117th Cong. (2021).

2021 Fla. Sess. Law Serv. Ch. 2021-32 (SB 7072) (West).

European Court of Human Rights

Appleby and Others v. the United Kingdom, no. 44306/98, ECHR 2003-VI, 6 May 2003, ECLI:CE:ECHR:2003:0506JUD004430698.

Lindon, Otchakovsky-Laurens and July v. France [GC], no. 21279/02, 36448/02, ECHR 2007-IV, 22 October 2007, ECLI:CE:ECHR:2007:1022JUD002127902.

Ahmet Yıldırım v. Turkey, no. 3111/10, ECHR 2012-VI, 18 December 2012, ECLI:CE:ECHR:2012:1218JUD000311110.

Delfi AS v. Estonia, no. 64569/09, 10 October 2013, ECLI:CE:ECHR:2013:1010JUD006456909.

Cengiz and Others v. Turkey, no. 48226/10 and 14027/11, ECHR 2015-VIII, 1 December 2015, ECLI:CE:ECHR:2015:1201JUD004822610.

Delfi AS v. Estonia [GC], no. 64569/09, ECHR 2015-II, 16 June 2015, ECLI:CE:ECHR:2015:0616JUD006456909.

Magyar Tartalomszolgáltatók Egyesülete and Index.hu Zrt v. Hungary, no. 22947/13, 2 February 2016, ECLI:CE:ECHR:2016:0202JUD002294713.

Kablis v. Russia, no. 48310/16 and 59663/17, 30 April 2019, ECLI:CE:ECHR:2019:0430JUD004831016.

Engels v. Russia, no. 61919/16, 23 June 2020, ECLI:CE:ECHR:2020:0623JUD006191916.

OOO Flavus and Others v. Russia, no. 12468/15, 23489/15 and 19074/16, 23 June 2020, ECLI:CE:ECHR:2020:0623JUD001246815.

Vladimir Kharitonov v. Russia, no. 10795/14, 23 June 2020, ECLI:CE:ECHR:2020:0623JUD001079514.

European Court of Justice

Judgment of the Court (Fourth Chamber) of 27 March 2014 in *Case C-314/12, UPC Telekabel Wien GmbH v Constantin Film Verleih GmbH and Wega Filmproduktionsgesellschaft mbH*, ECLI:EU:C:2014:192.

Judgment of the Court (Grand Chamber) of 13 May 2014 in *Case C-131/12, Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González*, ECLI:EU:C:2014:317.

Judgement of the Court (Third Chamber) of 3 October 2019 in *Case C-18/18, Glawischnig-Piesczek*, ECLI:EU:C:2019:821.

Judgement of the Court (Grand Chamber) of 23 March 2010 in *Case C-236/08, C-237/08 and C-238/08, Google France SARL and Google Inc. v. Louis Vuitton Malletier SA*, ECLI:EU:C:2010:159.

Judgement of the Court (Grand Chamber) of 12 July 2011 in *Case C-324/09, L'Oréal SA and Others v. eBay International AG and Others*, ECLI:EU:C:2011:474.

Judgment of the Court (Seventh Chamber) of 11 September 2014 in *Case C-291/13, Sotiris Papasavvas v. O Fileleftheros Dimosia Etaireia Ltd and Others*, ECLI:EU:C:2014:2209.

Judgment of the Court (Second Chamber) of 14 June 2017 in *Case C-610/15, Stichting Brein v Ziggo BV and XS4All Internet BV*, ECLI:EU:C:2017:456.

Judgment of the Court (Grand Chamber) of 22 June 2021 in *Case C-682/18 and C-683/18, Frank Peterson v. Google LLC and Others and Elsevier Inc. v Cyando AG*, ECLI:EU:C:2021:503.

Opinion of Advocate General Poiares Maduro of 22 September 2009 in *Case C-236/08, C-237/08 and C-238/08, Google France SARL and Google Inc. v. Louis Vuitton Malletier SA*, ECLI:EU:C:2009:569.

Opinion of Advocate General Saugmandsgaard Øe of 16 July 2020 in *Case C-682/18, C-683/18, Frank Peterson v. Google LLC, YouTube LLC, YouTube Inc., Google Germany GmbH (C-682/18) and Elsevier Inc. v. Cyando AG (C-683/18)*, ECLI:EU:C:2020:586.

Opinion of Advocate General Saugmandsgaard Øe of 15 July 2021 in *Case C-401/19, Poland v. Parliament and Council*, ECLI:EU:C:2021:613.

United States Case Law

Smith v. People of the State of California, 80 S.Ct. 215 (1959).

Globe Intern., Inc. v. Khawar, 119 S.Ct. 1760 (1999).

Manhattan Community Access Corporation v. Halleck, 139 S.Ct. 1921 (2019).

Jane Doe No. 1 v. Backpage.com, LLC, 817 F.3d 12 (1st Cir. 2016).

Zeran v. America Online, Inc., 129 F.3d 327 (3rd Cir. 1997).

Doe v. MySpace, Inc., 528 F.3d 413 (5th Cir. 2008).

Doe v. SexSearch.com, 551 F.3d 412 (6th Cir. 2008).

Jones v. Dirty World Entertainment Recordings LLC, 755 F.3d 398 (6th Cir. 2014).

Doe v. GTE Corp., 347 F.3d 655 (7th Cir. 2003).

Chicago Lawyers' Committee for Civil Rights Under Law, Inc. v. Craigslist, Inc., 519 F.3d 666 (7th Cir. 2008).

Batzel v. Smith, 333 F.3d 1018 (9th Cir. 2003).

Carafano v. Metrosplash.com, Inc., 339 F.3d 1119 (9th Cir. 2003).

Yahoo! Inc. v. La Ligue Contre Le Racisme et l'antisémitisme (LICRA), 433 F.3d 1199 (9th Cir. 2006).

Fair Housing Council of San Fernando Valley v. Roommates.com, LLC, 521 F.3d 1157 (9th Cir. 2008).

Barnes v. Yahoo!, Inc., 570 F.3d 1096 (9th Cir. 2009).

Fair Housing Council of San Fernando Valley v. Roommate.com, LLC, 666 F.3d 1216 (9th Cir. 2012).

Doe v. Internet Brands, Inc., 824 F.3d 846 (9th Cir. 2016).

Ben Ezra, Weinstein, and Company, Inc. v. America Online Inc., 206 F.3d 980 (10th Cir. 2000).

F.T.C. v. Accusearch Inc., 570 F.3d 1187 (10th Cir. 2009).

Colon v. Twitter, Inc., 14 F.4th 1213 (11th Cir. 2021).

Stratton Oakmont, Inc. v. Prodigy Services Co., 23 Media L Rep 1794 (N.Y. Sup. Ct. 1995).

Khawar v. Globe Intern., Inc., 965 P.2d 696 (Cal. S.C. 1998).

Barrett v. Rosenthal, 146 P.3d 510 (Cal. S.C. 2006).

Cubby, Inc. v. CompuServe, Inc., 776 F.Supp. 135 (S.D. New York 1991).

Blumenthal v. Drudge, 992 F.Supp. 44 (D.D.C. 1998).

Search King Inc. v. Google Technology, Inc., 2003 WL 21464568 (W.D.Okla. 2003).

Anthony v. Yahoo Inc., 421 F.Supp.2d 1257 (N.D.Cal. 2006).

Doe v. SexSearch.com, 502 F.Supp.2d 719 (N.D.Ohio 2007).

Dart v. Craigslist, Inc., 655 F.Supp.2d 961 (N.D.Ill. 2009).

Goddard v. Google, Inc., 640 F.Supp.2d 1193 (N.D.Cal. 2009).

Fields v. Twitter, Inc., 200 F.Supp.3d 964 (N.D.Cal. 2016).

Lancaster v. Alphabet Inc., 2016 WL 3648608 (N.D.Cal. 2016).

e-ventures Worldwide, LLC v. Google, Inc., 2017 WL 2210029 (M.D.Fla. 2017).

Mezey v. Twitter, Inc., 2018 WL 5306769 (S.D.Fla. 2018).

Ebeid v. Facebook, Inc., 2019 WL 2059662 (N.D.Cal. 2019).

Cohoon v. Konrath, 2021 WL 4356069 (E.D.Wis. 2021).

Doe v. Google LLC, 2021 WL 4864418 (N.D.Cal. 2021).

NetChoice, LLC v. Moody, 2021 WL 2690876 (N.D. Florida 2021).

Dutch Case Law

HR, 13 November 2015, ECLI:NL:HR:2015:3307, *Nederlandse Jurisprudentie* 2018/110, m.nt. P.B. Hugenholtz.

HR, 14 February 2017, ECLI:NL:HR:2017:220 (concl. P.C. Vegter), *Nederlandse Jurisprudentie* 2017/259, m.nt. E.J. Dommering.

Gerechtshof Amsterdam, 2 June 2020, ECLI:NL:GHAMS:2020:1421.

Rb. Amsterdam (vzr.), 9 September 2020, ECLI:NL:RBAMS:2020:4435, *Jurisprudentie geneesmiddelenrecht* 2020/36, m.nt. M.D.B. Schutjens.

Rb. Amsterdam (vzr.), 13 October 2020, ECLI:NL:RBAMS:2020:4966, *Computerrecht* 2021/66, m.nt. M. Klos (*Facebook*).

Rb. Amsterdam (vzr.), 9 September 2020, ECLI:NL:RBAMS:2020:4435 (*YouTube*).

Rb. Amsterdam (vzr.), 18 August 2021, ECLI:NL:RBAMS:2021:4308 (*BLCKBX/Google*).

Rb. Noord-Holland (vzr.), 6 October 2021, ECLI:NL:RBNHO:2021:8539 (*Kamerlid/LinkedIn*).

Rb. Midden-Nederland (vzr.), 8 October 2020, ECLI:NL:RBMNE:2020:4348, *Computerrecht* 2021/65, m.nt. M.G.A. Berk.

Summary

Wrongful moderation deals with the question of how internet intermediary service providers (providers) that offer a service for users to provide information of their own and receive information from other users are provided with a legal incentive to overregulate or underregulate user-provided information based on its illegal, unlawful or otherwise considered harmful content. Providers that offer functionalities for user-provided information are criticised for failing to counter illegal content and overregulating content that is not illegal but considered harmful by the provider. In the United States of America (US) and the European Union (EU), legislation is proposed to remedy such overregulation and underregulation. However, as argued in *Wrongful moderation,* both overregulation and underregulation may be tied to the liability regimes that are in place that regulate when the provider is exempted from liability for the content of user-provided information.

The central question of this dissertation is to what extent (proposals for) regulation applicable to providers laid down in the e-Commerce Directive in the EU and Section 230 in the US provide a legal incentive to overregulate or underregulate user-provided information. The focus lies on so-called hosting service providers that offer an online platform for user-provided information. These providers, as argued, can moderate and curate user-provided information based on its content and are therefore popular targets for (state) regulation which is called internet intermediary regulation. Content regulation may target content that can be considered illegal, unlawful, or harmful. Underregulation and overregulation that may arise from such regulation may raise freedom of expression concerns.

The first chapter discusses how internet providers are regulated differently from offline information intermediaries. This chapter sets out that regulation is (and must) be differentiated between offline information intermediaries and providers to prevent overregulation and underregulation based on the content of the information. Next to this, the argument is advanced that it is necessary to distinguish different types of providers. While providers that offer platform functionalities for user-provided information can impose content-based restrictions, this is not true for all providers. Providers that are pivotal to the internet infrastructure should not engage in content-based restrictions – especially not when these restrictions include removal. When state regulation targets this last category of providers to impose content regulation, there is a clear risk of overregulation.

The second chapter explores who can be targeted by content regulation, the potential instruments, the remedies, and how internet intermediary regulation can cause extraterritorial overregulation. In this chapter, the argument is made that the actor (the provider or the user), the instruments (how are rule violations uncovered) and the remedies all contribute to (potential) overregulation or underregulation. Making providers responsible for regulating user-provided information may cause (extraterritorial) overregulation when the provider is required to remove user-provided information that violates the law in one jurisdiction and also for users in a jurisdiction in which the content is not illegal. This chapter also deals with the distinction between curation and moderation. As far curation occurs as a sanction following a rule violation, this is considered similar to moderation.

The third and fourth chapters discuss the internet intermediary liability regimes in the US and the EU. In the US, providers are excluded from civil liability for moderation decisions taken in "good faith" and liability from the content of user-provided information as a publisher or

speaker. In the EU, providers are merely exempted from liability for user-provided information as the provider had no knowledge or awareness of its illegal content. Besides, active providers that can be said to have gained knowledge or control over user-provided information cannot count on safe harbours. This approach provides a legal incentive to refrain from voluntary moderation because this may amount to knowledge or control and thus potential liability. Removing or disabling access to user-provided information with illegal content can exempt the provider from liability when this is the case. New proposals for regulation leave the EU safe harbours largely unaffected.

The fifth chapter reviews these approaches. Based on the previous chapters, three factors contribute to overregulation or underregulation: ambiguity, the required instruments, and the required remedy. So far, overregulation seems a far more significant risk than underregulation. This chapter discusses that strict liability regimes hold the most risks for overregulation and underregulation. However, strict liability regimes are less common in the US and the EU. Conditional liability regimes (most common in the EU) know fewer risks for underregulation and a decreased risk for overregulation than strict liability regimes. However, conditional liability regimes based on knowledge-based liability (as the EU) may cause providers to refrain from voluntary moderation efforts. Immunity from legal liability is the riskiest concerning underregulation because providers cannot be held liable for illegal content on their services. While there is no risk of becoming liable, immunity regimes do not provide a legal incentive to regulate. Because immunity regimes do not require providers to deploy instruments or a specific set of sanctions, immunity regimes allow a more refined set of remedies that may follow a rule violation. This fourth possibility is characterised as non-liability regulation. While voluntary codes are not legally enforceable, they may be viewed as the first step toward regulation which may cause providers to overregulate to prevent such regulation.

The argument in *Wrongful moderation* is made that a conditional liability approach stands out in balancing between overregulation and underregulation and users' freedom of expression rights. The 'right' conditional liability approach may prevent wrongful moderation consisting of overremoval or underremoval. However, the conditional liability approach is also particularly vulnerable to ambiguous legislation. When the provider has to fulfil a set of unclear or complicated conditions, a conditional liability approach may lose its competitive value. Unclear conditions may lead to very similar effects as strict liability approaches. Because the provider is unsure whether the provider could rely on the exemptions from liability, the provider may engage in extensive overmoderation, cease (parts) of the service, or cease all moderation efforts to prevent gaining knowledge.

Summary in Dutch: *Onrechtmatige moderatie*

Wrongful moderation behandelt de vraag hoe aanbieders van hostingdiensten waarbij zij informatie van gebruikers opslaan, een juridische prikkel krijgen om inhoud verstrekt door gebruikers met mogelijke een illegale, onrechtmatige of schadelijke inhoud te overreguleren of onderreguleren. Providers worden vaak bekritiseerd omdat zij nalaten illegale inhoud tegen te gaan en inhoud die niet illegaal is maar door de aanbieder als schadelijk wordt beschouwd wel verwijderen. In de Verenigde Staten van Amerika (VS) en de Europese Unie (EU) is en wordt wetgeving voorgesteld om dergelijke overregulering en onderregulering tegen te gaan. Echter, zoals betoogd in *Wrongful moderation*, kunnen zowel overregulering als onderregulering veroorzaakt worden door juridische prikkels die voortkomen uit het aansprakelijkheidsregime die van toepassing is op de aanbieder. Dergelijke aansprakelijkheidsregimes voorzien meestal in de voorwaarden waaronder de aanbieder niet aansprakelijk is of kan worden gesteld voor de inhoud van informatie geplaatst door gebruikers.

De centrale vraag van dit proefschrift is in hoeverre (voorstellen voor) regulering van providers zoals neergelegd in de Richtlijn inzake elektronische handel van de EU en Sectie 230 in de VS een juridische prikkel geven tot over- of onderregulering van door gebruikers geplaatste informatie. De nadruk ligt op zogenaamde hosting service providers die een onlineplatform bieden voor gebruikers om informatie te plaatsen en te ontvangen van andere gebruikers. Deze aanbieders kunnen de inhoud van de door gebruikers geplaatste informatie modereren en cureren en zijn daarom populaire doelwitten voor (overheids)regulering. Deze regulering kan gericht zijn op inhoud die als illegaal, onwettig of schadelijk wordt beschouwd. Onder- en overregulering die uit dergelijke regulering kan voortvloeien kan het uitoefenen van het recht op vrijheid van meningsuiting door gebruikers negatief beïnvloeden.

In het eerste hoofdstuk wordt besproken hoe internettussenpersonen anders worden gereguleerd dan offline informatietussenpersonen. In dit hoofdstuk wordt uiteengezet dat de regulering gedifferentieerd is (en moet zijn) tussen offline informatietussenpersonen en internettussenpersonen om over- en onderregulering van informatie geplaatst door gebruikers te voorkomen. Daarnaast wordt besproken dat het noodzakelijk is om verschillende soorten internettussenpersonen te onderscheiden. Internettussenpersonen die platformfunctionaliteiten bieden voor door gebruikers geplaatste zijn het meest geschikt om de inhoud van deze informatie te reguleren. In contrast, internettussenpersonen die een spilfunctie vervullen in de internetinfrastructuur dienen zich te onthouden van dergelijke regulering – zeker wanneer deze regulering leidt tot het verwijderen van gebruikersinhoud. Wanneer overheidsregulering zich op deze laatste categorie van aanbieders richt om inhoudelijke beperkingen op te leggen, bestaat er een duidelijk risico van overregulering.

Het tweede hoofdstuk onderzoekt wie het doelwit van inhoudsregulering kan zijn, de mogelijke instrumenten, de remedies, en hoe regulering van tussenpersonen op het internet extraterritoriale overregulering kan veroorzaken. In dit hoofdstuk wordt besproken dat de keuze voor de te reguleren actor (de provider of de gebruiker), de instrumenten (hoe worden overtredingen van regels ontdekt) en de mogelijke remedies allemaal bijdragen aan (potentiële) over- of onderregulering. Het verantwoordelijk maken van internettussenpersonen voor het reguleren van door gebruikers verstrekte informatie kan leiden tot (extraterritoriale) overregulering. Dit is vooral het geval wanneer de internettussenpersoon verplicht wordt om door gebruikers geplaatste informatie die in strijd is met de wet in de ene jurisdictie ook te verwijderen

voor gebruikers in een jurisdictie waar de inhoud niet illegaal is. In dit hoofdstuk wordt ook ingegaan op het onderscheid tussen curatie en moderatie. Voor zover curatie plaatsvindt als sanctie na een regelovertreding, wordt dit beschouwd als vergelijkbaar met moderatie.

In het derde en vierde hoofdstuk worden de aansprakelijkheidsregimes voor internettussenpersonen in de VS en de EU besproken. In de VS zijn internettussenpersonen uitgesloten van civiele aansprakelijkheid voor te goeder trouw genomen moderatiebesluiten en van aansprakelijkheid voor de inhoud van de door gebruikers geplaatste informatie indien. In de EU worden internettussenpersonen enkel vrijgesteld van aansprakelijkheid voor door gebruikers geplaatste informatie indien de internettussenpersonen niet op de hoogte of bewust was van de illegale inhoud ervan. Bovendien kunnen te actieve internettussenpersonen waarvan kan worden gezegd dat zij kennis of controle hebben verworven over door de gebruiker verstrekte informatie, niet op deze uitzondering rekenen. Deze benadering vormt een juridische prikkel om af te zien van vrijwillige moderatie omdat dit kan neerkomen op kennis of controle en dus op potentiële aansprakelijkheid. Het verwijderen of ontoegankelijk maken van door de gebruiker verstrekte informatie met illegale inhoud kan de internettussenpersoon alsnog rekenen op een vrijstelling voor aansprakelijkheid. Nieuwe voorstellen voor regelgeving laten deze "veilige havens" van de EU grotendeels ongemoeid.

In het vijfde hoofdstuk worden deze benaderingen geëvalueerd. Op basis van de voorgaande hoofdstukken zijn er drie factoren die bijdragen tot over- of onderregulering: ambiguïteit, de vereiste instrumenten en de vereiste remedies. In dit hoofdstuk wordt besproken dat strikte aansprakelijkheidsregimes de meeste risico's voor overregulering en onderregulering inhouden. Strikte aansprakelijkheidsregimes komen echter minder vaak voor in de VS en de EU. Voorwaardelijke-aansprakelijkheidsregimes (het meest gangbaar in de EU) kennen minder risico's voor onderregulering en een kleiner risico voor overregulering dan strikte aansprakelijkheidsregimes. Voorwaardelijke-aansprakelijkheidsregimes op basis van kennis (zoals in de EU) kunnen er echter toe leiden dat aanbieders afzien van vrijwillige moderatieinspanningen. Immuniteitsregimes vormen het grootste risico voor onderregulering, omdat aanbieders niet aansprakelijk kunnen worden gesteld voor illegale inhoud op hun diensten. Hoewel er geen risico is om aansprakelijk te worden gesteld, geven immuniteitsregelingen geen juridische prikkel om te reguleren. Omdat bij immuniteitsregelingen de aanbieders niet verplicht zijn instrumenten of een specifieke remedie in te zetten, maken immuniteitsregelingen een meer verfijnde benadering mogelijk om schendingen van wet- en regelgeving te adresseren. Het vierde regime besproken wordt gekarakteriseerd als niet-aansprakelijkheidsregulering. Hoewel vrijwillige codes niet wettelijk afdwingbaar zijn, kunnen zij worden beschouwd als de eerste stap in de richting van regulering, wat ertoe kan leiden dat aanbieders overreguleren om nieuwe wetgeving te voorkomen.

In *Wrongful moderation* wordt betoogd dat een voorwaardelijke-aansprakelijkheidsbenadering de beste kaarten heeft tegen overregulering en onderregulering en daarmee ook de beste waarborgen voor de vrijheid van meningsuiting van gebruikers. Een voorwaardelijke-aansprakelijkheidsbenadering kan – indien juist uitgevoerd – een sterke juridische prikkel vormen tegen onrechtmatige moderatie bestaande uit over- of onderverwijdering voorkomen. Voorwaardelijke-aansprakelijkheidsregimes zijn echter wel bijzonder kwetsbaar voor dubbelzinnige wetgeving. Wanneer de aanbieder moet voldoen aan een reeks onduidelijke of ingewikkelde voorwaarden, kan een voorwaardelijke-aansprakelijkheidsregime leiden tot overregulering of onderregulering vergelijkbare met strikte aansprakelijkheid. Omdat de aanbieder niet zeker weet of geslaagd een beroep kan worden gedaan op de vrijstellingen van

aansprakelijkheid, kan de aanbieder overgaan tot overmoderatie, (delen van) de dienst staken of moderatie-inspanningen staken om te voorkomen dat de aanbieder kennis verwerft van illegale inhoud.

Curriculum vitae

Michael Klos (1991) studied Public Administration (BSc) and Law (LLB) at Leiden University. Klos completed his master's degree in Jurisprudence and Philosophy of Law (LLM) in September 2017. Klos has been working as Lecturer/Researcher at the Department of Jurisprudence (Institute Metajuridica) at Leiden University from December 2017 to December 2021. Michael started his PhD research in September 2018 as an external PhD candidate at Leiden Law School of Leiden University under the supervision of Prof. dr. Paul Cliteur and mr. dr. Gelijn Molier. Since December 2021, Michael Klos has been associated with the department as a Researcher.