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

Klos, M.

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.