UCLA

JOURNAL OF LAW & TECHNOLOGY

SPECIAL ISSUE: GOVERNING THE DIGITAL SPACE

REGULATING THE EXPANDING CONTENT MODERATION UNIVERSE: A EUROPEAN PERSPECTIVE ON INFRASTRUCTURE MODERATION

Christoph Busch*

ABSTRACT

Much of the public and scholarly debate around content moderation focuses on user-facing platforms such as Facebook, Twitter, or YouTube. More recently, however, battles over content have shifted deeper into the internet stack, from the application layer to the infrastructure layer. As a consequence, hosting companies, domain registrars, ad networks, payment processors, and app stores are playing an increasingly important role in the battle over illegal and harmful content. Recent examples include the removal of the Parler app from iOS and Android app stores in the wake of

* Professor of Law and Director of the European Legal Studies Institute, University of Osnabrück, Germany; Visiting Fellow, Information Society Project, Yale Law School; Member of the European Commission's Expert Group for the EU Observatory on the Platform Economy; Reporter for the European Law Institute's Model Rules on Online Platforms. I am very grateful to the participants of the Yale ISP Writing Workshop for valuable advice on an earlier version of this Article. Special thanks to the editorial team of UCLA JOLT for their outstanding editing and thoughtful suggestions.

the January 6th riot at the Capitol. Similarly, Amazon suspended Parler from its webhosting service.

Against this background, this Article explores the various contexts and shapes of content moderation at the infrastructure layer, and examines how infrastructure moderation differs from content moderation at the application layer. One important difference is that infrastructure moderation is usually not about individual content items but rather about meaningful moderation practices (or the lack thereof) at higher levels in the content moderation stack. In this sense, infrastructure moderation can be characterized as a sort of "meta-moderation."

Building on these findings, the Article further examines how regulators react to the ongoing "infrastructural turn" and the expansion of the content moderation ecosystem. In doing so, the Article focuses on the latest regulatory developments in the European Union (EU), in particular the forthcoming Digital Services Act and the planned revision of the Code of Practice on Disinformation. The analysis shows that the EU is not only adapting the existing regulatory framework in response to the expansion of content moderation practices, but also actively promoting infrastructure moderation in the fight against disinformation.

In conclusion, the Article argues that there is an urgent need for elaborating principles tailored to the specifics of infrastructure moderation and ensuring subsidiarity, transparency, and procedural fairness. Such principles could provide guidance for providers of technical or financial infrastructure services when engaging in content moderation. They could also serve as a basis for the future development of a regulatory framework for responsible infrastructure moderation

TABLE OF CONTENTS

INTRODUCTION

On January 6, 2021, the assault on the U.S. Capitol sent a shockwave across the Internet. In the wake of the events, then-President Trump was banned from Twitter[1] and suspended by Facebook.[2] A few days later, YouTube blocked Trump's official channel.[3] Twitter also banned more than 70,000 accounts promoting conspiracy theories.[4] But the impact of the events in Washington D.C. was not limited to the major social media platforms. The shockwaves were also felt at the deeper layers of the Internet. Within forty-eight hours after the events, the Google Play Store and the Apple App Store confirmed that they would be suspending downloads of the Parler mobile application.[5] One day later, Amazon Web Services announced that it would stop providing cloud hosting services to Parler, which temporarily took the application offline.[6] The "great deplatforming"[7] of January 2021 also involved digital payment providers

---

1. *See, e.g.*, Sarah E. Needleman, *Twitter Bans President Trump's Personal Account Permanently*, WALL ST. J. (Jan. 8, 2021, 11:47 PM), https://www.wsj.com/articles/twitter-says-it-is-permanently-suspending-account-of-president-trump-11610148903.

2. *See, e.g.*, Mike Isaac & Kate Conger, *Facebook Banned Trump From Its Platforms For the Rest of His Term For Inciting Violence*, N.Y. TIMES (Jan. 7, 2021), https://www.nytimes.com/2021/01/07/us/politics/facebook-banned-trump-from-its-platforms-for-the-rest-of-his-term-for-inciting-violence.html.

3. Brian Fung, *YouTube Is Suspending President Donald Trump's Channel*, CNN BUS. (Jan. 13, 2021, 5:15 PM), https://edition.cnn.com/2021/01/12/tech/youtube-trump-suspension/index.html.

4 Kate Conger, *Twitter, in Widening Crackdown, Removes Over 70,000 QAnon Accounts*, N.Y. TIMES (Jan. 20, 2021), https://www.nytimes.com/2021/01/11/technology/twitter-removes-70000-qanon-accounts.html.

5. Brian Fung, *Parler Has Now Been Booted by Amazon, Apple and Google,* CNN BUS. (Jan. 11, 2021, 6:54 AM), https://www.cnn.com/2021/01/09/tech/parler-suspended-apple-app-store/index.html; *see also*

Apple, Google and Amazon Kick Parler Off Their Platforms, N.Y. TIMES (Jan. 9, 2021), https://www.nytimes.com/interactive/2021/01/09/us/parler-amazon-apple-google-responses.html.

6. Fung, *supra* note 5.

7. Adam Thierer, *The Great Deplatforming of 2021*, DISCOURSE (Jan. 14, 2021), https://www.discoursemagazine.com/politics/2021/01/14/the-great-deplatforming-of-

such as PayPal, which decided to block groups that helped Trump supporters travel to Washington D.C.[8]

These examples illustrate that content moderation is not limited to major social media platforms, but is a much broader phenomenon which exists in many contexts and takes many forms. So far, most of the public and scholarly debate around content moderation has focused on user-facing platforms such as Facebook, Twitter, and YouTube.[9] Less often discussed, however, are content moderation decisions taken at deeper layers of the Internet by providers of seemingly neutral infrastructure services, such as hosting companies, domain registrars, and networks, or payment processors.[10]

Recently, however, there has been a growing interest in content moderation at the infrastructure level.[11] Some legal scholars have expressed skepticism about infrastructure moderation. In this sense, Jack Balkin has argued that providers of infrastructure services should not engage in content moderation at all and "should concern themselves only with legality or illegality of transactions."[12] Similarly, others have suggested that

---

2021/; *see also* ROBERT SPRAGUE, NORMALIZING DE-PLATFORMING: THE RIGHT NOT TO TOLERATE THE INTOLERANT 2 (2021), https://ssrn.com/abstract=3915739.

8. Jennifer Surane, *PayPal Blocks Group That Helped Trump Supporters Travel to D.C.*, BLOOMBERG (Jan. 7, 2021, 2:13 PM), https://www.bloomberg.com/news/articles/2021-01-07/paypal-blocks-group-that-helped-trump-supporters-travel-to-d-c.

9. *See generally* James Grimmelmann, *The Virtues of Moderation*, 17 YALE J.L. & TECH. 42 (2015); Jack M. Balkin, *Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation*, 51 U.C. DAVIS L. REV. 1149 (2018); Kate Klonick, *The New Governors: The People, Rules, and Processes Governing Online Speech*, 131 HARV. L. REV. 1598 (2018); Hannah Bloch-Wehba, *Automation in Moderation*, 53 CORNELL INT'L L.J. 41 (2020).

10. *See generally* Natasha Tusikov, *Defunding Hate: PayPal's Regulation of Hate Groups*, 17 SURVEILLANCE & SOC'Y 46 (2019).

11. *See, e.g.*, Tarleton Gillespie et al., *Expanding the Debate About Content Moderation: Scholarly Research Agendas for the Coming Policy Debates*, 9 INTERNET POL'Y REV. 1 (2020), https://doi.org/10.14763/2020.4.1512; *see also* Jonathan Zittrain, *The Inexorable Push for Infrastructural Moderation*, TECHDIRT: TECH POL'Y GREENHOUSE (Sept. 24, 2021, 12:09 PM), https://www.techdirt.com/articles/20210924/12012347622/inexorable-push-infrastructure-moderation.shtml.

12. Jack M. Balkin, *How to Regulate (and Not Regulate) Social Media*, 1 J. FREE SPEECH L. 71, 73 (2021) ("For basic internet services the regulatory answer is pretty simple: non-

infrastructure providers should adopt a policy of "content agnosticism."[13] This Article, in contrast, accepts the reality that content moderation is "an expanding socio-technical phenomenon,"[14] which inevitably is spreading horizontally across the application layer and vertically creeping down the Internet stack towards the infrastructure layer.

Starting from this premise, the Article explores the different shapes and contexts of content moderation by providers of infrastructure services and attempts to shed some light on this underexamined but rapidly evolving part of the content moderation ecosystem. In doing so, the Article makes several contributions to the content moderation debate.

1. The survey of content moderation practices at different levels of the Internet stack shows that infrastructure moderation is usually not about individual content items; it is instead about meaningful moderation practices (or the lack thereof) at higher levels in the content moderation stack. In this sense, infrastructure moderation can be characterized as a sort of meta-moderation or second-order moderation.
2. The Article adds a European perspective to the debate. It shows how the EU is both adapting the existing regulatory framework in response to the expansion of content moderation practices, and actively promoting infrastructure moderation in the fight against disinformation.
3. The Article argues that there is an urgent need for the elaboration of principles tailored to the specifics of infrastructure moderation, ones that ensure subsidiarity, transparency, and procedural fairness. Such principles could provide guidance for providers of technical or financial infrastructure services when engaging in content moderation. Moreover, they could also serve as a basis for the future development of a regulatory framework for responsible infrastructure moderation.

This Article proceeds in three Parts. Part I provides a brief overview of the content moderation stack and surveys how content moderation is expanding into the Internet infrastructure layer. Part II turns to the question

---

discrimination. Let the bits flow freely and efficiently. Don't engage in content regulation at this level.").

13. Annemarie Bridy, *Remediating Social Media: A Layer-Conscious Approach*, 24 B.U. J. SCI. & TECH. L. 193 (2018).

14. Gillespie et al., *supra* note 11, at 3.

of whether and how the EU's recent regulatory initiatives for platform economy address the emerging issue of infrastructure moderation. In doing so, Part II focuses on the latest regulatory developments in the European Union, particularly the forthcoming Digital Services Act and the planned revision of the Code of Practice on Disinformation. Building on this analysis, Part III seeks to identify several elements that might inform foundational principles of responsible infrastructure moderation.

I. CONTENT MODERATION STACK

Internet architecture is usually represented as a layered model consisting of interrelated technologies that build and depend on one another to create the network we know as the Internet. There are different versions of this "Internet stack" and there is no single definitive model.[15] For purposes of this article, we can use a simplified model that distinguishes between the application layer (which includes websites, social media and other public-facing platforms) and the infrastructure layer (which includes everything beneath the application layer, from hosting services, content delivery networks, domain registries and registrars, all the way down to ISPs). For some actors (e.g., app stores) it is difficult to say where exactly they are located in the stack as they combine elements of user-facing and infrastructural services. The same applies to providers of financial services, such as PayPal or Stripe. Traditionally, these services are not considered as part of the Internet stack. But when it comes to content moderation, they play an increasingly important role as chokepoints for online speech. From this perspective, providers of financial services can also be considered part of the expanding content moderation ecosystem.

Before taking a closer look at the different layers of the content moderation stack, it is necessary to clarify the kinds of policies and practices to which the term content moderation refers. Some scholars broadly define content moderation as "the detection of, assessment of, and interventions taken on content or behaviour deemed unacceptable by platforms or other information intermediaries, including the rules they impose, the human labour and technologies required, and the institutional mechanisms of adjudication, enforcement and appeal that support it."[16] The proposal for a

---

15. *See, e.g.*, ULRIKE UHLIG ET AL., HOW THE INTERNET REALLY WORKS: AN ILLUSTRATED GUIDE TO PROTOCOLS, PRIVACY, CENSORSHIP, AND GOVERNANCE 76–77 (2021).

16. Gillespie et al., *supra* note 11; *see also* Grimmelmann, *supra* note 9, at 47 (defining content moderation as "governance mechanisms that structure participation in a community to facilitate cooperation and prevent abuse.").

Digital Services Act (DSA), published by the European Commission in December 2020, uses the following definition:

> "[C]ontent 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.[17]

This definition is rather broad in two respects: first, it not only covers decisions to remove content and disable user accounts, but also other measures that influence the "availability, visibility and accessibility" of content, such as downranking and deamplification. This underlines that, depending on the context and the position in the Internet stack, moderation decisions can be much more nuanced than a binary remove-or-not decision.[18] Second, the definition is not limited to "illegal content,"[19] but also includes measures taken with regard to information which is incompatible with the terms and conditions of the online service provider. In other words, the DSA definition includes any activities used by service providers to enforce their private "house rules" (often referred to as "Community

17. *Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and Amending Directive 2000/31/EC* art. 2(p), at 45, COM (2020) 825 final (Dec. 15, 2020) [hereinafter DSA]. The DSA also sets out detailed procedural requirements for content moderation by platforms. *Id.* at 51–52.

18. *See generally* Eric Goldman, *Content Moderation Remedies*, 28 MICH. TECH. L. REV. 1 (2021), http://dx.doi.org/10.2139/ssrn.3810580 (describing dozens of different remedies that Internet services have imposed as part of the content moderation activities).

19. DSA, *supra* note 17, art. 2(g), at 45 (defining "illegal content" as "any information, which, in itself or by its reference to an activity, including the sale of products or provision of services is not in compliance with Union law or the law of a Member State, irrespective of the precise subject matter or nature of that law").

Standards" or "Acceptable Use Policies") with regard to content that is legal, but nevertheless considered objectionable ("lawful, but awful").[20]

A. Application Layer

Much of the public debate around content moderation focuses on social media platforms such as Facebook, Twitter, or YouTube, which play a key role in public discourse. The evolution of content moderation on social media platforms has been well documented from various perspectives.[21] Despite all the justified criticism of content moderation practices by prominent social media platforms, it is probably fair to say that the governance of content moderation on these platforms is quite developed compared to content moderation practices at the infrastructure level. Today, Facebook, Twitter, and YouTube all have reasonably detailed community standards. They regularly publish transparency reports. Some platforms have even taken steps toward institution building, as shown by the Facebook Oversight Board.

Of course, the debate about content moderation in the application layer is no longer limited to the largest social media platforms but expanding to other user-facing services: Recently, messaging services, such as WhatsApp and Telegram, have come into focus. Messaging services provide a technical platform for private communication not only between individuals, but also among small and closed groups. In this sense, a messaging service like WhatsApp "can be understood as social media insofar as content sharing among small and large groups, public communication, interpersonal connection, and commercial transactions converge in key features of the app."[22] Indeed, some messaging services are

---

20. Eric Goldman & Jess Miers, *Online Account Termination/Content Removals and the Benefits of Internet Services Enforcing Their House Rules*, 1 J. FREE SPEECH L. 191, 194–95 (2021).

21. *See, e.g.*, Kate Klonick, *The Facebook Oversight Board: Creating an Independent Institution to Adjudicate Online Free Expression*, 129 YALE L.J. 2418, 2428–48 (2020) (discussing the history of content moderation at Facebook); *see also* SARAH T. ROBERTS, BEHIND THE SCREEN: CONTENT MODERATION IN THE SHADOWS OF SOCIAL MEDIA (2019) (offering an ethnographic study of commercial content moderation on social media platforms).

22. Ariadna Matamoros-Fernández, *Encryption Poses Distinct New Problems: The Case of WhatsApp*, 9 INTERNET POL'Y REV. 7, 7 (2020).

increasingly used for content sharing among larger groups and via open channels.

In Germany, for example, Telegram has become an influential forum during the pandemic for organizing offline protests against lockdowns, mask mandates, and vaccination programs: Since June 2020, it has been used to organize at least 4,300 protests, some of which have been linked to Telegram channels related to the QAnon conspiracy theory.[23] The controversy about Telegram culminated in December 2021, when the German police arrested a group of extreme anti-vaccine campaigners in connection with an alleged murder plot against the state premier of Saxony, which had been shared on Telegram.[24] According to news reports, some 130 members on a Telegram chat group had been sharing messages encouraging people to oppose government measures "with armed force if needed."[25] In response to these reports, German politicians demanded that Apple and Google remove Telegram from their app stores.[26]

The rise of messaging services as new fora for hate speech and disinformation illustrates that issues relating to content moderation are not limited to social media platforms. Indeed, "there's nothing stopping people from choosing to gather and have a conversation within World of Warcraft,

---

23. Jordan Wildon & Kristina Gildejeva, *Assessing the Scale of German Language Disinformation Communities on Telegram*, LOGICALLY (Sept. 10, 2021, 9:26 AM), https://www.logically.ai/articles/german-language-disinformation-telegram (reporting that the largest German-language QAnon channel on Telegram has at least 152,000 subscribers); *see also* Mark Scott, *Ahead of German Election, Telegram Plays Radicalizing Role*, POLITICO (Sept. 22, 2021, 9:16 AM), https://www.politico.eu/article/german-telegram-election-misinformation/.

24. Erika Solomon, *Olaf Scholz Warns of Threat from Germany's Extreme Anti-Vaccine Campaigners*, FIN. TIMES (Dec. 15, 2021), https://www.ft.com/content/dcd6eef5-d7c8-4407-84ac-d16c15098f12.

25. *Id.*

26. Max Hoppenstedt, *Was die Politik gegen Telegram unternehmen kann*, SPIEGEL (Dec. 14, 2021), https://www.spiegel.de/netzwelt/telegram-was-die-politik-gegen-die-gefaehrliche-chat-app-unternehmen-kann-a-7245e6fd-b057-41b9-9b50-69095458cd54; *see also* Axel Kannenberg, *Innenministerin Faeser: Telegram-App soll aus App-Stores entfernt werden*, HEISE (Jan. 19, 2022, 7:05 PM), https://www.heise.de/news/Innenministerin-Faeser-Telegram-App-soll-aus-App-Stores-entfernt-werden-6332582.html (reporting that the German Federal Interior Secretary Nancy Faeser called upon Apple and Google to take their "social responsibility" seriously and delete the Telegram app from their app stores).

merely admiring the view of the game's countryside as they chat about sports, politics or alleged terrorist schemes."[27] Therefore, the scholarly and public debate about content moderation needs to widen its scope to a broader range of applications that could be used for the dissemination of problematic content.

B. Infrastructure Layer

Moderation decisions about illegal or objectionable content are not only made at the application layer: There is growing evidence that content moderation is "bleeding" from social media into the Internet's infrastructure layer.[28] The content-related decisions taken at this level are often even more opaque than the decisions made by user-facing platforms such as Facebook or Twitter.[29]

The next two Subparts discuss various forms of content moderation by providers of infrastructure services. The first Subpart focuses on decisions to remove content made by providers of technical infrastructure services, the so-called "plumbers" of the Internet (sometimes referred to as "deep deplatforming").[30] The second Subpart turns to content moderation decisions made by providers of financial services (sometimes referred to as "financial deplatforming"),[31] which also can be considered a type of infrastructure moderation in the broader sense.

1. Technical Infrastructure

Beneath the application layer of the Internet with its user-facing applications and websites, there is a plethora of technical services that keep the global network running. It would be naive to assume that these infrastructures are apolitical. On the contrary, the design of Internet infrastructure and its administration "internalize the political and economic values that ultimately influence the extent of freedom and innovation."[32] It

---

27. Zittrain, *supra* note 11.

28. *Id.*

29. Gillespie et al., *supra* note 11, at 6.

30. Will Duffield, *A Brief History of 'Deep Deplatforming'*, CATO INST.: CATO AT LIBERTY (Jan. 22, 2021, 4:21 PM), https://www.cato.org/blog/brief-history-deep-deplatforming.

31. *Id.*

32. Laura DeNardis, *Hidden Levers of Internet Control: An Infrastructure-Based Theory of Internet Governance*, INFO., COMMC'N & SOC'Y 720, 721 (2012).

is, therefore, not surprising that the battle over which content is and is not acceptable is also being fought at the infrastructure level.

a. App Stores

A key role in infrastructure moderation is played by app stores such as the Google Play Store and the Apple App Store. In a sense, app stores are situated at the interface between the application layer and the infrastructure layer. On the one hand, the app stores provide user-facing services which end users can directly access via their smartphones. On the other hand, from the perspective of app developers, app stores constitute essential infrastructure for the distribution of their apps.[33]

This bottleneck position affords Apple and Google the power to decide which apps are available in their app stores, and also allows them to indirectly determine how content is distributed through moderating their apps. For example, Apple's App Store Review Guidelines stipulate that apps should not include "objectionable content," a concept which is rather vaguely circumscribed in the guidelines.[34] Further, and more interestingly, they also prescribe how apps themselves should deal with user-generated content:

To prevent abuse, apps with user-generated content or social networking services must include:

- A method for filtering objectionable material from being posted to the app
- A mechanism to report offensive content and timely responses to concerns
- The ability to block abusive users from the service
- Published contact information so users can easily reach [the developer].[35]

---

33. *See, e.g.*, Nikolas Guggenberger, *Essential Platforms*, 24 STAN. TECH. L. REV. 237, 268 (2021) (arguing that the two major app stores' apps are essential infrastructures as app developers lack practical and reasonable alternatives to the offers by Apple and Google's offerings).

34. *App Store Review Guidelines*, APPLE DEV.: APP STORE ¶ 1.1, https://developer.apple.com/app-store/review/guidelines/ (Oct. 22, 2021).

35. *Id.* ¶ 1.2.

In other words, Apple defines a set of minimum requirements for content moderation for third-party apps. This could be characterized as a kind of meta-moderation or second-order moderation through which Apple, as a provider of infrastructure services, exercises indirect control over content moderation policies implemented at the application level.

The app stores' considerable influence over content moderation is perhaps most conspicuously demonstrated by the deplatforming of Parler in January 2021. But this is not the only example. In July 2021, the Apple App Store banned Unjected, a mobile dating app specifically aimed at connecting people who are unvaccinated against COVID-19.[36] The app, which became known as "Tinder for anti-vaxxers," allegedly violated Apple's COVID-19 guidelines.[37] Another example of both Google and Apple using more targeted interventions was when the Telegram application blocked German far-right conspiracy theorist Attila Hildmann from their Android and iOS apps in June 2021 after receiving pressure from both Google and Apple to do so.[38]

b. Web Hosting and Cloud Services

Another important category of infrastructure is web hosting and cloud storage providers. Typical examples are infrastructure-as-a-service (IaaS), which provides servers for cloud computing and storage, and platform-as-a-service (PaaS), which supplies a digital environment for developing, running, and managing applications on the provider's cloud service. Major cloud computing services include Amazon Web Services (AWS), Google Cloud, Microsoft Azure, and Oracle, but many smaller and niche players also exist in the market.

An early example that shed a bright light on the role of web hosting providers as chokepoints for online content was the 2010 deplatforming of

---

36. Chance Miller, *Apple Boots 'Tinder for Anti-Vaxxers' App from the App Store for Violating COVID-19 Guidelines*, 9TO5MAC (July 31, 2021, 12:34 PM), https://9to5mac.com/2021/07/31/apple-boots-tinder-for-anti-vaxxers-app-from-the-app-store-for-violating-covid-19-guidelines/. At the time of writing, the app seems to be available again in the Apple app store (last visited Jan. 1, 2022).

37. *Id.*

38. *Zugang zu Telegram-Kanälen von Attila Hildmann gesperrt*, FRANKFURTER ALLGEMEINE (June 9, 2021, 3:05 PM), https://www.faz.net/aktuell/gesellschaft/kriminalitaet/zugang-zu-telegram-kanaelen-von-attila-hildmann-gesperrt-17380316.html.

Wikileaks by AWS.[39] More recent examples of deplatforming by hosting services include the eviction of Gab, a platform known for its far-right userbase, from Microsoft Azure in and AWS in the wake of the 2017 shooting at a Pittsburgh synagogue.[40] Gab is no longer relying on a cloud service provider—its data are stored on rented servers in an undisclosed data center.[41]

Compared to the relatively detailed specifications in Apple's App Store Review Guidelines, AWS' house rules are rather vague. Among other things, AWS's "Acceptable Use Policy" prohibits the use of its services "for any illegal or fraudulent activity, to violate the rights of others, to threaten, incite, promote, or actively encourage violence, terrorism or other serious harm."[42] In particular, AWS does not stipulate any specific content moderation requirements for user-generated content hosted on its servers. In summary, content related decisions by web hosting providers seem to be more opaque and ad hoc than those made by app stores.

c. Content Delivery Networks and Security Services

Among the lesser known, but important, infrastructure services are content delivery networks (CDNs) such as Cloudflare, Akamai, Peer5, and Amazon Cloudfront, which increasingly play a role in content moderation. CDNs are large, distributed systems of multiple servers which improve the speed, security, and performance of websites. In addition to accelerating the delivery of content delivery, CDNs often provide additional security services, in particular protection against distributed denial of service (DDoS) attacks that overwhelm a server with fake traffic.

These security services are particularly important for disinformation and hate sites, which are frequently targeted by DDoS attacks carried out

---

39. Duffield, *supra* note 30 (reporting that Wikileaks, in search of an alternative web hosting service, found shelter with a Swiss provider).

40. José Van Dijck et al, *Deplatformization and the Governance of the Platform Ecosystem*, NEW MEDIA & SOC'Y 1, 6, 8 (2021), doi: 10.1177/14614448211045662.

41. Robert McMillan & Aaron Tilley, *Parler Faces Complex, Costly Route to Getting Back Online*, WALL ST. J. (Jan. 12, 2021, 1:12 PM), https://www.wsj.com/articles/parler-faces-obstacles-to-getting-back-online-11610474343.

42. *AWS Acceptable Use Policy*, AMAZON WEB SERVS., https://aws.amazon.com/de/aispl/aup/ (July 14, 2021).

by activists. This became clear when Cloudflare decided to terminate its services to the Daily Stormer, a white supremacist website, These security services are particularly important for disinformation and hate sites, which are frequently targeted by DDoS attacks carried out by activists. This became clear when Cloudflare decided to terminate its services to the Daily Stormer, a white supremacist website, in 2017 and 8chan, a website notorious for hosting lawless message boards, in 2019.[43] In response to the termination of their services for the Daily Stormer, which took place in the aftermath of the Charlottesville riots of August 11, 2017, the CEO and Co-Founder of Cloudflare published an influential blog post about the role of infrastructure providers in content moderation. He explained why the company had dropped the Daily Stormer and, at the same time, highlighted the "risks of a company like Cloudflare getting into content policing."[44] Similarly, when Cloudflare terminated its services for 8chan after the shooting in El Paso on August 3, 2019, the CEO stressed that he felt "incredibly uncomfortable about playing the role of content arbiter," and that he did not plan to exercise it often.[45]

d. Domain Registrars

Even farther down the Internet stack, the Domain Name System (DNS) is where one finds domain registrars such as GoDaddy, Tucows, DreamHost, and Epik. Registrars are companies accredited by the Internet Corporation for Assigned Names and Numbers to sell domain names that allow users to easily access a website. In doing so, they serve as middlemen between the registry operators (like Verisign for .com and .net) and the registrant of a domain name. Registrars can block a website's domain name by removing its registration from the registry. This makes them an important bottleneck for controlling what is and is not accessible online.

In May 2020, a group of major registrars published a "Framework to Address Abuse" in the DNS, which outlines the types of content upon

---

43. Tim Elfrink, *'A Cesspool of Hate': U.S. Web Firm Drops 8chan After El Paso Shooting*, WASH. POST (Aug. 5, 2019), https://www.washingtonpost.com/nation/2019/08/05/chan-dropped-cloudflare-el-paso-shooting-manifesto/.

44. Matthew Prince, *Why We Terminated Daily Stormer*, CLOUDFLARE (Aug. 17, 2017), https://blog.cloudflare.com/why-we-terminated-daily-stormer/.

45. Matthew Prince, *Terminating Service for 8Chan*, CLOUDFLARE (Aug. 4, 2019), https://blog.cloudflare.com/terminating-service-for-8chan/.

which a domain registrar should act.[46] The Framework defines five broad categories of harmful activities to which registrars should respond (malware, botnets, phishing, pharming, and spam). In contrast to these activities referred to as "DNS abuses," the Framework emphasizes that so-called "Website Content Abuses"—a term referring to harmful or illegal content of a particular website—do not warrant the suspension or blocking of a domain.[47] According to the signatories of the Framework, "this distinction is critical in order for the Internet to remain open for free expression"; they underline that acting at the DNS level to address content-related issues "in general is a disproportionate remedy that can cause significant collateral damage."[48]

An exception to this principle is, however, made for illegal content related to "the physical and often irreversible threat to human life," such as: (1) child sexual abuse materials; (2) illegal distribution of opioids online; (3) human trafficking; and (4) specific and credible incitements to violence.[49] This last exception was invoked by GoDaddy when it suspended the Daily Stormer's domain in the aftermath of the 2017 Charlottesville riots. Likewise in 2018 when GoDaddy terminated its services for Gab, a right-wing social network, after it had emerged that a man who killed 11 people at a Pittsburgh synagogue used Gab to spread anti-Semitic messages.[50] Of course, the standards applied by registrars differ significantly: Soon after the Gab was evicted by GoDaddy, the controversial domain registrar Epik announced that it had "welcomed the domain Gab.com."[51] Similarly, in

46. Framework to Address Abuse, DNS ABUSE FRAMEWORK (May 29, 2020), https://dnsabuseframework.org/media/files/2020-05-29_DNSAbuseFramework.pdf.

47. *Id.* at 3.

48. *Id.*

49. *Id.*

50. Sam Byford, *Gab.com Goes Down After GoDaddy Threatens to Pull Domain*, VERGE (Oct. 28, 2018, 11:11 PM), https://www.theverge.com/2018/10/28/18036520/gab-down-godaddy-domain-blocked.

51. Rob Monster, *Why Epik Welcomed Gab.com*, EPIK (Nov. 3, 2018), https://www.epik.com/blog/why-epik-welcomed-gab-com.html.

January 2021, it was reported that Epik had also become the domain registrar for Parler.[52]

e. Internet Service Providers

Finally, below domain registrars and registries are Internet service providers (ISPs), such as Comcast, AT&T, and Verizon. They offer Internet access to users via broadband or cellular networks. While it is not uncommon for ISPs to block certain illegal content (such as piracy websites), they have mainly tried to stay out of the more political content moderation debate. However, this does not mean that ISPs cannot be used for politically motivated content moderation. For example, in March 2021 it was reported that the Russian government would slow down access to Twitter after accusing the social network of failing to remove illegal content.[53]

2. Financial Infrastructure

Websites and apps that distribute content not only rely on technical infrastructures, but also on financial infrastructures. This also applies to disinformation and hate sites that need to fund their activities. In particular, online advertising providers such as Google Ads and payment processors such as PayPal and Patreon play an important role in this context.[54] The sale of merchandise on e-commerce websites and online marketplaces also makes an important contribution to the funding of disinformation campaigns.

As a consequence, financial infrastructure services have increasingly come into focus as potential chokepoints for online speech. In particular, the high concentration and strict regulation of financial service providers

---

52. Matt Binder, *Parler Transfers Domain Name to Epik, Domain Registrar of Choice for the Far Right*, MASHABLE (Jan.11, 2021), https://mashable.com/article/parler-domain-name-epik; *see also* Geoffrey A. Fowler & Chris Alcantara, *Gatekeepers: These Tech Firms Control What's Allowed Online*, WASH. POST (Mar. 24, 2021), https://www.washingtonpost.com/technology/2021/03/24/online-moderation-tech-stack/.

53. Anton Troianovski & Andrew E. Kramer, *Russia Says It Is Slowing Access to Twitter*, N.Y. TIMES (Oct. 22, 2021), https://www.nytimes.com/2021/03/10/world/europe/russia-twitter.html.

54. Catherine Han et al., *On the Infrastructure Providers that Support Misinformation Websites*, ZAKIRD, https://zakird.com/papers/misinfo-infra-preprint.pdf (last visited Mar. 8, 2022).

makes the denial of payment processing an effective tool for indirect content moderation via "financial deplatforming."[55] Using this perspective, providers of financial services can also take the role of "meta-moderators"[56] and shape the content policies of platforms that rely on them. The power of financial deplatforming became apparent in 2010 when Bank of America, VISA, MasterCard, PayPal, and Western Union all prohibited donations to Wikileaks.[57] And in 2018, the social network Gab was banned by Paypal and Stripe in the aftermath of the Pittsburgh synagogue shooting.[58]

The practices of content moderation by financial infrastructure providers are not limited to whistleblowing platforms and right-wing social networks, but also affect other categories of content that are considered objectionable. In August 2021, for example, the Internet content subscription service OnlyFans announced that it would be blocking content creators from posting explicit photos and videos at the request of its "banking partners and payout providers."[59] This decision was reversed a

55. Will Duffield, *Bankers as Content Moderators*, CATO INST. (Sept. 27, 2021), https://www.cato.org/commentary/bankers-content-moderators ("There are hundreds of domain registrars, but only a handful of major payment processors. This disparity makes the denial of payment processing one of the most effective levers for controlling speech."); *see also* Charles Arthur, *WikiLeaks Claims Court Victory Against Visa*, GUARDIAN (July 12, 2012, 12:40 PM), https://www.theguardian.com/media/2012/jul/12/wikileaks-court-victory-visa.

56. Duffield, *supra* note 55.

57. Arthur, *supra* note 55.

58. Tim Bradshaw, *Stripe Steps Away from Gab Network After Synagogue Shooting*, IRISH TIMES (Oct. 28, 2018, 7:12 PM), https://www.irishtimes.com/business/technology/stripe-steps-away-from-gab-network-after-synagogue-shooting-1.3678990; *see also After Pittsburgh Synagogue Shooting, PayPal Bans Gab Social Network*, CNET (Oct. 27, 2018, 4:17 PM), https://www.cnet.com/tech/services-and-software/after-pittsburgh-synagogue-shooting-paypal-bans-gab-social-network/.

59. Taylor Lorenz & Alyssa Lukpat, *OnlyFans Says It Is Banning Sexually Explicit Content*, N.Y. TIMES (Sept. 9, 2021), https://www.nytimes.com/2021/08/19/business/onlyfans-porn-ban.html.

few days later, but only after OnlyFans' banking partners had assured the company that it "can support all genres of creators."[60]

While the past decisions on financial deplatforming tended to be taken on an ad hoc basis, there have recently been signs of a more systematic approach by financial content moderators. In July 2021, it was reported that the online payment processor PayPal would be partnering with the Anti-Defamation League to investigate and share information about accounts that the ADL considers too extreme.[61]

Of course, websites and platforms that have been ejected by mainstream payment processors can resort to peer-to-peer payment systems and cryptocurrencies. For example, WikiLeaks turned to Bitcoin donations after being deplatformed by PayPal and other payment processors.[62] Gab likewise shifted to cryptocurrencies after being barred from Paypal, Venmo, Square, and Stripe.[63] But there are signs that financial content moderation is slowly expanding into the realm of cryptocurrencies. In 2019 it was reported that Coinbase, a major digital currency exchange, closed an account held by Gab.[64] Thus, even within the decentralized

---

60. Timothy Bella & Lateshia Beachum, *OnlyFans Reverses Ban on Sexually Explicit Content After Wide Backlash from Its Users*, WASH. POST (Aug. 25, 2021, 12:50 PM), https://www.washingtonpost.com/business/2021/08/25/onlyfans-reversal-sex-porn-ban/.

61. *PayPal Partners with ADL to Fight Extremism and Protect Marginalized Communities*, ANTI-DEFAMATION LEAGUE (July 26, 2021), https://www.adl.org/news/press-releases/paypal-partners-with-adl-to-fight-extremism-and-protect-marginalized; *see also* Anna Irrera, *PayPal to Research Transactions That Fund Hate Groups, Extremists*, REUTERS (July 26, 2021, 1:43 PM), https://www.reuters.com/business/finance/paypal-research-blocking-transactions-that-fund-hate-groups-extremists-2021-07-26/.

62. Roger Huang, *How Bitcoin and WikiLeaks Saved Each Other*, FORBES (Apr. 26, 2019, 1:26 PM), https://www.forbes.com/sites/rogerhuang/2019/04/26/how-bitcoin-and-wikileaks-saved-each-other/.

63. Michael del Castillo, *The Alt-Right's Favorite Social Network Gab's Plan to Use Blockchain to Make Itself Indestructible*, FORBES (Oct. 31, 2018, 6:14 AM), https://www.forbes.com/sites/michaeldelcastillo/2018/10/31/the-alt-rights-favorite-social-network-gabs-plan-to-use-blockchain-to-make-itself-indestructible/; see also Tess Owen, *Gab is Back in Business After Finding a Payments Processor Willing to Work with the Alt-Right*, VICE (January 23, 2019, 5:14 PM), https://www.vice.com/en/article/eve43n/gab-is-back-in-business-after-finding-a-payments-processor-willing-to-work-with-the-alt-right.

64. Erin Carson, *Gab Says It Was Kicked Off Coinbase*, CNET (Jan. 7, 2019, 12:16 PM), https://www.cnet.com/news/gab-says-it-was-kicked-off-coinbase-again/

blockchain ecosystem, there are certain choke points which can be used for financial content moderation. The next stage of the development could be a migration of platforms like Gab to decentralized crypto exchanges which enable users to buy and sell cryptocurrencies through peer-to-peer transactions relying on automated smart contracts and without any interference from a third party.

II. EU PLATFORM REGULATION MEETS INFRASTRUCTURE MODERATION

Part I has provided an overview of the different layers that make up the content moderation stack and has surveyed the different contexts and shapes in which infrastructure moderation happens. It has also shown that infrastructure moderation differs from content moderation at the application level in that it is usually not about individual content items, but rather about moderation practices (or the lack thereof) at higher levels in the content moderation stack. Building on these findings, Part II turns to the question of whether and how the EU's recent initiatives to regulate the platform economy address the emerging issue of infrastructure moderation. Answering this question requires a closer look at two recent EU proposals: the DSA[65] unveiled in December 2020, and the Guidance on Strengthening the Code of Practice on Disinformation[66] published in May 2021. Both initiatives show that the European regulatory framework is slowly adapting to the expansion of content moderation practices into the infrastructure layer.

A. Digital Services Act

On December 15, 2020, the European Commission published its much anticipated proposed DSA. The draft regulation is part a of a more comprehensive legislative package, which also includes the Digital Markets Act (DMA).[67] Together, the DSA and the DMA aim to create a new regulatory framework for the governance of digital services in the European Union. The DMA introduces *ex ante* rules applicable only to large

---

65. DSA, *supra* note 17.

66. *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions on European Commission Guidance on Strengthening the Code of Practice on Disinformation*, COM (2021) 262 final (May 26, 2021).

67. *Proposal for a Regulation of the European Parliament and of the Council on Contestable and Fair Markets in the Digital Sector (Digital Markets Act)*, COM (2020) 842 final (Dec. 15, 2020).

online platforms which act as "gatekeepers."[68] The DSA has a much broader scope—it aims to update the existing rules on platform responsibilities in the provision of digital services by means of revising and complementing the e-Commerce Directive 2000/31/EC (ECD).[69] Unlike earlier legislative reforms and regulatory initiatives regarding platform responsibilities, which focused on specific issues[70] or specific platforms,[71] the DSA adopts a horizontal approach that covers a broad range of issues and digital intermediaries. It seeks to update the existing European rules on digital intermediaries, and also aims to counter an emerging fragmentation of the regulatory framework for digital services in the EU.

---

68. *See* Alexandre de Streel et al., *The European Proposal for a Digital Markets Act: A First Assessment*, CTR. ON REG. IN EUR. (CERRE) 11 (Jan. 19, 2021), https://cerre.eu/publications/the-european-proposal-for-a-digital-markets-act-a-first-assessment/.

69. Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on Certain Legal Aspects of Information Society Services, in Particular Electronic Commerce, in the Internal Market (Directive on Electronic Commerce), 2000 O.J. (L 178) 1.

70. *See*, *e.g.*, Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on Copyright and Related Rights in the Digital Single Market and Amending Directives 96/9/EC and 2001/29/EC, 2019 O.J. (L 130) 92 (copyright); *Proposal for a Regulation of the European Parliament and of the Council on a Temporary Derogation from Certain Provisions of Directive 2002/58/EC of the European Parliament and of the Council as Regards the Use of Technologies by Number-Independent Interpersonal Communications Service Providers for the Processing of Personal and Other Data for the Purpose of Combatting Child Sexual Abuse Online*, COM (2020) 568 final (Sept. 10, 2020) (child abuse); *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions on a Strategy for a More Effective Fight Against Child Sexual Abuse*, COM (2020) 607 final (July 27, 2020) (child abuse); EUROPEAN COMM'N, CODE OF CONDUCT ON COUNTERING ILLEGAL HATE SPEECH ONLINE (2016), https://ec.europa.eu/newsroom/document.cfm?doc_id=42985 (illegal hate speech); European Commission, *Proposal for a Regulation of the European Parliament and of the Council on Preventing the Dissemination of Terrorist Content Online*, COM (2018) 640 final (Sept. 12, 2018) (terrorist content).

71.*See* Directive (EU) 2018/1808 of the European Parliament and of the Council of 14 November 2018 Amending Directive 2010/13/EU on the Coordination of Certain Provisions Laid Down by Law, Regulation or Administrative Action in Member States Concerning the Provision of Audiovisual Media Services (Audiovisual Media Services Directive) in View of Changing Market Realities, 2018 O.J. (L 303) (audio-video sharing platforms).

In recent years, several EU member states have enacted national laws that set rules for content moderation for social media platforms. Prominent examples are the German NetzDG,[72] the French Loi Avia,[73] and the Austrian Communication Platforms Act.[74] The DSA aims to replace these national regulations with a uniform European regulatory framework.

1. Regulatory Approach Towards Content Moderation

Generally speaking, the DSA makes three contributions to the regulatory framework for content moderation by: (1) setting out general rules of liability for providers of intermediary services; (2) establishing a regime of due diligence obligations, with a special focus on online platforms including social media; and (3) strengthening the cooperation between national authorities in charge of the public enforcement of online regulation.[75]

The liability rules for providers of online intermediary services set out in Chapter II of the DSA form the legal backdrop for content-related decisions taken by the various actors in the content moderation stack.[76] These rules determine under which circumstances online intermediaries, such as ISPs, hosting providers, or social media platforms, are legally required to remove content. The DSA abstains from the difficult task of drawing a line between legal and illegal content. Instead, it defines a number of liability exemptions by establishing when a provider of online

---

72. Netzwerkdurchsetzungsgesetz [NetzDG] [Network Enforcement Act], Sept. 1, 2017, BUNDESGESETZBLATT, Teil I [BGBL I] at 1352, last amended by Gesetz [G], June 3, 2021, BGBL I art. 1, at 1436, (Ger.), https://www.gesetze-im-internet.de/netzdg/BJNR335210017.html.

73. Loi 2020-766 du 24 juin 2020 visant à lutter contre les contenus haineux sur internet [Law 2020-766 of June 24, 2020 Aimed at Fighting Hateful Content on the Internet], Journal Officiel de la République Française [J.O.] [Official Gazette of France], June 25, 2020, p. 1, https://www.legifrance.gouv.fr/eli/loi/2020/6/24/JUSX1913052L/jo/texte.

74. KOMMUNIKATIONSPLATTFORMEN-GESETZ [KOPI-G] [COMMUNICATION PLATFORMS ACT] Sept. 3, 2022 BUNDESGESETZBLATT, Teil I [BGBL I] No. 151/2020, https://www.ris.bka.gv.at/GeltendeFassung.wxe?Abfrage=Bundesnormen&Gesetzesnummer=20011415 (Austria).

75. *See, e.g.*, Martin Eifert et al., *Taming the Giants: The DMA/DSA Package*, 58 COMMON MKT. L. REV. 987 (2021).

76. *See* DSA, *supra* note 17, arts. 3–9, at 46–49.

intermediary services *cannot* be held liable in relation to third-party content. The safe harbor regime of the DSA broadly follows the existing model of the E-Commerce Directive.[77]

The DSA defines liability exemptions based on an intermediary's specific functions: mere conduit,[78] caching,[79] and hosting.[80] For the first two categories of providers, the DSA establishes a broad liability exemption as long the providers are "in no way involved with the information transmitted."[81] In contrast, for hosting providers, legal immunity is based on a knowledge standard. Unlike Section 230 of the Communications Decency Act,[82] which gives absolute immunity to publishers (other than immunity from Federal criminal law), the DSA shields hosting providers from liability only if they do not know that they are hosting illegal content. This general rule is supplemented by two important limitations: First, Member states must not impose a general obligation to monitor content on providers.[83] Second, it incorporates a Good Samaritan rule, whereby providers who carry out self-initiated investigations in order to detect and remove illegal content will not lose their liability exemption for this reason alone.[84]

The above liability rules are complemented by a number of due diligence obligations for online intermediaries. Because of the broad range of intermediary services covered by the new rules, the DSA does not apply a one-size-fits-all approach. Instead, the proposal follows a "risk-based

---

77. *See* Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on Certain Legal Aspects of Information Society Services, in Particular Electronic Commerce, in the Internal Market (Directive on Electronic Commerce), 2000 O.J. (L 178) 1, 4, arts. 12–13; see also Sebastian Felix Schwemer et al., *Liability Exemptions of Non-Hosting Intermediaries: Sideshow in the Digital Services Act?*, 8 OSLO L. REV. 4, 27 (2021) (discussing the liability exemptions for non-hosting providers under the ECD).

78. DSA, *supra* note 17, art. 3, at 46.

79. *Id.* art. 4, at 46.

80. *Id.* art. 5, at 47.

81. *Id.* recital 21, at 22.

82. 47 U.S.C. § 230.

83. DSA, *supra* note 17, art. 7, at 47.

84. *Id.* art. 6, at 47.

approach" and formulates a catalogue of "asymmetric due diligence obligations."[85] In doing so, the proposal distinguishes four levels of regulation:[86]

- *Level 1* contains a basic set of rules, which applies to the broadest category, i.e. all providers of online intermediary services. This category includes all providers of mere conduit, caching and hosting services.[87]
- *Level 2* applies to all providers of hosting services, such as cloud storage providers and webhosting services.[88]
- *Level 3* contains some additional provisions that apply only to online platforms,[89] defined as a provider of hosting services which, at the request of a recipient of the service, stores and disseminates to the public information.[90] This category includes online marketplaces, social networks and app stores.
- *Level 4* adds some specific due diligence obligations for "very large online platforms" (VLOPs),[91] which have more than 45 million average monthly users (i.e. roughly 10 percent of EU citizens). In contrast, very small platforms are exempt from most due diligence obligations.[92]

---

85. *Proposal for a Regulation of the European Parliament and of the Council on Contestable and Fair Markets in the Digital Sector (Digital Markets Act)*, at 6, COM (2020) 842 final (Dec. 15, 2020).

86. The final version of the DSA could see even further differentiations. In its General Approach of November 18, 2021, the Council of the European Union has suggested to add specific due diligence obligations for two additional categories of service providers, i.e., online search engines and online marketplaces. *General Approach of the Council of the European Union on the Proposal for a Regulation of the European Parliament and of the Council on a Single Market for Digital Services (Digital Services Act) and Amending Directive 2000/31/EC*, Doc. 13203/21, recital 56, at 60 (Nov. 18, 2021), https://data.consilium.europa.eu/doc/document/ST-13203-2021-INIT/en/pdf.

87. *See* DSA, *supra* note 17, arts. 10–13, at 49–51.

88. *See id.* arts. 14–15, at 51–52.

89. *See id.* arts. 16–24, at 53–59.

90. *See id.* art. 2(h), at 45.

91. *See id.* arts. 25–33, at 59–64.

92. *See, e.g., id.* art. 16, at 53.

This model of asymmetric regulation, under which the rules become more numerous and stricter as we go from Level 1 to Level 4, is an expression of the principle of proportionality. In terms of substance, the DSA introduces, for the first time, a number of due diligence obligations with regard to content moderation. In particular, hosting providers must provide harmonized notice-and-action mechanisms and justify removal decisions with a statement of reasons.[93] In addition, platforms must provide users with meaningful possibilities to challenge decisions to remove or label content via an internal complaint system and an external out-of-court dispute resolution mechanism.[94] And VLOPs are subject to additional rules to ensure more comprehensive public oversight of their content moderation practices. In this sense, VLOPs are obliged to develop appropriate tools for assessing and managing systemic risks and take measures to protect the integrity of their services against manipulation, including disinformation campaigns or interference with electoral processes.[95]

Most of the rules regarding content moderation in the DSA have been drafted with user-facing platforms such as Facebook, Twitter, or YouTube in mind. But as the following two sections explain, a closer examination reveals that the DSA responds to the expansion of content moderation practices by expanding its content moderation rules, both in the horizontal and vertical dimensions.

2. Horizontal Expansion: Messaging Services as Social Media Platforms

The first extension of the regulatory framework is horizontal and takes into account that content moderation is relevant not only for social media platforms, but also for other actors on the application layer of the Internet. As explained above, there has been a recent shift from social media platforms to messaging services as distribution channels for harmful and illegal content.[96] In particular, groups and channels on messaging services that apply a more lenient policy towards content moderation, such as

93. *Id.* arts. 14–15, at 51–52.

94. *Id.* arts. 17–18, at 53–54.

95. *Id.* arts. 25–33, at 59–64.

96. *See supra* text accompanying notes 22–27; *see also* Matthias C. Kettemann & Martin Fertmann, *Viral Information: How States and Platforms Deal with Covid-19-Related Disinformation: An Exploratory Study of 20 Countries* 8–9 (GDHRNet, Working Paper No. 1, 2020).

Telegram or Viber, have become fora for COVID-19-related disinformation.[97]

This development raises the question of whether messaging services can be considered as online platforms (or, as the case may be, even very large online platforms) for the purposes of the DSA. At first glance, the answer seems to be negative, as its definition of "online platforms" only covers intermediary services which disseminate information "to the public."[98] And "dissemination to the public" means making information available, at the request of the recipient of the service who provided the information, to a potentially unlimited number of third parties.[99] Therefore, as a matter of principle, "interpersonal communication services"[100] (for example, email or private messaging services) fall outside the scope of the DSA.[101] But Recital 14 creates a certain degree of legal uncertainty because "the mere possibility to create groups of users of a given service should not, in itself, be understand to mean that the information disseminated in that manner is not disseminated to the public";[102] only "closed groups" which consist of a "finite number of pre-determined persons" shall be excluded from the DSA.[103]

In an effort to clarify the applicability of the DSA to messaging services, the Council of the European Union suggested to amend Recital 14 as follows: "where access to information requires registration or admittance to a group of recipients of the service, that information should be considered to be disseminated to the public only where recipients of the service seeking to access the information are automatically registered or admitted without

97. Kettemann & Fertmann, *supra* note 96, at 8–9.

98. DSA, *supra* note 17, art. 2(h), at 45.

99. *Id.* art. 2(i), at 45.

100. Directive (EU) 2018/1972 of the European Parliament and the Council of 11 December 2018 Establishing the European Electronic Communications Code (Recast), 2018 O.J. (L 321) 1, 36.

101. DSA, *supra* note 17, recital 14, at 21.

102. *Id.*

103. *Id.*

a human decision or selection of whom to grant access."[104] The proposed amendment also provides that the DSA "may apply" to "public groups or open channels" provided by messaging services.[105] We will have to wait and see how the scope of the DSA is ultimately tailored. But it is certainly apparent that the European legislature is making efforts to extend the content moderation rules at least to some features of messaging services.

3. Vertical Expansion: Moving Down the Content Moderation Stack

While the horizontal extension of content moderation rules to messaging services may seem like a rather small change, the DSA extends the scope of its due diligence obligations for content moderation much farther in the vertical dimension. Chapter III of the DSA sets out a number of due diligence obligations for providers of online intermediary services, which include specifications for content moderation.[106] As already mentioned, the DSA differentiates the due diligence obligations according to risk and size. Accordingly, requirements of varying degrees of detail apply to the different levels of infrastructure moderation. As the due diligence requirements imposed by the DSA increase from the bottom to the top of the Internet stack, the following Subparts will discuss its impact on infrastructure moderation from a bottom-up perspective, starting with "technical auxiliary functions"[107] and then moving up the stack, from hosting providers to app stores.

a. Technical Auxiliary Functions

Recital 27 acknowledges that since the enactment of the Electronic Commerce Directive in 2000, "new technologies have emerged that improve the availability, efficiency, speed, reliability, capacity and security"[108] of Internet services. Such services, which facilitate the proper functioning of the Internet at the infrastructure level (technical auxiliary

104. *General Approach of the Council of the European Union on the Proposal for a Regulation of the European Parliament and of the Council on a Single Market for Digital Services (Digital Services Act) and Amending Directive 2000/31/EC*, Doc. 13203/21, recital 14, at 18 (Nov. 18, 2021), https://data.consilium.europa.eu/doc/document/ST-13203-2021-INIT/en/pdf.

105. *Id.*

106. DSA, *supra* note 17, arts. 10–37, at 49–67.

107. *Id.* recital 27, at 23–24.

108. *Id.*

functions), include DNS services and top-level domain name registries, registrars,[109] and CDNs. Recital 27 also underlines that providers of such infrastructure services can also benefit from the liability exemptions set out elsewhere in the DSA "to the extent that they qualify as 'mere conduits', 'caching' or 'hosting' services."[110] But it is doubtful whether the services related to the DNS and CDNs fit into one of these three categories.[111] This, of course, creates uncertainty regarding the extent to which providers of DNS-related services are shielded from liability. To avoid this ambiguity, it might be sensible for the final version of the DSA to expressly specify that DNS services and CDNs also benefit from the liability exemptions.[112]

As mentioned above, the DSA provides for a graduated regime of due diligence obligations for different categories of digital service providers, depending on the nature of their services and their size (e.g., intermediary services, hosting services, online platforms and very large online platforms). Under this "asymmetric approach,"[113] providers of basic services at the infrastructure level only have to comply with a limited set of obligations set out in Articles 10 to 13 DSA.

In particular, all providers of intermediary services, including those fulfilling technical auxiliary functions, have to designate a single

---

109. The Commission Proposal of 15 December 2020 only mentions DNS services, TLD registries and CDNs, but the General Approach adopted by the Permanent Representatives of the Member States on 12 November 2021 adds a reference to registrars. *Id.* recital 83, at 38.

110. *Id.* recital 27, at 23–24.

111. Schwemer et al., *supra* note 77, at 27 (arguing that DNS operators and CDNs neither transmit data in a communications network nor provide access to these).

112. *See Draft Report of the Committee on the Internal Market and Consumer Protection of the European Parliament on the Proposal for a Regulation of the European Parliament and of the Council on a Single Market for Digital Services (Digital Services Act) and Amending Directive 2000/31/EC, Amendments 183–376*, amend. 308, at 116 (July 8, 2021), https://www.europarl.europa.eu/doceo/document/IMCO-AM-695150_EN.pdf (suggesting to add a the following sentence at the end of Recital 27 of the DSA: "Domain name system (DNS) registration services can also benefit from the exemptions from liability set out in this Regulation."); *see also* Schwemer et al., *supra* note 77, at 27 (suggesting that a specific liability exemption for "auxiliary network intermediaries" should be included in the DSA).

113. DSA, *supra* note 17, at 11.

(electronic) point of contact allowing for direct communication with Member States' authorities, the European Commission, and the Board of Digital Service Coordinators which will coordinate the enforcement of the DSA.[114] Moreover, service providers from outside the European Union have to designate a legal representative in one of the EU Member States.[115] This representative is not merely a "mailbox" for communications by national enforcement agencies, but can be held liable for non-compliance with the obligations under the DSA.[116] And Article 10 is closely linked to Article 1(3), pursuant to which the DSA will apply to intermediary services provided to recipients in the European Union, regardless of the jurisdiction from which the service is provided.[117] Similar obligations to designate a "representative" or a "responsible person" have been stipulated in other EU laws with an extra-territorial reach, such as the General Data Protection Regulation,[118] the Market Surveillance Regulation,[119] and the recent proposal for a General Product Safety Regulation.[120]

With regard to content moderation at the infrastructure level, one of the most interesting provisions is Article 12, pursuant to which all providers of intermediary services must:

> [I]nclude information on any restriction that they impose . . . to the use of their service in respect of information provided by the recipients of the service, in their terms and conditions. That information shall include

---

114. *Id.* art. 10(1), at 49.

115. *Id.* art. 11(1), at 49.

116. *Id.* art. 11(3), at 50.

117. *Id.* art. 1(3), at 43.

118 Regulation 2016/679, of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119), 48–49.

119 Regulation 2019/1020, of the European Parliament and of the Council of 20 June 2019 on Market Surveillance and Compliance of Products and Amending Directive 2004/42/EC and Regulations (EC) No 765/2008 and (EU) No 305/2011, 2016 O.J. (L 169), 14.

120 European Commission, *Proposal for a Regulation of the European Parliament and of the Council on General Product Safety, Amending Regulation (EU) No 1025/2012 of the European Parliament and of the Council, and Repealing Council Directive 87/357/EEC and Directive 2001/95/EC of the European Parliament and of the Council*, at 42, COM (2021) 346 final (June 30, 2021).

information on any policies, procedures, measures and tools used for the purpose of content moderation, including algorithmic decision-making and human review. It shall be set out in clear and unambiguous language and shall be publicly available in an easily accessible format.[121]

While this provision seems to be primarily aimed at service providers at the application level of the Internet, such as social media platforms, it also applies to service providers at the infrastructure level. Therefore, Article 12(1) also requires providers of infrastructure services to be more transparent about their content moderation policies and procedures. From a practical perspective, this will mean that providers of DNS services and CDNs who offer their services in the EU will have to adapt their Acceptable Use Policies.

The transparency requirements set out in Article 12 are supplemented by Article 13, which stipulates detailed reporting obligations.[122] In particular, it requires providers of intermediary services to publish a yearly report on any content management they engaged in during the relevant time period.[123] The report shall not only provide information about orders received from Member States' authorities to act against illegal content,[124] or to provide information about specific individual recipients of intermediary services,[125] but also about any content moderation the service provider engaged in at its own initiative.[126] Already, many user-facing platforms such as Facebook and Twitter publish periodic transparency reports, which

---

121. DSA, *supra* note 17, art. 12, at 50. The General Approach adopted by the Permanent Representatives of the Member States on 18 November 2021 adds that the terms and conditions must be made available also in a "machine-readable" format. *General Approach of the Council of the European Union on the Proposal for a Regulation of the European Parliament and of the Council on a Single Market for Digital Services (Digital Services Act) and Amending Directive 2000/31/EC*, Doc. 13203/21, Article 12(1), at 125 (Nov. 18, 2021), https://data.consilium.europa.eu/doc/document/ST-13203-2021-INIT/en/pdf.

122. "[P]roviders of intermediary services that qualify as micro or small enterprises within the meaning of . . . Recommendation 2003/361/EC" are exempt from the reporting obligations. DSA, *supra* note 17, art. 13(2), at 51.

123. *Id.* art. 13(1), at 50.

124. *See id.* art. 8, at 47–48.

125. *See id.* art. 9, at 48–49.

126. *Id.* art. 13(1)(c), at 50–51.

typically disclose aggregate data about their content moderation practices.[127] These reports, however, vary greatly in detail.[128]

While Articles 12(1) and 13 merely set out transparency and reporting requirements regarding content moderation policies, Article 12(2) goes one step further and stipulates how providers of infrastructure services shall implement their content moderation procedures:

> Providers of intermediary services shall act in a diligent, objective and proportionate manner in applying and enforcing the restrictions referred to in paragraph 1, with due regard to the rights and legitimate interests of all parties involved, including the applicable fundamental rights of the recipients of the service as enshrined in the Charter.[129]

First, Article 12(2) sets a benchmark for content moderation decisions at all levels of the content moderation stack. All service providers are obliged to take due consideration of fundamental rights enshrined in the EU Charter of Fundamental Rights (CFR),[130] particularly the freedom of expression and information, the freedom to conduct a business, and the right to non-discrimination.[131]

Second, and more importantly, the reference to the principle of proportionality in Article 12(1) could provide some guidance as to which layer of the content moderation stack content-related decisions should be made. Since content moderation decisions at the infrastructure level are typically broader and more severe, decisions about problematic content

---

127. *See* Daphne Keller & Paddy Leerssen, *Facts and Where to Find Them: Empirical Research on Internet Platforms and Content Moderation*, *in* SOCIAL MEDIA AND DEMOCRACY: THE STATE OF THE FIELD AND PROSPECTS FOR REFORM 220, 228 (Nathaniel Persily & Joshua A. Tucker eds., 2020) (providing an overview of reporting practices by major platforms).

128. *See, e.g.*, Nicolas P. Suzor et al., *What do We Mean When We Talk About Transparency? Toward Meaningful Transparency in Content Moderation*, 13 INT'L J. COMMC'N. 1526 (2019).

129. DSA, *supra* note 17, art. 12(2), at 50.

130. *See* Charter of Fundamental Rights of the European Union, 2000 O.J. (C 364) [hereinafter CFR].

131. DSA, *supra* note 17, art. 12(2), at 50. *See also id.* recital 3, at 18; CFR, *supra* note 130, arts. 11, 16, 21, at 11–13.

should preferably be made higher up in the stack.[132] This idea is also reflected in Recital 26, according to which any requests or orders for content removal by online intermediaries "should, as a general rule, be directed to the actor that has the technical and operational ability to act against specific items of illegal content, so as to prevent and minimise any possible negative effects for the availability and accessibility of information that is not illegal content."[133]

b. Hosting Providers

The general requirements for online intermediaries explained above also apply to web hosting and cloud storage providers. This means, for example, that providers like AWS or Microsoft Azure must provide, in their terms and conditions, detailed information on their content moderation policies.[134] It is doubtful that AWS's current, rather brief, explanations in the Acceptable Use Policy will meet these requirements.[135] As explained above, additional obligations also arise from Article 13, which requires detailed reports on content moderation to be published at least once a year.

In addition to these general rules applicable to all providers of intermediary services, webhosting and cloud storage providers are subject to more stringent procedural requirements regarding content moderation. In particular, hosting providers must set up a user-friendly electronic notice-and-action mechanism that allows "any individual or entity to notify them of the presence on their service of specific items of information that the individual or entity considers to be illegal content."[136] If a hosting

---

132. *See also* Schwemer et al., *supra* note 77, at 27 (arguing that "more remote intermediaries should not be targeted or only be targeted as a last resort").

133. DSA, *supra* note 17, recital 26, at 23.

134. *See id.* art. 12(1), at 50.

135. *See* AWS Acceptable Use Policy, AMAZON WEB SERVS., INC., https://aws.amazon.com/de/aispl/aup/ (July 14, 2021).

136. DSA, *supra* note 17, art. 14(1), at 51. Some hosting providers already offer ways to report illegal content. For example, Microsoft Azure allows users to submit an "abuse report." *See* Submit Abuse Report (CERT), MICROSOFT, https://msrc.microsoft.com/report/abuse (last visited Feb. 24, 2022). A similar facility for reporting abuses is available for AWS. *See* Report Amazon AWS Abuse, AMAZON WEB SERVS., INC., https://support.aws.amazon.com/#/contacts/report-abuse (last visited Feb. 24, 2022).

provider decides to remove or disable access to the illegal content, Article 15 obligates the host toto provide the uploader of that content with a detailed statement of reasons for removal.[137]

The language of Articles 14 and 15 suggests that their drafters were thinking primarily of user-facing services at the application layer. For example, Article 15(1) refers to the removal of "specific items of information."[138] Yet cloud service providers can usually only block access to an entire service; they cannot remove individual pieces of content. Moreover, cloud data is often stored in an encrypted manner. Depending on the encryption method, only cloud customers will have access to the encryption keys, making it difficult or impossible for the cloud storage provider to locate the complained-of information. Obviously, applying the notice-and-action mechanism is likely to pose technical difficulties in such cases.

c. App Stores

Even more extensive regulations regarding infrastructure moderation will be applicable to app stores, which are likely to qualify as "online platforms" within the multi-level system of the DSA. Accordingly, app stores will be subject not only to the general rules for all intermediary services,[139] and for providers of hosting services,[140] but also the more stringent rules for online platforms.[141] In addition to being required to provide a user-friendly notice-and-action mechanism (as it is the case for webhosting and cloud storage providers), they are also obliged to set up an internal complaint-handling system which enables "recipients of the service" to lodge a complaint against decisions to remove content or terminate the service.[142] Furthermore, "recipients of the service" who are affected by the app store's decision must be given the opportunity to appeal

137. DSA, *supra* note 17, art. 15, at 52.

138. *Id.* art. 15(1), at 52.

139. *See id.* arts. 10–13, at 49–51.

140. *See id.* arts. 14–15, at 51–52.

141. *See id.* arts. 16–24, at 53–59.

142. *Id.* art. 17, at 53.

the decision to an impartial dispute resolution body whose decision is binding on the app store.[143]

Articles 14 and 15 do not clearly indicate whether such a complaint procedure must be made available only to an app developer whose app has been banned, or whether individual users affected by the decision to ban the app can lodge a complaint. Articles 14 and 15 grant the right to challenge the removal decision to the "recipient of the service," which is defined as "any natural or legal person who uses the relevant intermediary service."[144] It could well be argued that the app store's service is used both by app developers (for distribution) and by app users (for purchase and download). This broad interpretation of the right of complaint would also ensure that the rights of all affected parties would be taken into account. In any case, the lack of clarity about who can file a complaint if an app is banned from an app store suggests that the procedural rules of the DSA are not yet adequately tailored to cases of deep moderation by app stores.

Stricter requirements also apply to app stores with regard to the periodic transparency reports about content-related measures.[145] In particular, Article 23 requires app stores to provide information about "any use made of automatic means for the purpose of content moderation, including a specification of the precise purposes, indicators of the accuracy of the automated means in fulfilling those purposes and any safeguards applied."[146] While algorithmic content moderation is currently used primarily at the application level,[147] it cannot be ruled out that methods of automated content moderation will play a greater role at the infrastructure layer in the future.

---

143. *Id.* art. 18, at 53–55.

144. *Id.* art. 2(b), at 44.

145. *Id.* arts. 13, 23, at 50–51, 58.

146. *Id.* art. 23(1)(c), at 58.

147. *See, e.g.*, Robert Gorwa, Reuben Binns & Christian Katzenbach, *Algorithmic Content Moderation: Technical and Political Challenges in the Automation of Platform Governance*, BIG DATA & SOC'Y, Feb. 2020, at 1, https://journals.sagepub.com/doi/pdf/10.1177/2053951719897945.

Two of the most prominent app stores, Google Play and the Apple App Store, will probably also qualify as VLOPs,[148] for which the DSA provides even more extensive regulations.[149] Thus, these app stores will be required to undergo annual assessments to identify any significant systemic risks stemming from the services.[150] The scope of this risk assessment includes the dissemination of illegal content, any negative effects on the exercise of fundamental rights, and risks related to the protection of public health, electoral processes, or public security.[151] In order to counter such systemic risks, VLOPs shall take mitigation measures, e.g. by adapting their content moderation systems or their terms and conditions.[152] The effectiveness of these measures will be assessed at least annually by an independent audit.[153] Furthermore, VLOPs are required to publish transparency reports every six months, as opposed to once a year.[154]

B. Code of Practice on Disinformation

The trend towards an expansion of the regulatory framework for content moderation is not limited to the DSA Proposal. In particular, the current plans to revise the EU Code of Practice on Disinformation (CPD)[155] also aim to expand the existing self-regulatory framework in the direction of infrastructure moderation.

1. Background

148. According to Art. 25(1) DSA online platforms which provide their services to a number of average monthly active recipients of the service in the European Union equal to or higher than 45 million (i.e., roughly 10% of the EU population) are considered as VLOPs. DSA, *supra* note 17, art. 25(1), at 59.

149. *See id.* arts. 25–33, at 59–64.

150. *Id.* art. 26, at 59–60.

151. *Id.*

152. *Id.* art. 27, at 60–61.

153. *Id.* art. 28, at 61.

154. *Id.* art. 33, at 64. In particular, the report must provide details about the results of the risk assessment (Art. 26), the risk mitigation measures (Art. 27), and the independent audit (Art. 28). *Id.*

155. European Commission, *EU Code of Practice on Disinformation* (2018), https://digital-strategy.ec.europa.eu/en/policies/code-practice-disinformation (click "Code of Practice on Disinformation (.pdf)" on sidebar).

The CPD was unveiled in October 2018 as a centerpiece of the EU's fight against disinformation.[156] The CPD is conceived as a self-regulatory framework under which the signatories (including Facebook, Google, Twitter, TikTok, and several advertising industry associations) voluntarily commit to take measures to minimize the spread of online disinformation. In order to coordinate these efforts, the CPD contains a wide range of commitments: from transparency requirements for political advertising, to the termination of fake accounts, to the demonetization of disinformation campaigns.[157] The CPD defines disinformation as "verifiably false or misleading information" which: (a) "is created, presented and disseminated for economic gain or to intentionally deceive the public"; (b) "may cause public harm"; and is intended as "threats to democratic political and policymaking processes as well as public goods such as the protection of EU citizens' health, the environment or security."[158] Thus, the CPD has a much broader scope than the DSA: While the DSA only addresses the responsibility of online intermediaries for "illegal content,"[159] the CPD aims at tackling "harmful content," which may or may not be illegal (for example, disinformation regarding COVID-19).[160]

However, there is a dual relationship between the CPD and the DSA. The first concerns the terms of use of the online service providers. If a provider prohibits certain types of disinformation content in its terms of use, such content is considered "illegal content" under the DSA.[161] As a result, the rules of the DSA must be observed when the provider takes measures against the disinformation content. The second link is even stronger, and concerns VLOPs. As explained above, VLOPs have to take

156. *See, e.g.*, Joris van Hoboken & Ronan Ó Fathaigh, *Regulating Disinformation in Europe: Implications for Speech and Privacy*, 6 U.C. IRVINE J. INT'L, TRANSNAT'L, & COMPAR. L. 9, 12–16 (2021) (explaining the political background of the CPD).

157. European Commission, *supra* note 155, § 1.

158. European Commission, *supra* note 154, preamble, at 1 (internal quotation marks omitted).

159. *See* DSA, *supra* note 17, arts. 2(g), 2(p), at 45.

160. *See* European Commission, *supra* note 155, preamble, at 3.

161. *See* DSA, *supra* note 15, art. 2(g), at 45.

measures to mitigate "systemic risks".[162] One of the risk mitigation measures mentioned in the DSA is cooperation with other online platforms through codes of conduct[163]—such as the CPD. And an online platform's unexplained refusal to participate in such a code of conduct can be a factor relevant to determining whether the online platform has infringed its obligations under the DSA.[164] In other words, the DSA creates a strong incentive for VLOPs (including major app stores) to become signatories to the CPD.[165] Through this interplay with the binding rules of the DSA, the voluntary rules of the CPD will be upgraded from a self-regulatory regime to a co-regulatory regime.

In May 2021, the European Commission published its "Guidance on Strengthening the Code of Practice on Disinformation," which provides an ambivalent assessment of the CPD.[166] On the one hand, the Commission praised the CPD as a "substantial, first-of-its-kind achievement."[167] But it also acknowledged that the CPD has significant shortcomings,[168] including "limitations intrinsic to the self-regulatory nature of the Code, as well as gaps in the Code's commitments."[169] The Guidance made various suggestions on how these deficits could be overcome.[170] The next two

---

162. *See supra* text accompanying notes 147–153.

163. DSA, *supra* note 17, art. 27(1)(e), at 60.

164. *Id.* recital 68, at 35.

165. *See* Ruairí Harrison, *Tackling Disinformation in Times of Crisis: The European Commission's Response to the Covid-19 Infodemic and the Feasibility of a Consumer-Centric Solution*, UTRECHT L. REV., Oct. 2021, at 18, 29.

166 European Commission, *supra* note 66.

167. *Id.* at 1.

168. *See Commission Staff Working Document: Assessment of the Code of Practice on Disinformation - Achievements and Areas for Further Improvement*, at 7, SWD (2020)180 final (Sept. 10, 2020).

169. European Commission, *supra* note 66, at 1.

170. A comprehensive analysis of the proposed CPD revisions is, of course, beyond the scope of this Article. And since such analysis has been undertaken by other scholars, it would also be unnecessary. *See, e.g.*, Harrison, *supra* note 162 (providing a critical assessment of the planned shift from self-regulation to co-regulation); *see also* Ethan Shattock, *Self-Regulation 2:0? A Critical Reflection of the European Fight Against Disinformation*, HARV. KENNEDY SCH. MISINFORMATION REV. (May 31, 2021),

Subparts will focus on the Commission's plan to expand the scope of the CPD and promote cooperation between different players in the content moderation ecosystem. Notably, the CPD encourages the expansion of content moderation at the infrastructure level, particularly regarding the online advertising and financial services industries. It also actively promotes cooperation between actors in their content moderation activities.

2. Demonetizing Disinformation

In its 2021 Guidance, the European Commission set out its expectations regarding the revision of the CPD and calls for stronger commitments by the signatories, and for broader participation to the Code. One area where the European Commission found shortcomings in the Code of Practice is the demonetization of disinformation.[171] In particular, the Commission criticized that "online advertisements still continue to incentivize the dissemination of disinformation," and urged online platforms and other players of the online advertising ecosystem to "take responsibility and work together to defund disinformation."[172]

To increase the Code's impact on the demonetization of disinformation, the Commission called for broader participation from the online advertising ecosystem (including ad exchanges, ad-tech providers, communication agencies, and even brands with a substantial online advertising budgets).[173] Clearly, the Commission wants to bring together a broad-based coalition of the willing to join forces in the fight against disinformation. Also included are players providing services that may be used for monetizing disinformation (such as online payment services, e-commerce platforms, crowd-funding platforms, and donation systems).[174]

With regard to online advertising, the Commission's Guidance called on signatories of the Code " to improve transparency and accountability

---

https://misinforeview.hks.harvard.edu/article/self-regulation-20-a-critical-reflection-of-the-european-fight-against-disinformation/ (criticizing the changes suggested by the European Commission as "vaguely framed, and fail[ing] to address critical issues").

171. European Commission, *supra* note 66, at 2.

172. *Id.*

173. *Id.* at 5–6.

174. *Id.* at 6.

around ad placements".[175] For this purpose, ad-tech companies and other actors in the online advertising industry would adopt measures to avoid placing advertisements next to disinformation content or on websites that are notorious for spreading disinformation.[176] Advertisers may do this by removing disinformation ads or disabling advertising accounts. Similarly, brand owners should commit to take measures to avoid the placement of their advertising in a context of disinformation.[177] Considering that online ads are a primary monetization strategy for websites spreading disinformation,[178] this approach could indeed prove effective.

While the Commission's plans regarding advertising are quite specific, the Guidance is less clear on how providers of financial services, such as payment processing services, crowd-funding platforms, and donation systems, should contribute to the fight against disinformation. The Guidance emphasizes that a revised CPD should include "tailored commitments that correspond to the diversity of services provided by signatories and the particular roles they play in the ecosystem."[179] In a sense, this echoes the calls for a "layer-conscious approach" to content moderation.[180] In any case, the Commission does not call for outright "financial deplatforming" of disinformation sites.[181] As part of the revision of the CPD, it will be necessary to develop appropriate and proportionate tools tailored to the role of financial service providers.

3. Content Cartels

In addition to its call for broader participation in the CPD, the European Commission seeks to promote "close cooperation of different

---

175. *Id.* at 7.

176. *Id.*

177. *Id.* at 7–8.

178. *See* Han et al., *supra* note 54, at 6.

179. European Commission, *supra* note 66, at 6.

180. *See* Bridy, *supra* note 13.

181. *See* European Commission, *supra* note 155, at 3 ("Signatories should not be compelled by governments, nor should they adopt voluntary policies, to delete or prevent access to otherwise lawful content or messages solely on the basis that they are thought to be 'false'.").

players," particularly in the online advertising ecosystem,[182] to facilitate information sharing across the advertising value chain in order to identify purveyors of disinformation. As one concrete example of such a cross-platform cooperation, the European Commission suggests creation of a "common repository" of rejected disinformation ads.[183] Such a database would facilitate the exchange of information about ads considered to be disinformation, ensuring that ads banned on one platform are prevented from appearing on other platforms. Such a shared database could build on existing initiatives in the advertising industry. In June 2019, the World Federation of Advertisers launched the "Global Alliance for Responsible Media," which aims to develop a set of industry-wide definitions and standards regarding different types of harmful online content.[184]

In accordance with the European Commission's "follow-the-money" approach, the cross-industry cooperation should not be limited to the online advertising industry; it should include other "players active in the online monetization value chain, such as online e-payment services, e-commerce platforms, and relevant crowd-funding/donation systems".[185]

But, while the cooperation of different actors may increase the effectiveness of individual measures against disinformation, it could lead to the creation of a new type of "content cartel"; that is, "arrangements between platforms to work together to remove content or actors from their services without adequate oversight."[186] Indeed, broad cross-industry cooperation by actors in different parts of the content moderation ecosystem could compound the existing lack of accountability in content moderation. As one commentator has observed: "Content cartels exacerbate [the lack of accountability]—when platforms act in concert, the

182. European Commission, *supra* note 66, at 8.

183. *Id.*

184. *Global Alliance for Responsible Media*, WORLD FED'N ADVERTISERS, https://wfanet.org/leadership/garm/about-garm (last visited Mar. 9, 2022).

185. European Commission, *supra* note 66, at 8.

186. Evelyn Douek, *The Rise of Content Cartels,* KNIGHT FIRST AMEND. INST. (Feb. 11, 2020), https://knightcolumbia.org/content/the-rise-of-content-cartels; *see also* Gillespie et al., *supra* note 11, at 7 (warning that the acting in concert by different service providers may reduce the diversity of viewpoints in the entire information environment).

actual source of any decision is harder to identify and hold to account."[187] And since the Commission's proposal would expand cooperation into the (financial) infrastructure layer, it illustrates the phenomenon of "content cartel creep."[188] While earlier examples of cross-platform cooperation concerned clear cases of illegal content, such as child pornography or terrorist content,[189] the planned revision of the CPD ventures into the uncertain terrain of content that is deemed harmful, but is not necessarily illegal.

III. TOWARDS SANTA CLARA PRINCIPLES FOR INFRASTRUCTURE MODERATION?

In 2018, a group of academics and civil society organizations developed the Santa Clara Principles on Transparency and Accountability in Content Moderation.[190] These principles describe best practices for obtaining meaningful transparency and accountability for content moderation that user-facing platforms could voluntarily adopt. Notably, the principles define requirements regarding transparency and procedural due process for content removals and account suspensions. In December 2021, a second edition of the principles (Santa Clara Principles 2.0) was issued, which strengthens due process requirements and provides more elaborated guidelines for reporting on and notifying users about takedowns.[191]

The Santa Clara Principles have been drafted for moderation of user-generated content at the application level of the Internet. At the time of

187. Douek, *supra* note 186.

188. *Id.*

189. *See, e.g.*, Hany Farid, *Reining in Online Abuses*, 19 TECH. & INNOVATION 593 (2018) (discussing the use of the digital fingerprinting technology PhotoDNA in the fight against these categories of illegal content).

190. ACLU Found. N. Cal. et al., *Santa Clara Principles 1.0*, SANTA CLARA PRINCIPLES ON TRANSPARENCY & ACCOUNTABILITY IN CONTENT MODERATION, https://santaclaraprinciples.org/scp1/ (last visited Mar. 9, 2022).

191. Access Now et al., *Santa Clara Principles 2.0*, SANTA CLARA PRINCIPLES ON TRANSPARENCY & ACCOUNTABILITY IN CONTENT MODERATION, https://santaclaraprinciples.org (last visited Mar. 9, 2022); *see also EFF, Partners Launch New Edition of Santa Clara Principles, Adding Standards Aimed at Governments and Expanding Appeal Guidelines*, ELEC. FRONTIER FOUND. (Dec. 8, 2021), https://www.eff.org/press/releases/eff-partners-launch-new-edition-santa-clara-principles-adding-standards-aimed.

writing, there is no such set of principles for infrastructure moderation.[192] However, the expansion of content moderation into the infrastructure layer of the Internet described in Part I has shown that there is an urgent need for the elaboration of principles tailored to the specifics of infrastructure moderation. When attempting to develop principles for content-based decisions at the infrastructure level, the Santa Clara Principles, with their focus on transparency and due process, can serve as a model. However, it is important to note that principles which have been developed for the application layer cannot be applied one-to-one to the infrastructure layer. Rather, principles for content moderation at the infrastructure layer should be based on a "layer-conscious approach"[193] that takes into account the different possibilities and limitations of the various actors.

Such principles—possibly a version 3.0 of the Santa Clara Principles—could provide guidance for players engaged in infrastructure moderation as well as for the future development of the regulatory framework. As discussed in Part II has shown, the development of such an extended regulatory framework is already underway. Even if the current European regulatory proposals discussed in Part II are by all means less than perfect, they can provide some valuable starting points for the development of principles for infrastructure moderation. Building on the foregoing analysis, this final Part describes three key elements of a responsible infrastructure moderation framework: subsidiarity, transparency, and procedural safeguards.

A. Subsidiarity

Infrastructure moderation should be subsidiary to content moderation at the application level. The subsidiarity of infrastructure moderation follows from the different roles played by providers at the different levels of the Internet stack. It is also closely linked with the principle of proportionality. As we have seen, actors at the infrastructure layer are

192. *But see* Corynne McSherry, *Content Moderation Beyond Platforms: A Rubric*, TECHDIRT: TECH POL'Y GREENHOUSE (Sept. 23, 2021, 1:35 PM), https://www.techdirt.com/articles/20210923/12543647617/content-moderation-beyond-platforms-rubric/ (formulating seven guiding questions regarding the specific dangers of infrastructure moderation at various levels of the Internet stack).

193. *See* Bridy, *supra* note 13 (arguing for content-agnosticism at the infrastructure layer and content-awareness at the application layer).

usually unable to apply tailored and nuanced remedies aimed at individual content items[194]—they "only have a sledge hammer to deal with these questions, rather than a scalpel."[195] For example, a cloud service hosting a platform which is used for spreading illegal or harmful content may only have the option to deactivate access to the entire platform. Such an all-or-nothing decision to deactivate an entire platform can cause considerable collateral damage to perfectly legal and unobjectionable content.[196]

Accordingly, if measures to combat illegal or harmful content can be taken at the application level, action should not be taken at the infrastructure level. This does not, of course, mean that providers of infrastructure services should abstain from any content moderation.[197] Rather, infrastructure moderation has a different objective than content moderation at the application level. Infrastructure providers usually bear no responsibility for individual content items at the upper levels of the Internet stack. But they should care about whether actors at the upper levels engage in meaningful content moderation or not. Therefore, the responsibility of app stores, cloud providers, and other actors at the infrastructure level is better characterized as a sort of system responsibility that takes a systemic perspective on content moderation.

A focus on system responsibility entails that players further down in the stack can delegate the task of moderating individual content items to actors higher in the stack who are closer to the content itself.[198] In practical terms, this means that infrastructure providers should mainly limit

---

194. *See* Goldman, *supra* note 18, at 50.

195. Mike Masnick, *Welcome to the New Techdirt Greenhouse Panel: Content Moderation at the Infrastructure Level*, TECHDIRT: TECH POL'Y GREENHOUSE (Sept. 22, 2021, 12:00 PM), https://www.techdirt.com/articles/20210921/16345547609/welcome-to-new-techdirt-greenhouse-panel-content-moderation-infrastructure-level/.

196. *See* TARLETON GILLESPIE, CUSTODIANS OF THE INTERNET 176 (2018); *see also* Jack M. Balkin, *Free Speech Is a Triangle*, 118 COLUM. L. REV. 2011, 2016–18 (2018) (arguing that content moderation by providers of infrastructure services may lead to collateral censorship and has many features of prior restraint).

197. *But see* Balkin, *supra* note 12, at 73 (arguing that basic internet services should not engage in content regulation).

198. *See* Ben Thompson, *Moderation in Infrastruc*ture, STRATECHERY (Mar. 16, 2021), https://stratechery.com/2021/moderation-in-infrastructure/.

themselves to establishing minimum standards for content moderation which service providers at the application layer should apply. This task of setting standards for other content moderators could be also described as a sort of meta-moderation or second-order moderation. The minimum requirements for content moderation by third-party apps set out in the Apple's App Store Review Guidelines provide a good example of the form such second-order moderation could take.[199] Only if actors higher up in the stack fail to play their role as responsible content moderators should infrastructure providers themselves take action. This approach has also the advantage that it increases the consistency of content moderation policies and helps to avoid ad hoc decisions.

B. Transparency

Second, content moderation at the infrastructure level should be transparent. Transparency is probably the single most important element of responsible content moderation, both at the application layer and the infrastructure layer.[200] It is essential for accountability, non-discrimination, and proportionality, as well as for identifying potential conflicts of interest. Content moderation at the application level—which has long been notoriously opaque—is adopting transparency (though, perhaps, not with alacrity).[201] Facebook only published its Community Standards in 2018, after intense public criticism of its content moderation activities.[202] Similarly, when Facebook released the first edition of its Community Standards Enforcement Report (CSER), it was only the second social media

---

199. *App Store Review Guidelines*, APPLE: DEVELOPER, https://developer.apple.com/app-store/review/guidelines/ (Oct. 22, 2021).

200. *See* COUNCIL OF EUROPE STEERING COMMITTEE FOR MEDIA AND INFORMATION SOCIETY, CONTENT MODERATION: BEST PRACTICES TOWARDS EFFECTIVE LEGAL AND PROCEDURAL FRAMEWORKS FOR SELF-REGULATORY AND CO-REGULATORY MECHANISMS OF CONTENT MODERATION 46 (2021), https://rm.coe.int/content-moderation-en/1680a2cc18.

201. *See* Catherine Buni & Soraya Chemaly, *The Secret Rules of the Internet*, VERGE (Apr. 13, 2016), https://www.theverge.com/2016/4/13/11387934/internet-moderator-history-youtube-facebook-reddit-censorship-free-speech ("The details of moderation practices are routinely hidden from public view, siloed within companies and treated as trade secrets when it comes to users and the public.").

202. *See* Monika Bickert, *Publishing Our Internal Enforcement Guideline and Expanding Our Appeals Process*, META: NEWSROOM (Apr. 24, 2018), https://about.fb.com/news/2018/04/comprehensive-community-standards/.

platform to do so (after YouTube).[203] In the meantime, many other platforms have followed suit, increasing the transparency of content moderation practices at least to some degree.[204]

In contrast, at the infrastructure level, opacity continues to prevail. So far, few infrastructure service providers publish regular transparency reports about their content-related decisions.[205] As explained above, for providers who offer their services on the European market, the DSA will change this.[206] In the future, all categories of online intermediaries, regardless of their position in the content moderation stack, will be required to provide meaningful transparency, both regarding the criteria applied for decision making about content and their enforcement.[207] The relevant DSA provisions could be generalized and extended to other actors in the content moderation ecosystem (e.g. providers of financial services or advertising services).[208] In this sense, all relevant players engaging in content moderation could be required to publish clear, easily understandable, and sufficiently detailed explanations of their content moderation policies. Furthermore, they should be required to regularly publish reports on their content moderation activities.

C. Procedural Safeguards

Transparency alone is not enough to ensure responsible content moderation at the infrastructure layer. Just like at the platform level, infrastructure providers that engage in content moderation must embrace

---

203. Spandana Singh & Leila Doty, *The Transparency Report Tracking Tool: How Internet Platforms Are Reporting on the Enforcement of Their Content Rules*, NEW AM. (Dec. 9, 2021), https://www.newamerica.org/oti/reports/transparency-report-tracking-tool/.

204. *See id.*

205. *See, e.g.*, CLOUDFLARE, CLOUDFLARE TRANSPARENCY REPORT (2021), https://www.cloudflare.com/resources/assets/slt3lc6tev37/7yVWG0hLy1d9627Ia6PoeJ/fb348cc38d32b88df51f72f4e100f5e6/Transparency-Report-H2-2020.pdf (focusing on law enforcement requests received by the company regarding content removal and blocking).

206. *See supra* Part II.A.3.a.

207. *See supra* text accompanying notes 120–122.

208. *See also* Tusikov, *supra* note 10, at 51 (suggesting that payment service providers such as PayPal should publish transparency reports about their content moderation activities).

procedural safeguards that protect users of their services.[209] The design of these safeguards must take into account that blocking access to a website or an entire platform is more extreme than removing a single content item. Therefore, it may not be sufficient to provide for an appeal mechanism that allows users of infrastructure services to contest blocking decisions *ex post*. Rather, it seems more appropriate to establish *ex ante* procedural safeguards. From this perspective, a system of graduated response could help to ensure the principle of proportionality is respected when infrastructure providers engage in content moderation.[210]

Therefore, as the first step of a multi-step process that eventually may lead to a final removal, a warning should be issued. For example, in a court filing responding to Parler's motion for a temporary restraining order, Amazon argued that "AWS notified Parler repeatedly that its content violated the parties' agreement" and requested removal of content that threatened the public safety.[211] In a second step, a temporary blocking could be imposed to lend weight to the warning. Depending on the type of infrastructure service, other intermediate measures could be used. For example, payment service providers like Paypal and donation systems like Patreon could display warnings when a user wants to make a payment to a disinformation site operator. Similarly, instead of blocking a website, ISPs could choose to throttle the speed of certain services as part of a graduated response to violations of their acceptable use policy.[212] A permanent blocking should be considered only if these lesser steps are ineffective.

---

209. *See also* Bloch-Wehba, *supra* note 9, at 90 (emphasizing the importance of procedural safeguards for content moderation at the platform level).

210. *See* Peter K. Yu, *The Graduated Response*, 62 FLA. L. REV. 1373, 1426–27 (2010); *see also* Goldman, *supra* note 18, at 17–20 (discussing the "Graduated Response" scheme used to discourage copyright infringements).

211. Defendant Amazon Web Servs., Inc.'s Opposition to Parler LLC's Motion for Temporary Restraining Order at 2, *Parler LLC v. Amazon Web Servs.*, 514 F. Supp. 3d 1261 (W.D. Wash. 2021) (No. 2:21-cv-0031-BJR).

212. *See* Goldman, *supra* note 18, at 18–19 (citing examples where throttling Internet speed was used as a sanction under the Copyright Alert System).

Furthermore, just like at the platform level, decisions about content at the infrastructure level must follow general rules of procedural fairness.[213] This means that infrastructure providers must give reasoned explanations for their decisions and provide mechanisms for handling complaints. In this regard, the European Commission's proposal for a DSA falls short of the expectations for an effective regulatory framework for infrastructure moderation. While the DSA does contain detailed rules for internal complaint handling mechanisms and external out-of-court dispute settlement, these apply only to user-facing platforms, not to providers of intermediary services further down the Internet stack.[214] Thus, the DSA stops halfway in terms creating a balanced regulatory framework for infrastructure moderation.

CONCLUSION

Recent years have seen a robust debate—both among legal scholars and in the general public—regarding the moderation of online speech. So far, the debate has largely focused on user-facing platforms such as Facebook, Twitter, and YouTube. But these platforms are just the tip of the iceberg—there is a vast array of infrastructure providers on which user-facing platforms rely. It is only recently that scholarly attention has turned to content moderation at the infrastructure level.

Against this background, this Article makes several contributions to the discussion of infrastructure-level moderation: First, it explores the differences between moderation at the application layer and the infrastructure layer, and the various shapes and contexts of content moderation in different segments of the infrastructure ecosystem. Unlike content moderation at the application level, infrastructure moderation is usually not about individual items of illegal or objectionable content. In contrast, it is rather about meaningful moderation practices (or the lack thereof) at higher levels in the content moderation stack. In this sense, infrastructure moderation can be characterized as a sort of meta-moderation or second-order moderation. Second, the Article offers an analysis of recent regulatory developments in the European Union and shows that regulators are slowly adapting to the horizontal and vertical

---

213. *See* Balkin, *supra* note 12, at 93–94 (emphasizing the importance of procedural fairness regarding the enforcement of platform community standards); *see also* Rory Van Loo, *Federal Rules of Platform Procedure*, 88 U. CHI. L. REV. 829 (2021) (discussing due process requirements for dispute resolution mechanisms on digital platforms).

214. DSA, *supra* note 17, art. 16–17, at 53.

expansion of content moderation. In this sense, the expansion of content moderation to the infrastructure level is followed by an expansion of the regulatory framework for infrastructure moderation. Finally, the Article discusses how the Santa Clara Principles—a framework that was designed for content moderation decisions at the application layer—could be adapted and applied to content moderation at the infrastructure level.

It goes without saying that the DSA and revised CPD will not be the last word on infrastructure moderation. The next frontier is already on the horizon. As one might expect, the expansion of content moderation into the infrastructure layer of the Internet has providers of illegal or harmful content seeking ways to evade what they perceive as censorship. In this perspective, the decision of the alt-right platform Gab to shift to cryptocurrencies after being barred from Paypal and other payment processing services[215] might presage a more general trend that goes beyond decentralized finance. The expansion of infrastructure moderation could thus promote the creation of new decentralized means of distributing content online.[216] The decentralized architecture of a future Web 3.0 will bring new challenges for the regulatory design around content moderation.

215. *See supra* text accompanying note 63.

216. *See* Duffield, *supra* note 30.