Mitigating harmful content on social media via platform regulation: The Digital Services Act and content assessment

The Digital Services Act (DSA) implements a legal framework for digital services, including social media platforms, to ensure that they operate in responsible, accountable ways. With an objective involving three critical theoretical pillars—transparency, accountability, and responsibility—the DSA, among other functions, (mostly) holds online platforms liable for content that they publish and also imposes requirements that they mitigate and remove harmful content. However, from a critical standpoint, the DSA begs some pivotal questions. For one, how can such a legal document, even if binding, mitigate the severe societal, psychological, emotional, and even physical dangers and detriments experienced by victims of social media abuse? For another, how can a supranational regulation combat local disinformation campaigns and political propaganda? In this article, we encourage not only introducing, analyzing, and critically examining the DSA but also propose policy recommendations to ameliorate content moderation on social media platforms.

Keywords: *Digital Services Act, platform regulation, social media regulation, digital services, content moderation*

Acknowledgments

Gergely Gosztonyi was supported by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences. This article has been developed and published with the financial support of a fellowship from the Center for Advanced Internet Studies (CAIS) and the ADVANCED_24 funding scheme of the Ministry of Culture and Innovation of Hungary from the National Research, Development and Innovation Fund.

Author Information

Gergely Ferenc Lendvai, Pázmány Péter Catholic University, Faculty of Law
ORCID: 0000-0003-3298-8087, lendvai.gergely.ferenc@hallgato.ppke.hu
János Tamás Papp, Pázmány Péter Catholic University, Faculty of Law
ORCID: 0000-0001-8682-6900, papp.janos.tamas@jak.ppke.hu
Gergely Gosztonyi, Széchenyi István University, Faculty of Law
ORCID: 0000-0002-6551-1536, gosztonyi.gergely@sze.hu

How to cite this article:

Lendvai, Gergely Ferenc, János Tamás Papp, and Gergely Gosztonyi. "Mitigating harmful content on social media via platform regulation: The Digital Services Act and content assessment ". *Információs Társadalom* XXV, no. 2 (2025): 59–72.
https://dx.doi.org/10.22503/inftars.XXV.2025.2.4

1. Road to the Digital Services Act (DSA)

Social networking sites have become the dominant communication tools of the 21st century. As they continue to rapidly evolve, both technologically and in the number and types of users, it remains challenging to predict the direction that they will take. Legislation for such sites and social media platforms in general, as with all aspects of life, lags behind and rarely anticipates changes. Meanwhile, technology expands freedoms, and the law tries to catch up. Such trends are especially apparent regarding the regulation of networking sites, which billions of people now use on a daily basis. Indeed, those sites are the new city centers of the world and give each user their own set of soapboxes (Casey 2021, 33). The result has been problems that are difficult to address with legal regulation and now require an entirely new regulatory approach.

Legislating online harms on social networking sites and social media platforms is essential for a healthy democracy. The vast reach and influence of such sites and platforms have made them primary sources of information and interaction for billions of users. However, without proper regulation, they can quickly become conduits for the spread of harmful content, including disinformation, hate speech, and incitements to violence, all of which can have real-world repercussions. At the same time, they primarily operate based on self-regulation, which can lead to inconsistencies in content moderation, leave victims of virtual harms without effective redress, and generally allow the platforms to influence the course of public discourse as they see fit. At any time, a platform can decide that a particular type of opinion or an unpopular political group is somehow disadvantageous and consequently ban or filter them out, whether openly or without notice (Tutt 2014). That possibility creates an entirely new situation for the democratic public, because no guarantee exists that a private company with absolute control over a platform that hosts large volumes of public discourse will not significantly distort the public debate for its own interests (Harawa 2014, 396). That risk is particularly important because social networking sites provide some of the most convenient and arguably most effective ways to participate in social discourse. Regular users of social networking sites are also arguably far more politically active than others, meaning that any influence exerted on the platforms can have more direct political consequences than on other media platforms (Pew Research Center 2011).

There is a critical social need to not only preserve the freedom of social media platforms by permitting as little state influence as possible but also to regulate them in a reassuringly broad way and to curb their power to unduly influence the democratic public. Striking an appropriate balance between those two competing interests is complicated, however. If the internet is indeed an unregulated, free virtual world, then any attempt to regulate it entails the loss of certain freedoms. Regulating internet infrastructure can constitute a restriction on freedom of expression, even when not targeted at specific content or at what can and cannot be said. At present, it seems that not only legislators but also the platforms themselves are gesturing toward accepting regulation, even if their motivation remains unclear. Aside from any possible antidemocratic motives to restrict citizens' freedom of expression online,

which are necessarily incompatible with the ideals of free speech, most serious regulatory initiatives have sought to make the internet a safer, more enjoyable place for minors and adults alike. Despite the assumption that social media platforms think no differently, the primary motivation for regulatory intervention is perhaps more economical, for it is far more cost-effective to comply with a general set of rules than to comply with different laws in each country.

The current operation of online platforms is not free from specific legal regulation by the state but subject to several different rules. Platforms are partly covered by media regulations, electronic commerce law, contract and consumer protection law, data protections, and competition law, among others. However, the question of how to regulate social networking sites often raises the dilemma of whether local regulation at the national level or regional regulation at the level of the European Union (EU) is more appropriate. The law recognizes both forms of regulation because there is both national and transnational legislation regarding online platforms (Pillalamarri and Stanley 2021). However, one of the most significant legal challenges in the world of social media is precisely that the most popular platforms operate globally. Although most of those sites started and remain based in the United States (e.g., Facebook, YouTube, and X, formerly Twitter), the very nature of the internet means that they are available all over the world.

For that reason, social media platforms have to comply with an array of different, often conflicting, legal provisions in their operations. In recent years, the EU has recognized the dangers posed by online platforms and begun to employ various tools in response, including self-regulatory codes (e.g., the 2022 Code of Practice on Disinformation), directives (e.g., EU Directive 2017/541 on combating terrorism), and regulations (e.g., EU Regulation 2019/1150). However, the most significant piece of legislation—the real game-changer—has been the Digital Services Act (DSA; Gosztonyi 2025).

2. The Digital Services Act (DSA)

In view of content moderation-related polemics, especially in the context of digital safety, the European Commission issued a proposal in late 2020 titled "Shaping Europe's Digital Future." Among other things, it envisioned a comprehensive regulatory package on platform governance consisting of two major regulations: the Digital Markets Act and the DSA (Cauffman and Goanta 2021). Between them, the DSA amassed unprecedented attention from legal practitioners, stakeholders, and legal scholars owing to its innovative, holistic regulation of popular intermediary platforms such as Facebook, X (formerly Twitter), and TikTok. The core of the European regulation rests in a rather logical understanding of platform use—that very large online platforms (VLOPs), which have more than 45 million monthly active users in the EU (DSA, Article 33), wield extensive influence over public discourse and content dissemination and are therefore inherently and instrumentally accountable for and should be liable to impose effective mechanisms to provide safeguards against illegal and harmful content that may potentially reach millions of users (Papp 2022).

The principles to ensure that the platforms are indeed obliged to act against social media abuse stand on three critical pillars: transparency, accountability, and responsibility (Decarolis and Li 2023). However, even though the DSA represents a significant stride toward enhancing the digital landscape's safety and reliability, critical questions loom large. For one, how effectively can a legal document, albeit binding, mitigate the severe societal, psychological, emotional, or even physical dangers and detriments faced by victims of social media abuse? For another, can a supranational regulation effectively combat the localized scourges of disinformation campaigns and political propaganda? In this article, we examine the key obligations of the DSA and their effectiveness in the context of harms on social media.

2.1. DSA and the moderation of illegal content: A brief overview

As mentioned in the official communiqué of the European Commission (2023), a core value of the DSA is creating a safer online world by protecting users from illegal content, assessing cyberbullying and other forms of online harassment, and implementing a more transparent content moderation framework for platforms. The DSA defines *illegal content* as "any information that, in itself or in relation to an activity, including the sale of products or the provision of services, is not in compliance with Union law or the law of any Member State which is in compliance with Union law, irrespective of the precise subject matter or nature of that law" (DSA, Article 3h). Such a broad definition of *illegal content* is welcome in principle, for it can help to ensure that the legislation covers the most comprehensive possible range of harmful content online. Nevertheless, it can also lead to uncertainty in its application.

To ensure the above while not imposing an unnecessary burden to conform, the DSA does not stray far from European legislation already in effect, namely the E-Commerce Directive, but instead enforces it and aims to make its stipulations more user-centered. The Directive, enacted in 2000, proposes the renowned "notice-and-takedown" mechanism (DSA, Article 14), which permits intermediaries to enjoy immunity for illegal content on their platforms, whether it entails illegal substances, violence, or nudity, as long as they lack knowledge of such content or, upon receiving notice or otherwise becoming aware of it, remove it from their pages (Urban, Karaganis and Schofield 2016). At the same time, the notice-and-takedown provision also has the power to free platforms from conducting general monitoring and/or fact-finding, which practically waives their obligation to seek out illegal content on their pages.

Added to that, Article 16 of the DSA strengthens the notice-and-takedown mechanism, albeit with a slight twist. The mechanism proposed by the new regulation sets forth that service providers acting as hosts, including online platforms, are to implement additional mechanisms to allow users to report seemingly illegal content on their platforms or content that may violate the platform's terms of use. The latter option derives from the profoundly controversial liaison between users and platforms; although the equality of arms in the relationship is practically nonexistent due to the extreme financial power of platforms, users are essentially in a

contractual relationship with platforms once they accept their terms of use (Ortolani 2023a). Thus, given the principle of *pacta sunt servanda*, users are to submit to every provision implemented by the platform or else be restricted from using the platform's services (Quintais, Appelman and Ó Fathaigh 2023). Subsequently, to enforce transparency and adequate responsibility, platforms are required to provide a statement of reason if they decide to impose restrictions on the content or any user who has published such content after having received a notice (DSA, Article 17). *Restriction* thus serves as an umbrella term in the DSA, with applications that include making content less visible or accessible on the platform, removing user accounts and restricting services, and even demonetizing in some instances (Leersen 2023). To accentuate the principle of accountability in connection to transparency, statements of reason are required to state the deciding facts of the restriction, whether automated tools were used to determine the content's unlawfulness, whether the content violates the law or terms of use, and the information that users are given for redress concerning the restriction imposed upon them (Leersen 2023). Last, internal provisions for handling complaints are to be presented. The chief aim of those mechanisms is to allow platforms to review initial decisions and thereby provide an in-house solution for complaints lodged by sanctioned users. In practice, as Pietro Ortolani (2023a) has highlighted, internal complaint handling aims to conceptualize moderation systems to synthesize the difficulties of the task of (mostly human) content moderation and users' rights.

A final aspect that merits mention is the DSA's rigid differentiation of platforms by size and size of userbase. VLOPs, including Google or Facebook, bear more obligations than platforms with fewer than 45 million monthly active users. A crucial part of that discriminative (Husovec 2024) regulatory philosophy is risk management. The DSA proposes that VLOPs shall play an active role in assessing and mitigating *systemic risk*, a term vaguely defined by the DSA in four categories: dissemination of illegal content, adverse effects on fundamental rights, negative effects on democratic processes (e.g., elections) and public safety, and negative effects on a person's well-being or health (DSA, Recitals 80–83). To put the provisions into practice, systemic risks may exemplified as fake news propaganda during election campaigns, disinformation regarding COVID-19 or other public health emergencies, or the possibility to sell or buy illegal substances on the platform. VLOPs are therefore obliged to act when systemic risks are identified. They are to assess said risks—that is, identify and analyze possible risks emerging from the design or algorithmic activities of the platforms—and mitigate them by, for instance, taking effective, timely actions to prevent further risks from arising (DSA, Articles 34 and 35).

2.2. A critical approach: The major problems with the DSA's provisions regarding content moderation

As for statements of reason, the rationale behind the narrowness of the scope of reasoning to be provided remains unclear. As Ortolani (2023b) has highlighted, the DSA does not require reasons from platforms when they forgo taking moderation

measures—for example, leave user-generated content up or decide to not restrict an allegedly hateful account (cf. Lendvai 2024).

As for the notice-and-action mechanism, the chief practical problem may arise from the contradictory interpretation of the provision. What indeed should platforms do if user-posted content is not clearly unlawful but the terms of use strictly forbid it from being posted? For example, should Facebook remove graphic but educational videos of medical procedures? Or is selling "haunted items" with superpowers legal on Facebook Marketplace when the descriptions of the items are deceptive? It is also dubious how effective any network address translation (i.e., NAT) will be on different platforms. Though Meta (Clegg, 2023) introduced a number of new features to be in conformity with the DSA, X, for example, is actively seeking to disregard all European initiatives regarding the fight against fake news or platform governance (Miller 2023). In those ways, Article 16 introduced more questions than answers regarding practicality for both users and content moderators.

As for internal systems for handling complaints, Aleksandra Kuczerawy (2023) has speculated about how the nondiscriminatory and nonarbitrary nature of internal complaint handling should be interpreted. Such vague terms cause ambiguity for moderators and addressees and may lead to platforms handling different scenarios differently. For instance, should a high-profile account be "whitelisted," so to speak? From a practical viewpoint, the vague use of expressions in the DSA also breeds a different type of polemic. After all, what would happen, as Kuczerawy has imagined, if users were to appeal decisions en masse?

The DSA also fails to identify and mitigate specified forms of illegal content. As Asha Allen (2022) has highlighted, online gender-based violence (e.g., cyberstalking) and nonconsensual forms of content sharing (e.g., revenge porn), as described by the Proposal for a Directive for Combating Violence Against Women and Domestic Violence—both of which are highly prevalent but critically overlooked problems in the online sphere (European Institute for Gender Equality 2024)—are not considered to be systemic risks by the European regulation. As a consequence, VLOPs are not required to implement specific measures to assess such polemics (Allen 2022). It is also unclear how platforms would be able to evaluate gender-based violence when identified as a systemic risk, for most platforms have no straightforward tools other than restricting accounts or removing content. On top of that, moderators are not provided with adequate education on gender-based violence. Despite non-binding commitments, that issue in the context of systemic risks and content moderation remains to be resolved.

Last, risk assessment and mitigation are also problematic for several reasons. The DSA proposes an *ex ante* position for content moderation regarding systemic risks; risks are unavoidable and unpreventable on popular platforms, and the *prima habet* approach is therefore to mitigate preexisting risks. However, that regulatory philosophy may spawn two different but equally detrimental practices. The first is overzealous content moderation by platforms, including the constant, active monitoring of content and identifying and addressing polemic content as systemic risks so that they will not have to engage in mitigation procedures. That option effectively permits digital authoritarianism and may likely cause a chilling effect among users.

The second practice is the polar opposite of overregulation: platforms may choose to not identify systemic risks under the flag of understanding freedom of expression in a more liberal sense. As a consequence, disinformation about pandemics (cf. Kouzy et al. 2020) or political propaganda from foreign states may be able to bypass the provisions of the DSA (Lendvai 2023a). From a practical standpoint, the practice of the platforms will be instrumental. After all, the DSA provides little to no guidance on whether the marginalization of certain groups online or algorithmic bias resulting in harmful societal effects can be understood as systemic risks.

3. Alternative ways to reduce harmful content

The fight to secure freedom of expression has undoubtedly reached an exciting new stage in recent years, as the internet enables billions of people to express themselves freely. States have sought to retake control since the mid-2010s by regulating the disclosure of illegal content (Gosztonyi 2023, 182). Nevertheless, they have usually confronted a situation in which regulation can be interpreted only "at different and sometimes overlapping levels: from the local to the supra-national and global" (Raboy and Padovani 2010). Thus, "the vertical, centralized and State-based modes of traditional regulation have been complemented by collaborative, horizontal arrangements, leading to 'a complex ecology of interdependent structures with a vast array of formal and informal mechanisms working across a multiplicity of sites'" (Hintz 2015, 111). As Jack Balkin (2018, 1153) has put it, specifically regarding new methods of content regulation, "In addition to targeting speakers directly, States now target the owners of private infrastructure, hoping to coerce or co-opt them into regulating speech on the nation State's behalf." In another alarming sign, Adrian Shahbaz and Allie Funk (2021, 2) of Freedom House wrote as early as 2021 that "global norms have shifted dramatically toward greater government intervention in the digital sphere."

3.1. Years of cooperation in the EU in the 2010s

An essential step in the fight against harmful content in the EU was adopting and signing two codes of conduct. In that regard, Petra Láncos, Napoleon Xanthoulis, and Luis Jiménez (2023) have claimed that using soft law is a strategic means to provide a frame of reference for member states without taking a step toward harmonization in a field subject to interventionist measures. The first was the EU Code of Conduct on countering illegal hate speech online, whereby Facebook, Microsoft, Twitter, and YouTube—TikTok joined in 2020 and LinkedIn in 2021—committed to, among other things, reviewing valid reports of unlawful hate speech within 24 hours and removing or making such content inaccessible. The Code found the mentioned notice-and-takedown system to be an appropriate solution to remove such content. Not long after, in 2018, Facebook, Google, Twitter, and Mozilla signed the EU Code of Practice on Disinformation; Microsoft signed in 2019 and TikTok in 2020.

The adoption of those codes was preceded by a European Commission Communication on online misinformation that characterized platform providers' responsibility as a question to be carefully examined. However, such codes, which are only applicable to signatories, cannot be considered a real regulatory solution from a legal standpoint, for they have to "apply within the framework of existing laws of the EU and its member states and must not be construed in any way as replacing, superseding or interpreting the existing and future legal framework." The weakness may have been visible when Twitter, under its new owner, Elon Musk, quickly left the voluntary Code of Practice against disinformation in May 2023 (Krukowska 2023). Although a considerable advantage is that because the companies had to modify their terms of service to comply with their voluntary commitments, they essentially agreed to apply European rules to all their global users.

3.2. An alternative solution

Aside from the mentioned cooperation effort, tech companies have also sought to develop their own solutions. Facebook, for instance, published a white paper (Bickert 2020) in 2020 that highlighted four problems with internet regulation:

- The legal environment and speech standards differ worldwide, which can cause inconsistencies because the biggest tech companies operate internationally.
- Technology is constantly changing, as is speech.
- The implementation will always be imperfect because it has to be applied in many languages and dialects.
- Companies are merely intermediaries, not content creators.

On those bases, in 2020 Facebook announced the creation of an independent oversight board of recognized experts tasked with mitigating the enormous amount of problematic content. The board has been described as the company's in-house Supreme Court, and many have anticipated that the members' expertise would improve the quality of the company's decisions (Lendvai 2023b). However, in time, Facebook's first batch of decisions revealed that the sought-after solution was not found. Moreover, according to Kate Klonick (2021), the "star-studded panel and lavish funding will prevent regulation while allowing the company to outsource controversial decisions." A similar point was made by Laurence Tribe, that Facebook is "generating a patina of legitimacy that the board's composition or function cannot justify" (Morgan 2021). It is their shared view that Facebook's oversight board is merely designed to mislead the public—and even more so, U.S. and European politicians—and to serve as a distraction.

3.3. Expert recommendations on legislation

In Judit Bayer's (2019, 20–21) view, "Because of the international nature of the services, the regulation should take place at the supranational level of the EU and

extension of the rules to the global community should be sought." Bayer considers introducing a single but flexible concept necessary, which would greatly facilitate the unification and harmonization of legislation. Daphne Keller and Paddy Leerssen (2020, 223), meanwhile, have argued that content regulation should prevent harm, protect legitimate online speech, and support innovation. However, because those objectives are often presented in competing ways, they significantly undermine the willingness to regulate and, therefore, the effectiveness of regulation. Klonick (2020) envisages the solution as a public role for private actors, which permits big tech companies to do the dirty work of content regulation but also makes it mandatory for them to do so in view of the law.

Concerning the idea of a harmonized European legal system, Miriam Buiten, Alexandre de Streel, and Martin Peitz (2020) have noted that national liability rules are far more difficult to standardize than exceptions to liability. For that reason, they would expect voluntary proactivity by tech companies instead of monitoring certain uploaded content. Despite stressing that such proactivity should not mean a general obligation to monitor (Buiten, de Streel, and Peitz 2020), they do not cut the Gordian knot of Court of Justice of the European Union tied in *Glawischnig-Piesczek v. Facebook* (Gosztonyi 2020, 142–44). Their proposal echoes the suggestion of Natali Helberger, Jo Pierson, and Thomas Poell (2017), who advise shifting the emphasis from competitive responsibility to cooperative responsibility.

Among other expert recommendations, in 2019, four global entities—the United Nations Special Rapporteur on Freedom of Expression, the Organization for Security and Cooperation in Europe Special Rapporteur on Freedom of the Press, the Organization of American States Special Rapporteur on Freedom of Expression, and the African Commission on Human and Peoples' Rights Special Rapporteur on Freedom of Expression and Access to Information—issued a statement on the challenges for freedom of expression in the coming decade (OHCHR 2019). A core issue raised is the implementation of transparency and accountability requirements opposed to the privatized regulation and increasing transparency of content moderation by algorithms and artificial intelligence (OHCHR 2019).

3.4. Policy recommendations

The solution to the problem elaborated in this article—a complex framework combining legal, political, and economic aspects—remains to be established. As Paulina Wu (2015, 309) has described it, "Although private actors alone are insufficient for successful regulation, they must be included in the regulatory process," meaning that all three actors in the "platform governance triangle" need to be involved (Gorwa 2024). It is also clear from the practice of the international courts that large tech companies can no longer hide behind users, who increasingly do not produce content or conduct editorial activities. At the same time, the exponentially increasing amount of content can be controlled only by human moderators and AI with a view to proactivity.

Indeed, there is no ready or easy solution to such an enormously complex problem. Indeed, any legislation should be based on the following ten imperatives:

1. Respecting human rights and fundamental freedoms in the regulation and operation of the internet;
2. Respecting the characteristics of the internet as a complex, constantly changing, and diverse ecosystem, such that regulations have to be created "without destroying the massive benefits an open and diverse global internet can bring" (Suzor 2019, 114);
3. Ensuring the diversity, pluralism, and impartiality of content on the internet;
4. Ensuring the most comprehensive citizen, civil, and user participation in decision-making possible;
5. Facilitating cooperation between the parties involved and their radically greater transparency;
6. Hosting real tests of the effectiveness of different notification and content removal systems;
7. Precisely defining the active and passive roles of service providers (cf. Barata 2021);
8. Maintaining a general ban on monitoring to ensure freedom of expression and prevent excessive content removal;
9. Developing media literacy in all segments of society; and
10. Aiming overall to "deter unwanted speech but not to deter legal and socially valuable communications" (Sartor 2017, 12).

4. Conclusion

In this article, we have outlined the emergence, the historical and regulatory context, the content, and the future and possible development of the DSA. The DSA constitutes a notable attempt to address the concerns embedded within the ambit of platform governance in the European regulatory context. Its pronounced emphasis on transparency, accountability, and the attribution of responsibility delineates a promising framework for cultivating a more secure digital milieu. Nevertheless, as seen in this article's critical examination, the DSA also bears inherent limitations and raises questions regarding the pragmatic efficacy of legal instruments in combating the severe societal, psychological, and emotional harms inflicted by online abuse. Furthermore, apprehensions are rife concerning the practical implementation and potential inadvertent consequences associated with the DSA's provisions on content moderation and systemic risk assessment. Against that background, this article aims to contribute to the fruitful discussion on content moderation and scholarly research on platform governance, a subject just as prevalent to researchers as it is to users of online platforms.

References

Allen, Asha. "An Intersectional Lens on Online Gender Based Violence and the Digital Services Act." *Verfassungsblog: On Matters Constitutional*, November 1, 2022. http://www.doi.org/10.17176/20221101-215626-0

Balkin, Jack M. "Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation." *UC Davis Law Review* 51 (2018): 1149–210. http://dx.doi.org/10.2139/ssrn.3038939

Barata, Joan. "The Digital Services Act and the Reproduction of Old Confusions: Obligations, Liabilities and Safeguards in Content Moderation." *Verfassungsblog: On Matters Constitutional*, March 2, 2021. http://www.doi.org/10.17176/20210302-154101-0

Bayer, Judit. "Between Anarchy and Censorship: Public Discourse and the Duties of Social Media." *CEPS Papers in Liberty and Security in Europe* 3 (2019). https://www.ceps.eu/ceps-publications/between-anarchy-and-censorship/

Bickert, Monika. "Online Content Regulation." *Facebook Documents* (2020). https://about.fb.com/wp-content/uploads/2020/02/Charting-A-Way-Forward_Online-Content-Regulation-White-Paper-1.pdf

Buiten, Miriam C., Alexandre de Streel, and Martin Peitz. "Rethinking Liability Rules for Online Hosting Platforms." *International Journal of Law and Information Technology* 2 (2020): 139–66. https://doi.org/10.1093/ijlit/eaaa012

Casey, Rachel. "John Stuart Mill and Social Media: Evaluating the Ethics of De-Platforming." *University of Central Florida Department of Legal Studies Law Journal* 4 (2021): 15–34. https://heinonline.org/HOL/LandingPage?handle=hein.journals/ucflaegs4&div=6&id=&page=

Cauffman, Caroline, and Catalina Goanta. "A New Order: The Digital Services Act and Consumer Protection." *European Journal of Risk Regulation* 12, no. 4 (2021): 758–74. https://doi.org/10.1017/err.2021.8

Decarolis, Francesco, and Muxin Li. "Regulating Online Search in the EU: From the Android Case to the Digital Markets Act and Digital Services Act." *International Journal of Industrial Organization* 90 (2023): 102983. https://doi.org/10.1016/j.ijindorg.2023.102983

European Institute for Gender Equality. "Cyberbullying Restricts Young Women's Voices Online." October 11, 2018. https://eige.europa.eu/newsroom/news/cyberbullying-restricts-young-womens-voices-online

European Commission. "DSA: Making the Online World Safer." Accessed May 14, 2025. https://digital-strategy.ec.europa.eu/en/policies/safer-online

Gorwa, Robert. *The Politics of Platform Regulation: How Governments Shape Online Content Moderation.* Oxford: Oxford University Press, 2024. https://doi.org/10.1093/oso/9780197692851.001.0001

Gosztonyi, Gergely. "The Contribution of the Court of Justice of the European Union to a Better Understanding the Liability and Monitoring Issues Regarding Intermediary Service Providers." *Annales Universitatis Scientiarum Budapestinensis de Rolando Eötvös Nominatae. Sectio Iuridica* 59 (2020): 133–44. https://doi.org/10.56749/annales.elteajk.2020.lix.7.133

Gosztonyi, Gergely. "Possible Directions for the Future." In *Censorship from Plato to Social Media*. Law, Governance and Technology Series, vol. 61. Cham: Springer, 2023. https://doi.org/10.1007/978-3-031-46529-1_12

Gosztonyi, Gergely. “How the European Union Had Tried to Tackle Fake News and Disinformation with Soft Law and What Changed with the Digital Services Act?” *Frontiers in Law* 4 (2025): 102–13. https://doi.org/10.6000/2817-2302.2024.03.12

Harawa, Daniel S. “Social Media Thoughtcrimes.” *Pace Law Review* 35, no. 1 (2014): 366–97. https://doi.org/10.58948/2331-3528.1886

Helberger, Natali, Jo Pierson, and Thomas Poell. “Governing Online Platforms: From Contested to Cooperative Responsibility.” *The Information Society* 34, no. 1 (2017): 1–14. https://doi.org/10.1080/01972243.2017.1391913

Hintz, Arne. “Social Media Censorship, Privatized Regulation, and New Restrictions to Protest and Dissent.” In *Critical Perspectives on Social Media and Protest: Between Control and Emancipation*, edited by Lina Dencik and Oliver Leistert, XX–XX. London: Rowman & Littlefield, 2015.

Husovec, Martin. *Principles of the Digital Services Act*. Oxford: Oxford University Press, 2024. https://doi.org/10.1093/law-ocl/9780192882455.001.0001

Keller, Daphne, and Paddy Leerssen. “Facts and Where to Find Them: Empirical Research on Internet Platforms and Content Moderation.” In *Social Media and Democracy*, edited by Nathaniel Persily and Joshua A. Tucker, 220–51. Cambridge: Cambridge University Press, 2020. https://doi.org/10.1017/9781108890960.011

Klonick, Kate. “The Facebook Oversight Board: Creating an Independent Institution to Adjudicate Online Free Expression.” *Yale Law Journal* 129, no. 8 (2020): 2418–99. https://www.yalelawjournal.org/feature/the-facebook-oversight-board

Klonick, Kate. “Inside the Making of Facebook’s Supreme Court.” *The New Yorker*, February 12, 2021. https://www.newyorker.com/tech/annals-of-technology/inside-the-making-of-facebooks-supreme-court

Kouzy, Ramez, Joseph Abi Jaoude, Afif Kraitem, Molly B. El Alam, Basil Karam, Elio Adib, Jabra Zarka, Cindy Traboulsi, Elie Akl, and Khalil Baddour. “Coronavirus Goes Viral: Quantifying the COVID-19 Misinformation Epidemic on Twitter.” *Cureus* 12, no. 3 (2020): e7255. https://doi.org/10.7759/cureus.7255

Krukowska, Eva. “Twitter Withdraws From EU Disinformation Code, Commissioner Says.” *Time*, May 27, 2023. https://time.com/6283183/twitter-withdraws-from-eu-disinformation-code-commissioner-says

Kuczerawy, Aleksandra. “Remedying Overremoval.” In *Putting the DSA into Practice*, edited by Joris van Hoboken, Ilaria Buri, João Pedro Quintais, Ronan Fahy, Naomi Appelman, and Marlene Straub, 167–82. Berlin: Verfassungsbooks, 2023. http://www.doi.org/10.17176/20230208-093135-0

Láncos, Petra L., Napoleon Xanthoulis, and Luis Arroyo Jiménez. *The Legal Effects of EU Soft Law*. Edward Elgar Publishing, 2023. https://doi.org/10.4337/9781802208917.00005

Leerssen, Paddy. “An End to Shadow Banning? Transparency Rights in the Digital Services Act between Content Moderation and Curation.” *Computer Law & Security Review* 48 (2023): 105790. https://doi.org/10.1016/j.clsr.2023.105790

Lendvai, Gergely F. "Media in War: An Overview of the European Restrictions on Russian Media." *European Papers* 8, no. 3 (2023a): 1235–45. https://doi.org/10.15166/2499-8249/715

Lendvai, Gergely F. "'Pure Rat Country' – Reflections on Case Decision 2022-001-FB-UA of Facebook Oversight Board (Knin Cartoon Case)." *Journal of Digital Technologies and Law* 1, no. 3 (2023b): 612–28. https://doi.org/10.21202/jdtl.2023.26

Lendvai, Gergely F. "Taming the Titans? – Digital Constitutionalism and the Digital Services Act." *ESSACHESS* 17, no. 2 (2024): 169–84. https://doi.org/10.21409/0M3P-A614

Clegg, N. "New Features and Additional Transparency Measures as the Digital Services Act Comes Into Effect." *Meta*, August 22, 2023. https://about.fb.com/news/2023/08/new-features-and-additional-transparency-measures-as-the-digital-services-act-comes-into-effect/

Miller, Gabby. "Elon Musk Will Test the EU's Digital Services Act." *Tech Policy*, September 11, 2023. https://techpolicy.press/elon-musk-will-test-the-eus-digital-services-act/

Morgan, John. "Scholars Overseeing Facebook: Supreme Court or Fairness Washing?" *Times Higher Education*, March 12, 2021. https://www.timeshighereducation.com/news/scholars-overseeing-facebook-supreme-court-or-fairness-washing

OHCHR. "Twentieth Anniversary Joint Declaration: Challenges to Freedom of Expression in the Next Decade." Accessed May 14, 2025. https://www.ohchr.org/sites/default/files/Documents/Issues/Opinion/JointDeclaration10July2019_English.pdf

Ortolani, Pietro. "The Digital Services Act, Content Moderation and Dispute Resolution." *SSRN Electronic Journal*, 2023a. https://doi.org/10.2139/ssrn.4356598

Ortolani, Pietro. "If You Build It, They Will Come: The DSA 'Procedure Before Substance' Approach." In *Putting the DSA into Practice*, edited by Joris van Hoboken, Ilaria Buri, João Pedro Quintais, Ronan Fahy, Naomi Appelman, and Marlene Straub, 151–66. Berlin: Verfassungsbooks, 2023b. http://www.doi.org/10.17176/20230208-093135-0

Papp, János Tamás. *A Közösségi Média Szabályozása a Demokratikus Nyilvánosság Védelmében*. Budapest: Wolters Kluwer, 2022.

Pew Research Center. "Social Networking Sites and Our Lives." June 16, 2011. https://www.pewresearch.org/2011/06/16/social-networking-sites-and-our-lives/

Pillalamarri, Akhilesh and Cody Stanley. "Online Content Regulation: An International Comparison." *International Law and Policy Brief*, December 8, 2021. https://studentbriefs.law.gwu.edu/ilpb/2021/12/08/online-content-regulation-an-international-comparison/

Quintais, João Pedro, Naomi Appelman, and Ronan Ó Fathaigh. "Using Terms and Conditions to Apply Fundamental Rights to Content Moderation." *German Law Journal* 24, no. 5 (2023): 881–911. https://doi.org/10.1017/glj.2023.53

Raboy, Marc, and Claudia Padovani. "Mapping Global Media Policy: Concepts, Frameworks, Methods." *Communication, Culture & Critique* 3, no. 2 (2010): 150–69. https://doi.org/10.1111/j.1753-9137.2010.01064.x

Sartor, Giovanni. *Providers Liability: From the eCommerce Directive to the Future. In-Depth Analysis for the Directorate-General for Internal Policies*. Strasbourg: European Parliament, 2017. https://www.europarl.europa.eu/RegData/etudes/IDAN/2017/614179/IPOL_IDA(2017)614179_EN.pdf

Shahbaz, Adrian, and Allie Funk. "The Global Drive to Control Big Tech." *Freedom House*. 2021. https://freedomhouse.org/report/freedom-net/2021/global-drive-control-big-tech

Suzor, Nicolas P. *Lawless*. Cambridge: Cambridge University Press, 2019.

Tutt, Andrew. "The New Speech." *Hastings Constitutional Law Quarterly* 41, no. 2 (2014): 235–98. https://repository.uclawsf.edu/hastings_constitutional_law_quaterly/vol41/iss2/1/

Urban, Jennifer M., Joe Karaganis, and Brianna Schofield. "Notice and Takedown in Everyday Practice." *SSRN Electronic Journal*, 2016. https://doi.org/10.2139/ssrn.2755628

Wu, Paulina. "Impossible to Regulate? Social Media, Terrorists, and the Role for the U.N." *Chicago Journal of International Law* 16, no. 1 (2015): 281–311. https://chicagounbound.uchicago.edu/cjil/vol16/iss1/11/

Legal documents

Directive (EU) 2017/541 of the European Parliament and of the Council of 15 March 2017 on combating terrorism and replacing Council Framework Decision 2002/475/JHA and amending Council Decision 2005/671/JHA

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

Proposal for a Directive of the European Parliament and of the Council for combating violence against women and domestic violence

Regulation (EU) 2019/1150 of the European Parliament and of the Council of 20 June 2019 on promoting fairness and transparency for business users of online intermediation services

Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Markets Act)

Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services and amending Directive 2000/31/EC (Digital Services Act)

The 2022 Code of Practice on Disinformation (EU)