Dorina Gyetván*

Censorship and Freedom of Expression in the Age of Social Media

Abstract

Although social media platforms have altered the structure of the public sphere, they have also inherited some of its issues, notably the problem of censorship. The phenomenon has remained, just its methods and practices have changed: censorship used to be strictly connected to states, but in the digital age, it is exercised by multiple actors, such as states, private companies and individuals (users), posing a unique and multilevel threat to freedom of expression. Social media service providers are motivated by their own economic interests and pressured by vague laws that impose liability for third-party user content; the combination of these factors steers service providers to ignore human rights standards, err on the side of caution, and tend to remove, block or restrict any questionable content in order to avoid liability. Therefore, online freedom of expression faces problems, just as it did in offline formats in analogue times, and often even more severe ones. However, technological advances mean that new censorship methods often remain unperceived by users and, therefore, often avoid the harsh criticism surrounding traditional censorship. In 2020, Facebook (Meta) set up the Oversight Board, a uniquely positioned semi-independent expert group as a sort of court-like body to deal with some of the more high-profile, influential and complex social-media-related decisions and offer a remedy against contract-based content moderation (censorship).

Keywords: freedom of expression, censorship, Oversight Board, flooding, content moderation

I Preamble

Social media platforms have fundamentally altered the structure of the public sphere, allowing masses of people to post their opinions, learn about the opinions of others,

* Dr Dorina Gyetván, PhD Candidate, ELTE Doctoral School of Law. ORCID iD: 0000-0001-5361-0011.

and share other's expressions.[1] At the beginning of the Internet's rise, the latter was put on a pedestal as the ultimate embodiment of freedom, equality and total freedom from censorship. Given the above characteristics, it was perceived as almost inconceivable that any state could control and restrict the flow of information.[2] On the one hand, one could simply conclude that 'new media were born to be free, although, inevitably, they are not a lawless domain, they do not tolerate any censorship or authority, the main reason being that media cannot be controlled by the state due to the technological nature thereof'.[3] Although the method, motive, scope and effectiveness of content moderation on the Internet varies from state to state, it initially seemed like social media platforms would allow users to break free from the traditional 'top-to-bottom' nature of state power.[4]

However, '[...] the control of information on the Internet and Web is certainly feasible, and technological advances do not therefore guarantee greater freedom of speech. There are many tools available and still more in development.'[5] Therefore, the Internet is not inherently distinct from the traditional media world and has inherited many of its problems as well.[6] In fact, transforming previously existing restrictions and combining them with advanced technological tools creates a brand new environment for arbitrary state intervention while simultaneously allowing private service providers to intervene arbitrarily. The early phenomena of decentralisation and complete lack of regulation are continuously declining. Power is increasingly concentrated in the hands of a few influential service providers; the growing economic power and the unstoppable development of technology have brought about overbearing censorship in the online space, similar to the censorship of traditional mass communication.

The aim of this paper is to take the traditional phenomenon of censorship as a starting point and investigate online freedom of expression as mainly controlled by modern social media service providers, to outline the new forms of censorship, and to briefly discuss the significance of intermediary service provider liability in this context, and the possible role of

---

1 Dirk Voorhoof, Hannes Cannie, 'Freedom of Expression and Information in a Democratic Society. The Added but Fragile Value of the European Convention on Human Rights' (2010) 72 (4–5) International Communication Gazette, DOI: https://doi.org/10.1177/1748048510362711

2 Gabriella Szabó, 'Internetes portálok médiaszociológiai és politológiai elemzése' (2008) 17 (4) Politikatudományi Szemle 62–63; Judit Oszti, *Az elektronikus média szerepe korunk háborúinak társadalmi-politikai megítélésében és a közgondolkodás formálásában* (Zrínyi Miklós Nemzetvédelmi Egyetem 2009, Budapest) 25.

3 László Majtényi, Gábor Polyák, 'A szabadság hazai hagyományának megtagadása – új médiatörvények Magyarországon' (2011) 4 (1) Közjogi Szemle 4; András Koltay, 'Az internet mint médium, a sajtószabadság és a demokratikus nyilvánosság' (2014) 14 (4) Információs Társadalom 16, DOI: https://doi.org/10.22503/inftars.XIV.2014.4.1

4 Márton Iványi, 'Az online közösségi hálózatok és a véleménynyilvánítás pozitív és negatív szabadsága' (2015) 25 (3) Iskolakultúra 82, DOI: https://doi.org/10.17543/ISKKULT.2015.3.72

5 William H. Dutton and others, *Freedom of connection, freedom of expression: the changing legal and regulatory ecology shaping the Internet* (UNESCO Publishing 2011, Paris).

6 Koltay (n 3) 16.

the Oversight Board concerning the extent to which such an entity is able to counterbalance the shortcomings of content moderation (private censorship).

II Traditional Censorship in the Narrow Sense

Censorship is as old as mankind itself. The Old Testament showed how Jeremiah was condemned to death for criticising the leaders of the people and prophesying against Judah.[7] Censorship of certain forbidden subjects was also common practice in ancient Greece; for example, the Greek philosopher Socrates was condemned to death for his teachings in 399 BC. In the analogue age, the term 'censorship' originally referred to state intervention in the content published in media, which proved to be an effective tool prior to the digital age,[8] as all (most) major communication outlets involved some kind of centralised state control, and the prospect of severe punishment was sufficiently discouraging to instil fear in citizens due to its draconian rigour.

What is obvious is that censorship was primarily linked to the state as the embodiment of concentrated power. In the age of modern technology, however, the concept should be more broadly interpreted and generally used in a wider sense:[9] we should not only understand censorship to refer to restrictions imposed by the state(s) but also to encompass the ability and inclination of social media service providers, which are gaining power by the minute, to restrict content on the basis of their private, financially motivated interests, which, over time, encourages the presence of so-called self-censorship on an ever wider scale as a result of external and internal moral convictions.[10]

III Modern Censorship

It might seem sensible to simply argue that social media providers cannot censor, as they are private entities, and only the state has the authority to censor.[11] However, to avoid the looming liability posed by national and international laws, service providers have taken a number of steps towards enacting more extensive regulations, such as hiring whole armies of moderators, adopting and enforcing private (contractual) rules, and developing a wide

---

7 Bible. *Old Testament*. Jeremiah 26.

8 John Naughton, 'How China censors the net: by making sure there's too much information' (2018) The Guardian, <https://www.theguardian.com/commentisfree/2018/jun/16/how-china-censors-internet-information> accessed 15 October 2024.

9 Gergely Gosztonyi, 'A cenzúra tipizálása a politikai cenzúra rövid történetének tükrében' (2022) 23 (1) Médiakutató.

10 András Koltay, 'A médiatartalmak közzététel előtti korlátozásának lehetőségei: engedélyezés, regisztráció, cenzúra, végzések' (2014) 10 (1) Iustum Aequum Salutare 77.

11 Amitai Etzioni, 'Should We Privatize Censorship?' (2019) 36 (1) Issues in Science and Technology 19.

variety of algorithm-based solutions. One of the main drawbacks of such control is the lack of legitimacy and accountability behind the decisions of private entities, mainly due to a (total) lack of transparency.[12] One thing that is for sure, however, is that the rise of social media services has triggered a heated and ongoing debate about how these platforms moderate the content published, shared, or generated by users on their platforms. In light of the above, Jack M. Balkin's 'new school'[13] of speech regulation can be described by three determining features:

1. collateral censorship;
2. the intertwining of the operations of public and private service providers; and
3. global private regulation.[14]

The algorithmic society has also multiplied the entities that are able to restrict freedom of expression; the digital age has brought about a 'pluralist model of speech governance',[15] so that the user can now expect intervention by the state and private entities, either through indirect censorship imposed by state coercion or due to the service providers' own economic interests. Social media has the power to shape and change public opinion, and the decisions of social media service providers determine who can participate in online public discourse and to what extent. Such characteristics are very similar to the phenomenon of traditional censorship.

Censorship in the digital age and censorship in the traditional sense certainly have one thing in common: they both threaten freedom of expression.[16] However, a significant difference, which makes the phenomenon of private censorship an even more significant issue, is that while traditional media censorship is typically confined to geographical boundaries such as state or administrative borders, in the age of the Internet, it is increasingly common to see more extensive, worldwide restrictions on freedom of expression.

From New Zealand[17] to the United States of America,[18] social media has often been at the centre of attention due to the repercussions of numerous tragedies. Therefore, the obligations, liabilities and responsibilities of service providers seem to be drifting further and further away from the initial solution of awarding them full immunity. In addition to the obvious positive effects of social media, we also perceive that the provision of unprecedented

---

12 Haggart, Blayne, Keller, Clara Iglesias, 'Democratic legitimacy in global platform governance' (2021) 45 (6) Telecommunications Policy 2, DOI: https://doi.org/10.1016/j.telpol.2021.102152

13 Jack M. Balkin, 'Free speech in the algorithmic society. The new school of big data, private regulation and the regulation of expression' (2018) 118 (7) Columbia Law Review 1173.

14 Balkin (n 13) 1176.

15 Balkin (n 13) 1186.

16 Iványi (n 4) 83.

17 'Mészárlás Új-Zélandon: vészforgatókönyvet vett elő a YouTube' (2019) HVG, <https://hvg.hu/tudomany/20190319_uj_zeland_meszarlas_tomeggyilkossag_facebook_elo_video> accessed 15 October 2024.

18 Teddy Wayne, 'The Trauma of Violent News on the Internet' (2016) The New York Times, <https://www.nytimes.com/2016/09/11/fashion/the-trauma-of-violent-news-on-the-internet.html> accessed 15 October 2024.

publicity cannot (and does not) exist without limits: social media service providers, with their power, technological capabilities and position, also assist in silencing the opinions of users. Social media service providers are in their heyday. The phenomenon of private censorship is a recurrent problem, as one of the main issues is that service providers' content-related decisions are (partly) rooted in their own economic interests, often without a legal basis. They are fully or partially made by employing artificial intelligence, lacking transparency, guarantees and even the possibility of effective remedy (appeal).[19]

IV New Forms of Censorship

In recent years, more sophisticated methods have been deployed by service providers, so new methods of regulating and restricting freedom of expression other than traditional censorship have emerged. These show great variety, but the common ground is that their emergence is ultimately somehow linked to the rise of private service providers:

1. self-censorship has become even more prominent with the emergence of service providers;
2. reverse censorship has emerged, involving controlling public opinion by flooding the information space using private operators;
3. collateral censorship refers to the phenomenon whereby states use their power over social media providers to encourage moderation by imposing the principle of secondary liability.

1 Self-censorship

Censorship is usually only discussed in the context whereby one entity restricts another individual's expression.[20] However, one manifestation of indirect censorship is self-censorship, which refers to the phenomenon whereby individuals such as social media users 'voluntarily' act in ways that eliminate the need for a censor.

Within the category of self-censorship, we can distinguish between public and private self-censorship.[21] In the case of public self-censorship, we refer to the situation when the censor is a government or a public authority, the restricted person is a natural person or legal

[19] Thiago Dias Oliva, 'Content Moderation Technologies: Applying Human Rights Standards to Protect Freedom of Expression' (2020) 20 (4) Human Rights Law Review 612, DOI: https://doi.org/10.1093/hrlr/ngaa032

[20] Paul Sturges, 'Self-Censorship: Why We Do the Censors' Work For Them' (2008) LIBCOM Conference, 2, <https://www.ifla.org/wp-content/uploads/2019/05/assets/faife/publications/sturges/self-censorship.pdf> accessed 15 October 2024.

[21] Philip Cook, Conrad Heilmann, 'Two Types of Self-Censorship: Public and Private' (2013) 61 (1) Political Studies 179, DOI: https://doi.org/10.1111/j.1467-9248.2012.00957.x

entity, and the latter resorts to self-censorship because of looming sanctions or prior filters imposed by external entities. Such pressures may include, for example:

1. a compelling reason related to the owner of the service provider;
2. ex-post liability; or
3. prior restrictions such as licensing procedures or any kind of approval.[22]

Public self-censorship means that individuals internalise certain aspects of public censorship and then censor themselves.[23] The indirect result of the public sphere's 'intervention' is thus the abandonment of one's true wilful expression: ie, unwanted voluntary – and by voluntary, I mean here the absence of actual direct, external intervention – self-censorship without involving any legal dispute or potential costs. Through such indirect pressure, content potentially generating controversy will be removed from social media without any tangible intervention.[24] In contrast to public self-censorship, private self-censorship is, in fact, an internal self-regulatory process involving an irreconcilable antagonism between what an individual considers publicly permissible to express and what one actually wishes to express publicly.[25]

From the above, it is clear that the distinction between the two types of self-censorship is based on whether the censor and the restricted person are united in one or two distinct entities: in the case of public self-censorship, the censor is a separate entity from the individual, whereas, in case of private self-censorship, the censor and the restricted are the same person, thus we refer to the suppression of one's own attitudes.[26]

Self-censorship has become an even more significant threat to freedom of expression since the emergence of intermediary service providers, as the failure to express one's views on the platforms with the largest audiences cannot be replaced by expressing one's self on other, less popular platforms.[27] This is also why multiple, simultaneous censors are particularly threatening to freedom of expression (as much as certain codes of conduct or other contractual provisions are) – namely, because external regulators (ie, those other than state law or international law) can have such a chilling effect on users through their vague wording and lack of procedural guarantees against arbitrary decisions that it leads to self-censorship.[28]

---

22 Sturges (n 20).

23 Cook, Heilmann (n 21) 179.

24 András Koltay, *A szólásszabadság alapvonalai* (Századvég Kiadó 2009, Budapest) 202.

25 Cook, Heilmann (n 21) 179.

26 Cook, Heilmann (n 21) 194.

27 András Koltay, 'Szólásszabadság, avagy a social media jogi státusa – 5. rész' (2019) Jogászvilág, <https://jogaszvilag.hu/szakma/szolasszabadsag-avagy-a-social-media-jogi-statusa-5-resz> accessed 15 October 2024; János Tamás Papp, 'Recontextualizing the Role of Social Media in the Formation of Filter Bubbles' (2023) 11 (1) Hungarian Yearbook of International Law and European Law 136–150, DOI: https://doi.org/10.5553/HYIEL/266627012023011001012

28 Koltay (n 27).

2 Reverse Censorship

The phenomenon of what Eugene Volokh calls 'cheap speech' refers to the fact that, with the development of technology, anyone can form an opinion on any issue and distribute it for free (or at least without significant cost).[29] This opportunity is no longer limited to those who can navigate the decisions and expectations or prescriptions of old gatekeepers: '[t]oday we live in an environment where speech is cheap, where it is abundant, where the fundamental challenge is no longer finding speakers but rather finding attention for speech'.[30]

Perhaps no one would argue that censorship has a long history in China, but the rise of social media platforms is generating more than 30 billion different pieces of content every day,[31] a volume of content that has made control over public discourse an unimaginably complex task.[32] There are three possible ways to effectively control the floods of information on the Internet: by fear, causing traction, and flooding.[33] Causing fear is far too costly and can easily backfire by generating significant resistance due to social media's potential. This reverse outcome is often referred to as the 'Streisand effect' in the literature – the fact that attempts to hide certain information can often end up attracting more public attention to whatever the actor (here, the state) was initially trying to hide.[34] Therefore, Chinese media typically uses one of three methods – the Great Firewall, keyword blocking or flooding – to regulate online content, the latter of which is distinct from the censorship tendencies of any other state.[35]

Flooding,[36] which refers to the collective and structured method of disseminating masses of information,[37] is not about removing undesirable content as quickly as possible or imposing liability on the person responsible for the expression in question, but an approach designed to distract and eliminate the threat inherent in the power of the community and volume of information.[38] The flooding method is more economical and no less effective

---

29 Eugene Volokh, 'Cheap Speech and What It Will Do' (1995) 104 (7) Yale Law Journal 1805, 1849, DOI: https://doi.org/10.2307/797032

30 David A. Graham, 'The Age of Reverse Censorship' *The Atlantic*, 2018, <https://www.theatlantic.com/politics/archive/2018/06/is-the-first-amendment-obsolete/563762> accessed 15 October 2024.

31 Naughton (n 8).

32 Graham (n 30).

33 Margaret Earling Roberts, *Fear, Friction, and Flooding: Methods of Online Information Control* (Harvard University 2014, Cambridge, Massachusetts) 91.

34 Sue Curry Jansen, Brian Martin, 'The Streisand Effect and Censorship Backfire' (2015) 9 (1) International Journal of Communication 666.

35 Gergely Gosztonyi, 'Special models of internet and content regulation in China and Russia' (2021) 9 (2) ELTE Law Journal 87–99, DOI: https://doi.org/10.54148/ELTELJ.2021.2.87

36 Roberts (n 33) 32–33.

37 Roberts (n 33) 91.

38 Gary King, Jennifer Pan, Margaret Earling Roberts, 'Reverse-engineering censorship in China: Randomized experimentation and participant observation' (2014) 345 (6199) Science 1251722, DOI: https://doi.org/10.1126/science.1251722

than traditional censorship. This approach has the added advantage of being able to control community discourse through 'the most targeted suppression of expression'[39] while avoiding the outrage associated with direct censorship.[40]

3 Collateral Censorship – Content Moderation or Censorship?

The term 'collateral censorship' was first coined by Michael I. Meyerson[41] to refer to the situation when states target the transmitter of the opinion (eg, the social media service provider) in order to restrict the actual (primary) author of the expression, ie, the user. Cheap speech on the Internet is made possible by intermediary service providers, but the same service providers can potentially and arbitrarily silence expression as well.[42] Moreover, these technological giants are more in the public eye and easier to identify than users who communicate under pseudonyms or hide behind the veil of anonymity.[43]

Furthermore, with the technological solutions at the social media service providers' disposal, it is clearly more practical for states to encourage them to regulate online expression than directly restrict the users who are the actual authors of the expression. The prospect of holding intermediary service providers liable generates a chilling effect that can (and in practice does) incentivise them to 'over-block' content.[44] The latter, sometimes without distinction, also remove legitimate content from the public discourse if it is problematic (even slightly or seemingly controversial) simply to avoid being held liable.[45]

Private regulation (censorship) by service providers involves the drafting, enforcement and detection of breaches of community standards and other contractual provisions. During the course of such procedures, intermediary service providers monitor and make decisions about the permissibility of the online content (expression) uploaded by everyday users on the basis of contractual provisions with the assistance of automated and human resources that lack a proper legal basis, since the removal, blocking and other restrictions are usually not based on hard law but internal contractual provisions and semi-transparent standards and codes.

The root of the problem, beyond the fact that service providers do not act according to human rights standards, is that they tend to choose to err on the side of caution and

---

39 Gary King, 'Reverse Engineering Chinese Censorship' (2014) Talk at ESRC Research Methods Festival, 5.

40 Roberts (n 33) 41.

41 Michael I. Meyerson, 'Authors, Editors, and Uncommon Carriers: Identifying the "Speaker" Within the New Media' (1995) 71 (1) Notre Dame Law Review 116.

42 Felix T. Wu, 'Collateral Censorship and the Limits of Intermediary Immunity' (2013) 87 (1) Notre Dame Law Review 299.

43 Wu (n 42) 300.

44 Sheera Frenkel, 'Facebook Says It Deleted 865 Million Posts, Mostly Spam' (2018) The New York Times <https://www.nytimes.com/2018/05/15/technology/facebook-removal-posts-fake-accounts.html> accessed 15 October 2024.

45 Wu (n 42) 300.

remove, block or otherwise restrict any questionable content to avoid liability. However, the distinction between lawful and unlawful content is often not clear-cut; it cannot be made on the basis of preset criteria without exploring and understanding the expression's context. Accordingly, much lawful content and many lawful users fall victim to the moderation practices of social media service providers.[46]

V Intermediary Service Provider Liability and the Prohibition of a General Monitoring Obligation according to the ECD and the DSA

To reduce the risks and occurrence of collateral censorship described above, the European Union provides varying degrees of immunity for intermediary service providers depending on their type.[47] Prior to the Digital Services Act (DSA),[48] the regulatory framework for EU digital services was primarily based on the E-Commerce Directive (ECD).[49] As part of the EU legislation, it provided safe harbour immunity under certain conditions specified in Articles 12–14. However, this legislation is not able to prevent the evils of collateral censorship since the ECD does not eliminate private censorship due to its permissive and vague provisions. In fact, the ECD explicitly encouraged service providers to deploy private censorship.[50] Such permissive and encouraging wording was intended to be counterbalanced by the prohibition on imposing a general obligation of monitoring in Article 15, which would effectively amount to Internet censorship.[51] Article 15 explicitly prohibits the imposition of general monitoring obligations with regard to due diligence proceedings on intermediary service providers, ie, the general obligation to detect and prevent illegal content on their platforms.[52]

However, the ECD did not define the meaning of the term 'general monitoring', thereby creating undesirable uncertainties as to how this prohibition (limitation) should

---

46 Wu (n 42) 301.

47 Andrea Kovács, 'A közvetítő szolgáltatások meghatározásának egyes problémáiról' in Gergely Gosztonyi (ed), *A velünk élő történelmi cenzúra* (Gondolat Kiadó 2022, Budapest) 85–96.

48 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) [2022] OJ L277/1.

49 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] OJ L178/1.

50 ECD Article 12(3), Article 13(2) and Article 14(3).

51 Ádám Liber, 'A közvetítő szolgáltatók felelőssége a szellemi tulajdon megsértéséért az Európai Unióban' (2013) 118 (3) Iparjogvédelmi és Szerzői Jogi Szemle 31.

52 Martin Senftleben, Christina Angelopoulos, *The Odyssey of the Prohibition on General Monitoring Obligations on the Way to the Digital Services Act: Between Article 15 of the E-Commerce Directive and Article 17 of the Directive on Copyright in the Digital Single Market* (University of Amsterdam – University of Cambridge 2020, Amsterdam – Cambridge) 6.

be interpreted in relation to intermediary service providers. The findings in the case law of international forums,[53] which also encourage service providers to use automatic mechanisms to avoid liability for third-party content, may also be problematic from the point of view of fundamental rights as automatic solutions such as various algorithm-based filtering or ranking mechanisms are not able to adequately interpret the context of the content in question[54] and thus have a chilling effect on freedom of expression.

In 2020, the European Commission officially presented the DSA proposal as part of a comprehensive digital strategy, and certain provisions concerning the largest platforms became applicable at the end of August 2023.[55] The regulation has been applicable in its entirety since 17 February 2024.[56] The DSA sought to define a uniform set of conditions for all service providers in relation to their exemption from liability and due diligence obligations; however, to avoid disproportionality in the case of smaller service providers, the DSA also applies asymmetric regulation to smaller actors on the market.[57] Even though, as a regulation, the DSA is a set of directly applicable rules that now apply to digital services across the EU, the DSA further confirms[58] the prohibition of general monitoring defined in Article 15(1) of the ECD with the provision remaining relatively unchanged, preserving the previous notice-and-takedown-system (coined as a notice and action mechanism in the DSA). Recital 30 of the DSA indicates that monitoring obligations in specific cases would not be counter to the prohibition defined in Article 8 of the DSA: intermediary service providers should not be, either *de jure* or *de facto*, subject to a monitoring obligation with respect to obligations of a general nature (general active fact-finding obligation, or as a general obligation for providers to take proactive measures in relation to illegal content),[59] but this does not concern monitoring obligations in specific cases and, in particular, does not affect orders by national authorities in accordance with national legislation in compliance with Union law, as interpreted by the Court of Justice of the European Union, and in accordance

---

53 Gergely Gosztonyi, *Censorship from Plato to Social Media. The Complexity of Social Media's Content Regulation and Moderation Practices* (Springer 2023, Cham), 121– 145, DOI: https://doi.org/10.1007/978-3-031-46529-1_9

54 Giancarlo Frosio, Sunimal Mendis, 'Monitoring and Filtering: European Reform or Global Trend?' in Giancarlo Forsio (ed), *The Oxford Handbook of Online Intermediary Liability* (Oxford University Press 2019, Oxford) 21. DOI: https://doi.org/10.1093/oxfordhb/9780198837138.013.28

55 Martin Husovec, 'Rising Above Liability: The Digital Services Act as a Blueprint for the Second Generation of Global Internet Rules' (2024) 38 (3) Berkeley Technology Law Journal 883–920, DOI: https://doi.org/10.2139/ssrn.4598426

56 DSA Article 93.

57 European Commission: Questions and answers on the Digital Services Act, 23 February 2024. <https://ec.europa.eu/commission/presscorner/detail/en/QANDA_20_2348> accessed 15 October 2024.

58 DSA Article 8.

59 Herbert Zech, 'General and specific monitoring obligations in the Digital Services Act' (2021) Verfassungsblog, 2021 <https://verfassungsblog.de/power-dsa-dma-07> accessed 15 October 2024, DOI: https://doi.org/10.17176/20210902-113002-0

with the conditions established in the DSA.[60] As we can see, however, similarly to the preceding ECD, the legislator does not specify what would constitute a specific case.[61] Furthermore, one may find a so-called 'Good Samaritan' clause in the DSA, which stipulates that service providers are not to be held liable if they, in good faith and in a diligent manner, voluntarily carry out own-initiative investigations into or take other measures aimed at detecting, identifying and removing, or disabling access to, illegal content, or implement the measures necessary to comply with the requirements of Union law and national law in compliance with Union law, including the requirements set out in the DSA.[62]

Article 8 would only be an appropriate provision capable of stopping the proliferation of collateral censorship if the distinction between specific and general monitoring obligations were clearly defined and general monitoring was not necessary for exemption from liability.[63] The courts' insistence on maintaining the general-individual distinction is not without reason: the difference is not merely based on economic or proportionality reasons, but rather it is designed to act as a guarantee: 'by exerting pressure and imposing responsibility on those who control the technological infrastructure, [governments] create an environment in which [...] censorship of private partners is an inevitable result'.[64]

VI Is the Oversight Board Able to Counterbalance the Shortcomings of Private Censorship?

The Oversight Board is a 26-member expert group[65] set up by Facebook (Meta) to independently review some of the most difficult and significant content-related decisions of Facebook, Instagram and Threads. It is sometimes referred to as Facebook's 'Supreme Court'.[66] On the one hand, when users have exhausted the relevant platform's appeals process concerning the aforementioned services, they may challenge the latest decision about a piece of content by appealing to the Oversight Board; on the other hand, Meta may

---

60 Valentina Golunova, 'In Tech we Trust? Fixing the Evolutionary Interpretation by the Court of Justice of the Prohibition of General Monitoring in the Era of Automated Content Moderation' in Evangelia Psychogiopoulou, Susana de la Sierra (eds), *Digital Media Governance and Supranational Courts: Selected Issues and Insights from the European Judiciary* (Edward Elgar Publishing 2022, Cheltenham) 62.

61 Gergely Gosztonyi, Andrej Skolkay, Ewa Galewska, Challenges of Monitoring Obligations in the European Union's Digital Services Act. (2024) 12 (1) ELTE Law Journal, DOI: https://doi.org/10.54148/ELTELJ.2024.1.45

62 DSA Article 7.

63 Anupam Chander, 'When the Digital Services Act Goes Global' (2023) 38 (4) Berkeley Technology Law Journal, DOI: https://doi.org/10.15779/Z38RX93F48

64 Frosio, Mendis (n 54) 16; *Delfi AS v Estonia,* no. 64569/09 (ECHR, 16 June 2015), Joint Dissenting Opinion of Judges Sajó And Tsotsoria 17.

65 Oversight Board: Updates On Oversight Board Membership (2023) <https://www.oversightboard.com/news/771690787717546-updates-on-oversight-board-membership/> accessed 15 October 2024.

66 Tamás Pongó, 'Új korszak az online véleménynyilvánítás korlátozásában? Gondolatok a Facebook Oversight Board működéséről' (2020) 4 (147) Iustum Aequum Salutare.

also refer issues to the Board.[67] The Oversight Board started accepting cases in the Autumn of 2020. The main objective of such an experimental body was to oversee the platform's decisions and provide guidance on what content should or might be removed, what should be left online, and why. In other words, it is an ongoing experimental attempt to remedy the harm caused by private censorship.

Without going into the details of the functioning or the content of the Oversight Board's decisions in this paper,[68] I will simply highlight a few key characteristics that underpin or, on the contrary, confirm or refute the Board's ability to counteract the anomalies described in the previous sections of this paper.

With regard to the potential of the Oversight Board, in its almost four years of operation, it has generated a number of benefits that represent a positive step forward for the exercise of freedom of expression online. Based on its track record, it is safe to say that its relevance extends beyond the adjudication of individual cases, and its aim is to establish a precedent system over time that provides a human rights-based, fast and flexible forum[69] for users to challenge the platform's content decisions. In addition to specific decisions, it provides further guidance[70] for the platform to act more in line with human rights standards and laws in the future. By meticulously selecting complex cases, it can also serve as a guide and benchmark for other decisions on probably more straightforward issues, thus impacting the platform's processes significantly more than just deciding individual cases.

However, there are still many areas where the Board lacks real influence that could potentially, with certain improvements, provide real, significant solutions to the problem of private censorship. Despite initial efforts, it does not function as an independent body; it is inseparable from the platform in a number of crucial ways.[71] There are also procedural safeguarding concerns, such as the anonymity of the committee members who decide on individual cases, the lack of due diligence due to the tight deadlines for procedures and the fact that human rights considerations are secondary to 'lex Facebook',[72] which the procedure is primarily based on. The various limitations of the respective procedures are a further concern in terms of the number of cases actually picked out for examination, the decision-making concerning the caseload and demand, and the strict requirements for eligibility to initiate proceedings.[73]

---

67 Oversight Board Charter Article 2, Section 1.

68 Cf: Gergely Ferenc Lendvai, 'A Facebook Ellenőrző Bizottság működése és bíráskodása a gyűlöletbeszéd kontextusában' (2024) 13 (1) In Medias Res, DOI: https://doi.org/10.59851/imr.13.1.11

69 Oversight Board, Bylaws Article 2, Section 2.

70 Oversight Board, Bylaws Article 2 Section 2 and Article 1 Section 4.

71 Oversight Board, Bylaws Article 1, Section 1.1.2., Article 2, Section 1.3.1.

72 Lorenzo Gradoni, 'Constitutional Review via Facebook's Oversight Board' (2021) Verfassungsblog, <https://verfassungsblog.de/fob-marbury-v-madison> accessed 15 October 2024, DOI: https://doi.org/10.17176/20210210-235949-0

73 Rebecca Heilweil, 'You can finally ask Facebook's oversight board to remove bad posts. Here's how' (2021) Vox, <https://www.vox.com/recode/22381607/facebook-oversight-board-appeal-remove-post-ads> accessed 15 October 2024.

In terms of the effectiveness of the available legal remedy, access is of decisive importance, and based on this criteria, the Oversight Board, from the point of view of average users, lacks potential: the Board has complete discretion as to which matters it deliberates and adopt a decision on. As of July of 2024, the Board has only delivered 107 decisions,[74] while during the course of its operation until Q2 of 2023 more than 2.7 million cases were submitted[75] to it. This means that only 0.004% of the cases submitted to the Board have been actually decided by the Board. The various jurisdictional limits, such as restrictions on the type of explicit content (eg, spam), the type of decision (eg, decisions concerning intellectual property), and certain services (eg, Messenger),[76] severely limit users' ability to submit cases to the Board and completely exclude the possibility of review in many cases.

The Oversight Board can, in its binding decisions, instruct Facebook to remove or keep certain content intact and online, as well as to amend its decisions, which the platform is obliged to implement within seven days. In fact, according to Section 4 of the Articles of Association, Facebook is required to search for content identical to the content affected by the decision and apply the same procedure as was the subject of the original decision if the necessary technological tools and organisational resources are available.[77] This can be compared to the emerging trend related to the prohibition of general monitoring obligations,[78] according to which platforms are expected not only to take certain steps in the event of specifically contested matters but to search for specific harmful content that occurs in the same context and act against this as well.

VII Conclusion

On the one hand, social media has given people unprecedented opportunities and a unique platform to express their opinions and receive information, but it has also transformed the institution of censorship that was established during the era of traditional media.

Censorship was once concentrated in the hands of states, but now, in the social media age, the role of the censor has been partially taken over by a number of private entities with economic power and technological tools, which arbitrarily define the limits of freedom of expression. The state's desire to restrict social discourse has not ceased, but how it is achieved has been transformed. This has resulted in a system in which a number of (new) actors have emerged (alongside the actual person expressing their opinion) that, on the one hand, are able and keen on restricting freedom of expression and, on the other, forced to

[74] See <https://transparency.meta.com/oversight/oversight-board-cases> accessed 15 October 2024.
[75] Oversight Board, 2023 Q2 Transparency Report.
[76] Oversight Board, Bylaws Article 2, Section 1.2.1.
[77] Oversight Board, Bylaws Article 1, Section 2.3.1.
[78] Senftleben, Angelopoulos (n 52).

cooperate in unprecedented ways, thus increasing the presence of censorship as a threat to freedom of expression. Instead of the single bipolarity of relationships that used to exist regarding traditional media, there is now a complex network of multiple actors that threaten freedom of expression, either directly or indirectly (private censorship of platform providers and public self-censorship) or through reverse censorship. What these different methods of restriction have in common, however, is that they can be identified as new 21st-century forms of censorship, as they are likely to involve the deliberate and systematic curtailment of fundamental rights such as freedom of expression guaranteed by law, driven by the opaque, unforeseeable, arbitrary economic and other interests of a private company or multiple private companies.[79]

Accordingly, today, private censorship is no longer just a theoretical distant possibility; it is a reality in our everyday lives. Social media service providers themselves have the ability and means to decide about any content on their platforms, but they are dangerous actors in relation to states in an age when traditional censorship methods no longer provide sufficient solutions to limit freedom of expression.[80] The combination of these factors creates a new situation for democratic discourse and forums – an environment in which freedom of expression can be restricted by multiple actors, driven by a great variety of motives, often without any or at least without sufficient procedural safeguards, and mainly (especially in the case of flooding) without being perceived by the actual author of the expression, thus posing an unprecedented threat to freedom of expression.

[79] Gábor Megadja and others, *A Facebook-cenzúra ellen* (Századvég 2019, Budapest) 42.

[80] Folkert Wilman, 'Two emerging principles of EU internet law: A comparative analysis of the prohibitions of general data retention and general monitoring obligations' (2022) 46 (1) Computer Law & Security Review, DOI: https://doi.org/10.1016/j.clsr.2022.105728