Revista catalana de dret públic #61

www.rcdp.cat

FREEDOM OF EXPRESSION AND DIGITAL PLATFORMS: CHALLENGES FOR CONTENT REGULATION IN A GLOBAL ENVIRONMENT*

Joan Barata Mir**

Abstract

The fundamental rights to freedom of expression and freedom of information are recognized, protected and delimited in practically all the constitutions of the political systems of liberal democracies, including the Spanish Constitution. These rights are also part of the human rights framework established by the rules of international law. Intermediaries have become the main actors in the process of disseminating and distributing all types of content. They also play a very important role in ensuring the "visibility" of content produced by traditional media. The interests that modulate the decisions taken by intermediaries are not solely and necessarily the protection of the right to freedom of expression. Intermediaries are commercially oriented entities and follow business models based on advertising, among other things. The most influential legal systems in the world (particularly those in the European Union and the United States) encourage the moderation of content by platforms, ensuring that they are not penalized as a result of content provided by third parties or due to the adoption of certain measures aimed at eliminating illegal or objectionable content. For these reasons, intermediaries play a very important role in creating the conditions for the exercise of the right to freedom of expression online.

Key words: freedom of expression; content regulation; liability of intermediaries; self-regulation of media.

LLIBERTAT D'EXPRESSIÓ I PLATAFORMES DIGITALS: REPTES DE LA REGULACIÓ DE CONTINGUTS EN UN ENTORN GLOBAL

Resum

Els drets fonamentals a la llibertat d'expressió i a la llibertat d'informació es troben reconeguts, protegits i delimitats en la pràctica totalitat de les constitucions dels sistemes polítics de les democràcies liberals, entre elles la Constitució espanyola. Aquests drets formen igualment part del quadre fonamental de drets humans establert per les normes de dret internacional. Els intermediaris s'han convertit en els principals actors del procés de difusió i distribució de tot tipus de contingut. També tenen un paper molt destacat a l'hora d'assegurar la "visibilitat" del contingut produït pels mitjans tradicionals. Els interessos que modulen les decisions preses pels intermediaris no són únicament i necessàriament la protecció del dret a la llibertat d'expressió. Els intermediaris són entitats orientades comercialment i segueixen models de negoci basats en la venda de publicitat, entre altres aspectes. Els sistemes legals amb més influència en el món (particularment els de la Unió Europea i els Estats Units) incentiven la moderació del contingut per part de les plataformes, garantint que no són penalitzades com a conseqüència del contingut subministrat per tercers o a causa de l'adopció de determinades mesures orientades a eliminar contingut il·legal o contingut inadequat. Per aquests motius, els intermediaris tenen un paper molt important en la creació de les condicions per a l'exercici del dret a la llibertat d'expressió en línia.

Paraules clau: llibertat d'expressió; regulació de continguts; responsabilitat d'intermediaris; autoregulació de mitjans.

* This article is a translation of an original one in Catalan.

** Joan Barata Mir, Associate Researcher at the Cyber Policy Center of the University of Stanford, former Secretary General of the Catalan Audio-Visual Council and former Principal Adviser to the Representative on Freedom of the Media to the OECD. Freeman Spogli Institute for International Studies. Encina Hall. 616 Jane Stanford Way. Stanford, CA 94305-6055, USA). jbaratamir@gmail.com, @JoanBarata.

Article received: 13.09.2020. Blind review: 25.09.2020 and 04.10.2020. Final version accepted: 07.10.2020.

Recommended citation: Barata Mir, Joan. (2020). Freedom of expression and digital platforms: challenges for content regulation in a global environment. *Revista Catalana de Dret Públic*, *61*, 1-13. https://doi.org/10.2436/rcdp.i61.2020.3519.

Summary

1 Introduction

In today's world, protecting freedom of expression calls for an awareness of the technological changes affecting activities involving the search for and the provision and receipt of all manner of information, ideas and opinions. Even more important, however, is a proper understanding of the various actors and types of regulations involved in defining the system governing the dissemination of online content, particularly via the so-called "intermediary" platforms or services.

This system is based on the concurrence of principles stemming, firstly, from international human rights regulations and organisations, as well as the laws adopted by national authorities and, secondly, by the rules and standards established by the very platforms themselves to provide a framework of civility for user dialogue and coexistence.

This article seeks to analyse the main legal issues and problems associated with this field.

2 Freedom of expression: a right protected both online and offline

The fundamental rights of freedom of expression and freedom of information are recognised, protected and delimited in almost every liberal democratic constitution, including that of Spain (in its Article 20). They also form a nominal part of the constitutions of other kinds of regimes, of an authoritarian or semi-authoritarian nature, although, in these latter cases, these rights are often fairly restricted due to broad-based limits or completely open references to the will of the legislator.

These rights are also part of the core human rights recognised by international law, specifically, the Universal Declaration of Human Rights (Article 19) and, more comprehensively, in Article 19 of the International Covenant on Civil and Political Rights, both adopted by the United Nations. It is worth quoting the content of the latter:

> 1. Everyone shall have the right to hold opinions without interference.
>
> 2. Everyone shall have the right to freedom of expression; this right shall include freedom to seek, receive and impart information and ideas of all kinds, regardless of frontiers, either orally, in writing or in print, in the form of art, or through any other media of his choice.
>
> 3. The exercise of the rights provided for in paragraph 2 of this article carries with it special duties and responsibilities. It may therefore be subject to certain restrictions, but these shall only be such as are provided by law and are necessary:
>
> (a) For respect of the rights or reputations of others;
>
> (b) For the protection of national security or of public order, or of public health or morals.

Aside from Article 19, the Covenant also includes an Article (20) that was highly controversial during the drawing up of this legal instrument, given that some – mostly liberal – countries believed that Article 19 already contained all the provisions necessary to manage any abuse or excess involving the exercise of freedom of expression. However, in the end, negotiations concluded in favour of those proposing the prohibition, in a specific article, of certain forms of expression (Kearney, 2007: 128, 131). It is worth noting that the incorporation of this Article into international human rights law has given rise to numerous interpretative debates, especially with regard to the notion of "propaganda for war" (paragraph 1) and the terms under which paragraph 2 establishes the basic terms of what has become known as "hate speech":[1]

> 1. Any propaganda for war shall be prohibited by law.
>
> 2. Any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence shall be prohibited by law.

1 See the "non-paper" from the Office of the Representative on Freedom of the Media of the Organization for Security and Co-operation in Europe (OECD) entitled "Propaganda and Freedom of the Media".

It is important to note a series of key points arising from this international legal framework.

The first is that, leaving to one side for the moment the specific impact of Article 20, Article 19 contemplates the imposition of certain limitations on freedom of expression. However, it should be borne in mind that what is particularly characteristic of this precept (paragraph 3 in particular) is that it precisely defines the requirements upon which national authorities may actually establish these limits or conditions. In other words, Article 19 establishes the *limits of the limits* on freedom of expression.

More specifically, any restrictions, regulations and limitations affecting the rights to the freedoms of expression and information must be clearly and specifically established by the legal system, be designed to protect the rights and interests set forth in paragraph 3 of Article 19 and respect the principles of necessity and proportionality. In other words, the measures in question must be strictly necessary and proportional and entail the minimum possible restriction. This "triple test"[2] is an obligation arising from international law and is, therefore, applicable to all public policies, laws and regulations adopted at a national level. What is more, this test is also applied by regional human rights courts, using the associated instrument (in the case of Europe, for example, the jurisprudence of the European Court of Human rights based on the provisions of Article 10 of the European Convention on Human Rights) (Bychawska-Siniarska, 2017).

Secondly, the right to freedom of expression is recognised and protected "regardless of frontiers". This has significant implications in the age of the Internet, which may finally allow this potentially global and unlimited dimension of freedom of expression to become reality. Obviously, the Internet takes the distribution and dissemination of ideas, opinion and information to a whole new level compared with that of the (mainly national) legal regimes governing the traditional media.

Thirdly, Article 19 protects the exercising of freedom of expression by anyone and via any medium or technology, not only those existing at the time of the adoption of the Covenant (1966). As the UN's Human Rights Committee states:[3]

> Journalism is a function shared by a wide range of actors, including professional full-time reporters and analysts, as well as bloggers and others who engage in forms of self-publication in print, on the Internet or elsewhere.

Whatever the case, it is clear that we are dealing with a wider definition of the concepts of the media and journalism, and this has been the subject of new debates within the field of definition of international standards (Richter, 2016).

It should be remembered in this regard that the UN Human Rights Council declared, in Resolution 32/13, of 11 July 2016, that "[...] the same rights that people have offline must also be protected online, in particular freedom of expression, which is applicable regardless of frontiers and through any media of one's choice, in accordance with articles 19 of the UDHR and ICCPR", which is exactly the same line established by Resolutions 20/8, of 5 July 2012, and 26/13, of 26 June 2014, on the promotion, protection and enjoyment of human rights on the Internet.

3 The role of intermediaries and content moderation

In 2018, the Committee of Ministers of the Council of Europe (CoE) adopted a Recommendation on "the roles and responsibilities of Internet intermediaries",[4] which described them as a wide, diverse and rapidly evolving range of players with important functions that:

> facilitate interactions on the internet between natural and legal persons by offering and performing a variety of functions and services. Some connect users to the internet, enable the processing of information and data, or host web-based services, including for user-generated content. Others aggregate information and enable searches; they give access to, host and index content and services designed and/or operated by third parties.

2 United Nations Human Rights Committee, 102nd session. Geneva, 11-29 July 2011. General comment no. 34.

3 General comment no. 34, above.

4 Recommendation CM/Rec(2018)2, of 7 2018, of the Committee of Ministers to member States on the roles and responsibilities of internet intermediaries.

> Some facilitate the sale of goods and services, including audio-visual services, and enable other commercial transactions, including payments.

Intermediaries have become the leading players in the process of disseminating and distributing all manner of content online. They also have a very important role in ensuring the "visibility" of the content created by traditional media, such as the written press or even radio and television. Indeed, it is increasingly frequent to see this kind of content consumed as the result of a recommendation or it appearing on social media on all kinds of platforms, rather than being directly accessed by users.[5]

Nevertheless, it is also clear that the interests affecting the decisions taken by intermediaries (and particularly content-related ones) are not solely and necessarily associated with the protection of freedom of expression. These intermediaries are obviously businesses with commercial interests and follow business models based on the sale of advertising space (and this, can, on occasion, lead to the promoting of socially undesirable or harmful discourse). Bietti (2020) warns that intermediaries promote content to maximise user engagement, addiction, behavioural targeting, and polarisation. Platforms have tried to counter this kind of argument by proclaiming their commitment to providing users with content that is "meaningful" to them, above and beyond a classification system based on what people like and click on.[6]

There are also other significant pressures, not always economic in nature: there is a need to show responsibility when detailing with "problematic content" within the framework of differing societies, managing the pressures from parliamentary committees, the media and public opinion, dealing with "threats" of regulation by governments and legislatures and being able to provide an appropriate response to the behaviour of business rivals (Keller, 2018b). At the same time, the relevant international bodies and civil societies demand that the platforms comply with international human rights law not only when it comes to moderating content, but also when taking into account the legislation or requirements set by relevant national authorities. For example, in his 2018 thematic report to the Human Rights Council, David Kaye, UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, directly addressed platforms, demanding that they recognise that "the authoritative global standard for ensuring freedom of expression on their platforms is human rights law, not the varying laws of States or their own private interests, and they should re-evaluate their content standards accordingly".[7]

In any case, it must be noted that, however hard many platforms do work to prevent the presence of undesirable content (be this due to breaking the law or violation of internal behaviour standards), such platforms regularly make significant errors in classifying supposedly harmful content and taking decisions on it (York and Gullo, 2018; Perrigo 2020a, 2020b). In this area, the use of artificial intelligence mechanisms to moderate the aforementioned content is becoming an almost unavoidable requirement, in a context in which platforms are obliged to apply laws and internal standards on a large scale. In other words, without prejudice to the fact that, in certain cases (especially after obtaining a provisional "positive"), some content is subject to closer review by human moderators, platforms feel the need to detect improper content quickly and "in bulk", which makes it physically impossible to implement, in every single case, the kinds of analysis and evaluation systems that would be employed by an administrative or judicial authority.

One particularly graphic example of this is that of content featuring what we could call violent extremism or acts of terrorism. In view of situations such as the appearance of propaganda videos by the terrorist organisation ISIS on some platforms, or even the live streaming by terrorists themselves of their attacks, the leading platforms have developed content matching techniques using a digital fingerprint or "hash" database. Every time a platform identifies and removes content because it is regarded as terrorist-related, this fingerprint is added and the content is shared on the common database, available to the other partners in the initiative. More specifically, the Global Internet Forum to Counter Terrorism (GIFCT)[8] was created in 2017 by Facebook, Twitter, Microsoft and YouTube as an industry initiative to "prevent terrorists and violent

5 See, for example, the 2018 Pew Research Center report "News Use Across Social Media Platforms".

6 See Ezra Klein's interview with Mark Zuckerberg, "Mark Zuckerberg on Facebook's hardest year, and what comes next". *Vox*, 2 April 2018.

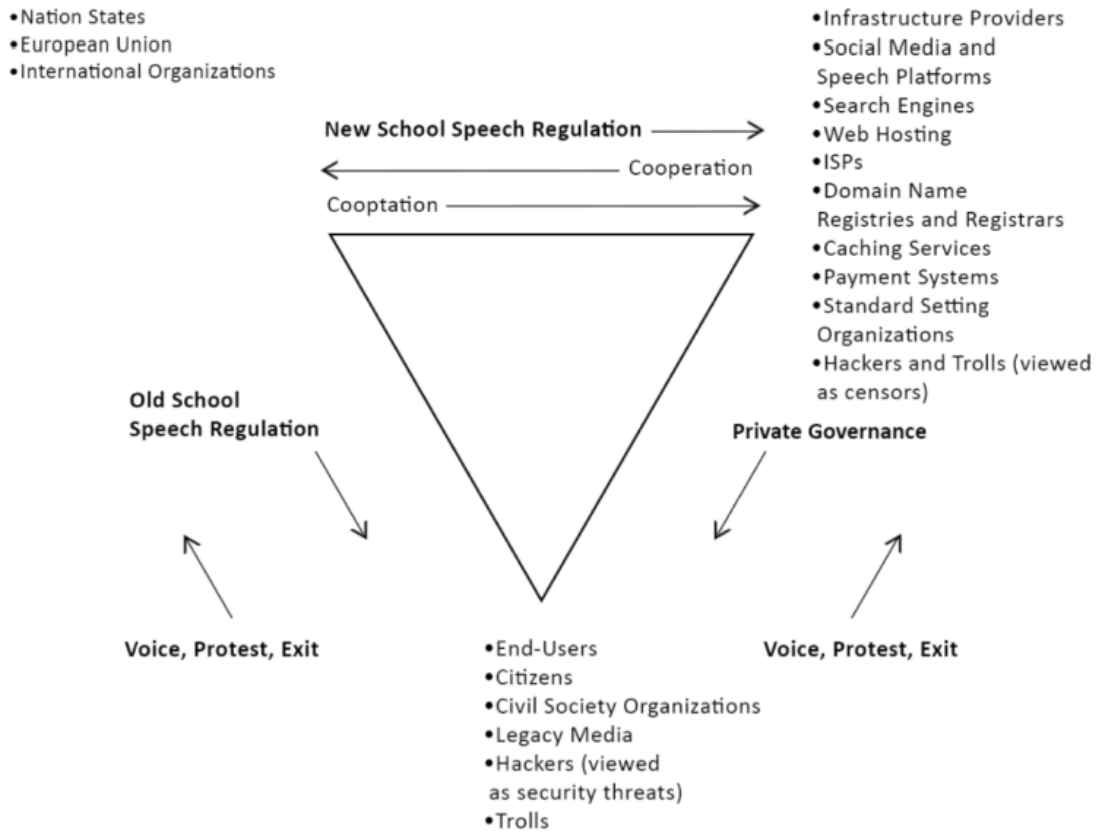
7 Available online at: www.ohchr.org/EN/Issues/FreedomOpinion/Pages/ContentRegulation.aspx.

8 https://www.gifct.org.

extremists from exploiting digital platforms". GIFCT is now on the way to becoming an organisation that is managed independently, rather than by the platforms themselves. Nevertheless, concerns around the lack of transparency, vulnerability to government pressure and the impact on civil groups activities (due to the consequences of possible errors or "false positives") make it a fairly controversial tool. Whatever the case, the impact of this database on freedom of expression is of great importance and, as some authors have noted, it is "possibly the most important acronym you've never heard of", but one that may, nevertheless, radically change the way the Internet is governed (Radsch, 2020).

It is against this general backdrop that Balkin (2018) created the well-known "free speech triangle" (Figure 1): on one corner are the nation states and international organisations, on another Internet companies (social media, search engines, ISPs, etc.) and, on the third, users, civil society and the traditional (legacy) media. This triangle illustrates how the relationship between the authorities and users and traditional media is the typical power relationship featuring a combination of familiar legal instruments and public or business protest or advocacy. For its part, the relationship between Internet companies and national bodies is more complex, due to the novelty of the area and to the problems of guaranteeing the application of the relevant laws. This is why the author makes reference to "new-school" regulation, also referred to in some environments as "smart regulation",[9] based on cooperation and co-regulation mechanisms, and in which the risks of possible co-optation by the public authorities would be greater than in traditional regulation. Lastly, as already noted, the relationship between Internet companies and the different kinds of users would be a closer fit with the idea of governance or private regulation.

Figure 1: The Pluralist Model of Speech Regulation[10]



Source: Balkin (2018).

9 See the February 2019 proposal made by a Google Senior Vice-President "Smart Regulation for combating illegal content".

One example of the complex interaction between platforms and different levels of political and regulatory powers with regard to online content is to be found in the European Union, within the framework of initiatives encouraging self-regulation. Specifically, Facebook, Microsoft, Twitter and YouTube (joined in September 2020 by TikTok) signed a code of conduct with the European Commission in May 2016 on countering illegal hate speech online.[10] This initiative is associated with the Communications and Recommendations formulated previously by the European Commission on this matter.[11] Similar initiatives have also taken place to tackle disinformation.[12]

Turning to national legislation, it is clear that global platforms also have to take into account and apply national laws and the decisions of national authorities (such as regulatory bodies and judges), as the content published and consumed by users can normally be associated with a specific territory. To give one example, on 1 January 2018, Germany saw the coming into force of its *Netzwerkdurchsetzungsgesetz* (Network Enforcement Act or NetzDG).[13] This law was subject to great debate and criticism, particularly by organisations defending fundamental rights, as they considered it disproportionate, inappropriate and promoting the private censorship of the Internet.[14] Briefly, NetzDG forces platforms (those with more than two million users in Germany) to provide users with mechanisms so that they may inform of a broad range of categories – 22, to be precise – already included in the country's criminal legislation. Once the relevant complaint has been received, it is up to the platform to establish whether the content really is illegal. Pursuant to NetzDG, platforms must remove cases of "obviously illegal" content within 24 hours. In more complicated cases, removal must take place within seven days. Breach of these obligations is punishable with fines of up to 50 million euros. Platforms are also obliged to publish "transparency reports" on content moderation decisions adopted.

NetzDG can be regarded as a truly unique law due both to the kind of obligations it imposes upon this sort of provider and to its scope and depth and the seriousness of the penalties it contemplates in cases of violation of its provisions. Although confirmation of NetzDG's effectiveness as a tool for combatting and eliminating certain forms of illegal online content still awaits reliable supporting data, the fact is that its fame has spread worldwide and it has been explicitly used as inspiration by authoritarian and semi-authoritarian regimes in other parts of the globe to justify significantly restrictive rules (Mchangama and Alkiviadou, 2020). These regimes include Russia, Venezuela, Belarus, Honduras, Malaysia, Singapore and the Philippines.

A law like NetzDG brings with it an obvious risk of the excessive removal of content. If a sword of Damocles in the form of huge fines is hanging over platforms in cases in which specific content is not correctly regarded as illegal, it is clear that these firms will feel pushed towards removing any content raising the slightest suspicion of being illegitimate. The impact on freedom of expression is obvious, particularly if we bear in mind the fact that we are not actually dealing with a purely private decision, but a limitation arising from a legal requirement, compliance with which is ultimately verified by the competent authorities. These cases entail nothing short of the exercising of the power to determine the legitimate scope of freedom of expression and – more importantly – acting accordingly, particularly with regard to the removal of the relevant content, ending up delegated to private corporations. Such a power had previously only been exercised by the courts or by regulatory authorities. Whilst it is true that all these decisions may be submitted to jurisdictional oversight, it will be difficult to offset the impact of the immediate removal of specific content with any subsequent judicial decision upholding an appeal by its creator.

There is also another emerging trend that would go further than the aforementioned regulatory framework, and that represents the adoption of legal rules aimed at obliging platforms to not only comply with general legal restrictions in fields such as hate speech, the dissemination of terrorist content or child protection, but also to adopt additional rules and principles that would entail, paradoxically, the legal requirement to implement rules restricting freedom of expression that would actually go beyond the limits of the law.

10 Available online at: https://ec.europa.eu/newsroom/just/item-detail.cfm?item_id=54300.

11 Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, tackling illegal content online and towards an enhanced responsibility of online platforms COMM (2017) 555 final (28 September 2017) and Commission Recommendation on measures to effectively tackle illegal content online, of 1 March 2018, C (2018) 1177 final.

12 See the different initiatives on the matter and, in particular, the code of practice signed with technology companies.

13 Published in the Federal Gazette on 1 September 2017, pp 3,352ff.

14 See the critical analysis by *ARTICLE 19*, the association for the global defence of freedom of expression.

The UK's initiative aimed at regulating so-called "online harms" is a good example of this trend. In April 2019, the British government published a broad-ranging white paper on the matter, which was the object of extensive public consultation.[15] February 2020 saw the publication of an initial response to the contributions received, which provides more specific details in some areas.[16] The main principle underlying this initiative is the imposition and verification by a specialist regulator of a "duty of care", on the basis of which platforms must adopt a series of measures, via their internal codes and standards, designed to eliminate or mitigate different kinds of content that is not only illegal, but also what the regulator regards as "harmful", that "threatens our way of life in the UK" or "unacceptable": in other words, content that "even if they may not be illegal in all circumstances, can also cause serious harm". More particularly, the initiative shows concern and a wish to deal with the issue of online disinformation, especially that coming from "hostile actors" using it to undermine "our democratic values".

In short, this is a model for the regulation of online content that, once again, includes significant "delegated powers" for the platforms with regard to the establishment of limits on the fundamental right of freedom of expression. Specifically, these powers must be exercised with regard to certain categories classified by the law as "problematic", but which it neither lists nor defines in detail, given that this is an evolving task that needs to be performed by an independent regulator based on the particular circumstances and risks existing at any given time within society. Whilst it is true that, in such cases, the final decision lies with the platform in question, it shall be the regulator's responsibility to ensure that said platform has undertaken "structural" measures or the proper procedures to detect and deal with problematic content, and that they are applied consistently and transparently.

In any case, there is great doubt implicit in this model as to what is or is not acceptable on any given platform. This issue, and the specific mechanism for content moderation eventually adopted by each platform, is left more or less up to an administrative regulatory authority. Such uncertainty does not create a favourable framework for exercising freedom of expression. Quite the contrary, in fact (Smith 2020a, 2020b).

4 Intermediaries and liability exemption systems

The world's most influential legal systems (particularly those of the European Union and the United States) encourage platforms to moderate content, guaranteeing that they will not be penalised for that supplied by third parties or due to the adoption of measures designed to remove illegal or inappropriate content.

In the United States, Section 230 of the Communications Act of 1934 (as amended by the Telecommunications Act of 1996) protects platforms from liability for third-party content. It also provides intermediaries with immunity with regard to any decision on content moderation, insofar as: a) it has been adopted "in good faith", and b) it refers to content or, more broadly, information that the platform regards as "objectionable". Section 230 has played a crucial role in the development of the Internet as we know it. Under the protection afforded by US legislation, platforms have had the incentive to run and expand their businesses under a predictable legal system and have had the capacity to intervene with regard to the information they distribute, especially in the case of content that is allegedly illegal or likely to be regarded as harmful in accordance with the internal standards established by each platform (Goldman, 2019).

In Europe, Articles 14 and 15 of the so-called Directive on electronic commerce[17] contain provisions that also affect content moderation by intermediaries. According to said provisions, platforms shall not be responsible for third-party content provided that a series of requirements and conditions are met. To retain their immunity, platforms cannot have actual knowledge of illegal activity or information and/or cannot be aware of any facts or circumstances from which the illegal activity or information is apparent (for civil liability claims). Additionally, EU legislation (specifically, Article 15 of the same Directive) also prohibits general content-monitoring obligations or obligations to actively seek facts or circumstances indicating illegal activity. In any

15 Available online at: https://www.gov.uk/government/consultations/online-harms-white-paper/online-harms-white-paper.

16 Available online at: https://www.gov.uk/government/consultations/online-harms-white-paper/public-feedback/online-harms-white-paper-initial-consultation-response.

17 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.

case, intermediaries can voluntarily adopt (as noted above) their own content moderation and oversight rules and enforce them without necessarily forfeiting said exemption (Barata Mir, 2020a).

So, current law in both the United States and the European Union (as well as other countries around the world, such as India and a number of Latin American nations) protect – to different degrees – the ability of platforms to decide how to organise, prioritise, downgrade or simply remove content. This protection is provided through exemptions from legal liability.

This legal framework means that platforms actually have a comprehensive range of internal rules that go much further than the legal limits generally established within the field of freedom of expression. Good examples of these are Facebook's Community Standards,[18] Twitter's rules and policies[19] and YouTube's Community Guidelines.[20]

This forms part of a very significant debate, particularly since the European Commission has undertaken to submit a draft Digital Services Act (DSA) package[21] that, amongst other things, would replace the provisions of the Directive on electronic commerce. Amongst other matters, this debate is focused on whether increasing the "editorial" control of intermediaries over content by means of moderation practices may justify the introduction of legal amendments that affect the degree of responsibility for third-party content. In other words, whether it would be possible, liability-wise, to go further in the distinction already pointed to (albeit a little vaguely) by the European Court of Justice between intermediaries hosting and/or providing content in a completely passive way and those that are somehow involved in the information supplied – still – by third parties.[22]

Previously, the European Union had already adopted some Directives directly affecting the system governing intermediary liabilities, with important debates around their impact on freedom of expression also taking place.

Directive (EU) 2019/79 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 contains – particularly in its Article 17 – a series of provisions that would compel platforms to play a leading role in detecting unlawful content as soon as it is uploaded by users, thereby creating a monitoring duty with a possible negative impact on freedom of expression (Quintais et al., 2019).

For its part, Directive (EU) 2018/1808 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, includes a series of duties applicable to audiovisual content hosted by the so-called video sharing platforms (such as YouTube). These duties affect areas such as the prevention and moderation of content constituting hate speech and child pornography that may impair the physical and mental development of minors, that breach obligations with regard to commercial communications or that can be regarded as incitement to terrorism. National authorities (principally independent regulatory bodies) are now responsible for ensuring that platforms have adopted "appropriate measures" to properly deal with the aforementioned types of content. This includes a guarantee that platforms review and properly apply their internal rules, have appropriate content reporting mechanisms, implement user age verification or classification and control systems, have transparent, easy-to-use and effective procedures for resolving users' complaints and provide media literacy tools. The Directive also allows the competent authorities to impose upon platforms the duty to restore content mistakenly removed due to incorrect application of the relevant legal criteria. These provisions give rise to a considerable number of doubts and uncertainties from the standpoint of their effective application (Barata Mir, 2020b).

18 https://www.facebook.com/communitystandards.

19 https://help.twitter.com/en/rules-and-policies.

20 https://www.youtube.com/howyoutubeworks/policies/community-guidelines.

21 https://ec.europa.eu/digital-single-market/en/digital-services-act-package.

22 See, principally, the following cases: L'Oréal (Judgement of 12 July 2011, case C-324/09) and Google France (Judgement of 23 March 2010, cases C-236/08 to C-238/08).

5 Platform users' freedom of expression

As noted above, both the Council of Europe and other international authorities have placed special emphasis on the fact that intermediaries play a highly important role in the creation of the conditions for the exercising of the right to freedom of expression online.

This viewpoint, based on the idea that intermediaries facilitate and permit the exercising of the right to freedom of expression by its users, cannot be found – with the same meaning and implications, at least – in the US legal system. More specifically, there is consolidated jurisprudence that holds that citizens or users may not exercise their First-Amendment right to freedom of expression in the light of decisions (the removal of certain content or even the deactivation of an account) taken by an intermediary. Quite the contrary, actually. This, in fact, forms part of the right to freedom of expression held by platforms themselves: the right to decide what content they host and supply to third parties, and what they don't (Keller, 2018a; Goldman, 2020a). One recent and particularly illustrative case of this is the judgement in the Prager University v. Google LLC case,[23] which analyses YouTube's decision to restrict access to a series of videos published by the organisation Prager University and demonetise them. Specifically, the court notes the following:

> Despite YouTube's ubiquity and its role as a public-facing platform, it remains a private forum, not a public forum subject to judicial scrutiny under the First Amendment […] PragerU runs headfirst into two insurmountable barriers—the First Amendment and Supreme Court precedent […]. The Internet does not alter this state action requirement of the First Amendment.

As we can see, the court rejects any possible use of the term "public forum", a place in which the expressions of any participant would find room and be protected. This concept is taken from a US Supreme Court decision upholding a judgement by the California Supreme Court,[24] which had raised this concept with regard to the possibility of expressing oneself freely in a publicly accessible space, specifically a shopping centre. It should be borne in mind, nevertheless, that this idea is taken not from the First Amendment, but from the Constitution of the State of California.

In other words, according to the PragerU precedent, the fact that intermediaries provide an open space accessible for talking does not strip said space of its intrinsically private nature or create any obligation to protect users' freedom of expression.

The debate is more complex outside the United States, mainly in Europe.

In Germany, the Munich Regional High Court ruled that a comment published by a right-wing politician that had, according to Facebook, violated its content rules, constituted the exercising of freedom of expression protected by the country's constitution. The Court therefore ordered Facebook to republish the comment and re-establish the user's account, which had been suspended for 30 days. It is unclear whether this decision will be upheld by higher courts, but it is important in that the first instance has established that Facebook users' freedom of expression must be protected in terms essentially equal to those applicable to national institutions, and that when intermediaries enforce their internal content moderation rules, they need to take into account (at least) these constitutional rules.[25]

Similar cases have occurred in other European countries and in other parts of the world, although they cannot be said to reflect a consolidated jurisprudential focus generally upheld by higher courts. The most recent and important example at the time of writing this article is the decision by a Rome civil court to order Facebook to reactivate the account of the far-right party CasaPound and pay it €800 for every day the account remained inactive following the order. As with the German case above, the account had allegedly violated Facebook's internal policies on hate speech. The court nevertheless held that the platform's decision represented an inacceptable exclusion or restriction of CasaPound's right to participate in Italian political debate.[26]

23 Prager University v. Google LLC, case no. 18-15712 (9th Circuit, 26 February 2020).

24 PruneYard v. Robins, 447 U.S. 74 (1980).

25 See a description and analysis of this case on the Columbia University Global Freedom of Expression Database.

26 See a description and analysis of this case on the Columbia University Global Freedom of Expression Database.

So it is that Europe provides us with cases in which the courts have ordered some platforms to adopt positive measures to protect certain content, based mainly on the argument that the intermediaries in question play a crucial role in facilitating the effective exercise of fundamental rights (particularly freedom of expression and the right to non-discrimination), or have an impact on basic national and European constitutional principles (such as pluralism) (Kettemann and Tiedeke, 2020). Clearly, the issue remains very open and is richly deserving of further examination by national constitutional courts and the European Court of Human Rights.

So, debate in Europe around the existence of any possible "positive" obligations on the part of states to protect freedom of expression against restrictive content moderation decisions taken by intermediaries is still at a very early stage, although some scholars have pointed to interesting connections between the principle of pluralism and diversity typical of the traditional system governing media such as radio and television, and the introduction of possible restrictions and obligations around how platforms manage content. Such restrictions and obligations would not principally entail the removal of illegal content, but rather the promotion of the free dissemination of and access to certain kinds of content (Helberger et al., 2019).

Authors such as Leerssen (2015) accept that, in general terms, the identification of direct state action currently remains essential for exercising or defending the right to freedom of expression before online service providers, particularly when this is carried out before courts or regulatory authorities. He also calls upon the Court of Justice of the European Union to further explore the horizontal dimension of free speech rights online, and depart from a strict distinction between public and private responsibilities in considering possible legal actions. There is no doubt that this is an issue that may lead to changes in some of the traditional constitutional fundamentals and focuses regarding the notion of and methods for protecting freedom of expression in our societies, which introduces us to what has been dubbed "digital constitutionalism" (De Gregorio, 2020).

6 Conclusions

As we have seen, there are currently numerous tensions between the different groups or sets of laws we have mentioned. Countries often regulate online content in accordance with their own priorities and interests, ignoring the guidelines and limits established by international human rights laws. This also leads to the creation of normative "compartments" that segment and impair the free flow of information and opinions across platforms with global scope, giving rise to what is commonly called the "Balkanisation" of the Internet or the "Splinternet". We can also see how, aside from establishing the aforementioned legislative frameworks, there is a growing trend of countries influencing the way the intermediary platforms themselves establish the principles and standards governing users' behaviour and the limits on expression within the specific framework of the particular platform. These limits are generally greater in scope than those defined by applicable law.

Finally, and related to the above, we are currently witnessing the opening shots of a debate as to whether public authorities (be they national or international) may establish and enforce mechanisms for protecting the fundamental rights of users of platforms, with regard to the decisions and measures adopted by the latter.

Reference list

Balkin, Jack. (2018). Free Speech is a Triangle. *Yale Law School, Public Law Research Paper, 640*.

Barata Mir, Joan. (2020a). Positive Intent Protections: Incorporating a Good Samaritan principle in the EU Digital Services Act. *Center for Democracy and Technology*.

Barata Mir, Joan. (February 25, 2020b). Regulating content moderation in Europe beyond the AVMSD. [Blog post]. *LSE Blog*.

Bietti, Elettra. (June 1, 2020). Free Speech is Circular. [Blog post]. *Medium*.

Bychawska-Siniarska, Dominika. (2017). *Protecting the right to freedom of expression under the European Convention of Human Rights. A handbook for legal practitioners*. Strasbourg: Council of Europe.

Goldman, Eric. (2019). Why Section 230 is better than the First Amendment. *Notre Dame Law Review Reflection*, *95*(1), 33-46.

Goldman, Eric. (March 2, 2020). YouTube Isn't a State Actor (DUH) – PragerU v. Google. [Blog post]. *Technology & Marketing Law Blog*.

De Gregorio, Giovanni. (2020). The Rise of Digital Constitutionalism in the European Union. *International Journal of Constitutional Law*. Awaiting publication.

Helberger, Natali, Leerssen, Paddy, & Van Drunen, Max. (May 29, 2019). Germany proposes Europe's first diversity rules for social media platforms. [Blog post]. *LSE Blog*.

Kearney, Michael G. (2007). *The Prohibition of Propaganda for War in International Law*. Oxford: Oxford University Press.

Keller, Daphne. (2018a). Who do you sue? *Hoover Institution. Aegis Series Papers*, *1902*.

Keller, Daphne. (2018b). Internet Platforms. Observations on Speech, Danger and Money. *Hoover Institution Aegis Series Paper*, *1807*.

Kettemann, Matthias C., & Tiedeke, Anne Sophie. (2019). Back up: Can Users Sue Platforms to Reinstate Deleted Content? A Comparative Study of US and German Jurisprudence on "Must Carry". *Internet Policy Review, 9(2)*.

Leerssen, Patrick. (2015). Cut Out by The Middle Man. The Free Speech Implications of Social Network Blocking and Banning in the EU. *JIPITEC 1*, 99-119.

Mchangama, Jacob, & Alkiviadou, Natalie. (2020). The Digital Berlin Wall: How Germany (accidentally) created a prototype for global online censorship – Part Two. Copenhagen: Justitia.

Perrigo, Billy. (2020a). These Tech Companies Managed to Eradicate ISIS Content. But They're Also Erasing Crucial Evidence of War Crimes. *Time*.

Perrigo, Billy. (May 4, 2020b). When Content Moderation Hurts. [Blog post]. *Mozilla Foundation*.

Quintais, João Pedro, Frosio, Giancarlo, van Gompel, Stef, Hugenholtz, P. Bernt, Husovec, Martin, Jütte, Bernd Justin, & Senftleben, Martin. (2019). Safeguarding User Freedoms in Implementing Article 17 of the Copyright in the Digital Single Market Directive: Recommendations from European Academics 10 (2020). *JIPITEC,* 277.

Radsch, Courtney. (2020). GIFCT: Possibly the Most Important Acronym You've Never Heard Of. Hint: It Involves the Internet and Terrorism. *Just Security*.

Richter, Andrei. (2016). Defining media freedom in international policy debates. *Global Media and Communication*, *12*(2), 127-142.

Smith, Graham. (June 24, 2020a). Online Harms Revisited. [Blog post]. *Cyberleague.*

Smith, Graham. (February 16, 2020b). Online Harms Deconstructed – the Initial Consultation Response. [Blog post]. *Cyberleague*.

York, Jillian C., & Gullo, Karen. (March 6, 2018). Offline/Online Project Highlights How the Oppression Marginalized Communities Face in the Real World Follows Them Online. [Blog post]. *Electronic Frontier Foundation*.