Plataforma por la Libertad de Información

...porque sin libertad de información no hay democracia

THE DIGITAL SERVICES ACT AND ITS IMPACT ON THE RIGHT TO FREEDOM OF EXPRESSION: SPECIAL FOCUS ON RISK MITIGATION OBLIGATIONS

Joan Barata

Joan Barata is a member of **PLI** since 2017. He is a a Fellow at the Cyber Policy Center of Stanford University. He works on freedom of expression, media regulation and intermediary liability issues. He teaches at various universities in different parts of the world and has published a large number of articles and books on these subjects, both in academic and popular press. His work has taken him in most regions of the world, and he is regularly involved in projects with international organizations such as UNESCO, the Council of Europe, the Organization of American States or the Organization for Security and Cooperation in Europe, where was the principal advisor to the Representative on Media Freedom. Joan Barata also has experience as a regulator, as he held the position of Secretary General of the Audiovisual Council of Catalonia in Spain and was member of the Permanent Secretariat of the Mediterranean Network of Regulatory Authorities.

Executive Summary

The proposed DSA, due to its general scope, could be an important tool in order to guarantee a proper protection of fundamental rights by sector specific legislation, and particularly regarding the impact that the imposition of certain duties on private platforms may have on the right to freedom of expression of their users and third parties.

Article 8 regulates possible orders to service providers from relevant judicial and administrative national authorities to act against a specific item of illegal content. The scope of these orders will be determined by the competent authority. National authorities are granted a very open and almost discretionary power to unilaterally impose a specific interpretation of international freedom of expression standards to third countries.

Article 14 of the proposal regulates notice and action mechanisms. It is necessary to guarantee that when considering notices and before adopting any decision regarding disabling access or removal, hosting providers are entitled and required to make their own good-faith assessment on the basis of the principles of legality, necessity and proportionality.

Duties and responsibilities regarding the assessment and mitigation of systemic risks enshrined in articles 26 and 27 may have an unnecessary and disproportionate impact on the right to freedom of expression. They incorporate a complex regime involving public bodies/State authorities (at the national and the EU level). In such a context, the proper introduction and application of principles and safeguards regarding the protection of human rights as freedom of expression becomes an unavoidable requirement.

ITHE DIGITAL SERVICES ACT AND ITS IMPACT ON THE RIGHT TO FREEDOM OF EXPRESSION: SPECIAL FOCUS ON RISK MITIGATION OBLIGATIONS

1. Introduction. Freedom of expression and platform regulation in the EU: general approach

Freedom of expression as a universal human right: implications

The right to freedom of expression is protected in Europe by article 11 of the EU Charter of Fundamental Rights and article 10 of the European Convention of Human Rights. According to article 6.1 of the Treaty on the European Union, fundamental rights are a source of primary law and therefore all secondary legislation (including the provisions that will be examined in this paper) need to fully respect and comply with them. It is also important to underscore that all EU member States are signatories to the Convention and bound by the case law of the European Court of Human Rights. As stated by article 6.4 of the Treaty, these regional standards, alongside constitutional traditions common to the Member States, constitute general principles of the Union's law.

Human rights law has been traditionally applied to the relations between individuals and States. The latter have the obligation not to establish unnecessary and disproportionate limits to the mentioned fundamental right, and also to ensure enabling environments for freedom of expression and to protect its exercise.

In the course of the recent years, new standards have been formulated aiming at extending the application of some of the mentioned protections to the relations between private individuals and, particularly, those between individuals and corporate businesses. The United Nations Guidelines on Businesses and Human Rights[1] constitute an important international document in this area, although it is neither binding nor even soft law. Some of the recommendations to businesses (particularly, avoid causing or contributing to adverse human rights impacts, make high-level policy commitments to respect the human rights of their users, conduct due diligence vis-à-vis actual and potential human rights impacts, engage in prevention and mitigation strategies, provide appropriate remediation) are frequently mentioned in international documents referring to the relationship between online platforms and their users. In his 2018 thematic report to the Human Rights Council[2], the United Nations Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression David Kaye directly addressed platforms, requesting them to recognize that "the authoritative global standard for ensuring freedom of expression on their

1 These principles were developed by the Special Representative of the Secretary-General on the issue of human rights and transnational corporations and other business enterprises. The Human Rights Council endorsed the Guiding Principles in its resolution 17/4 of 16 June 2011. Available at: https://www.ohchr.org/documents/publications/guidingprinciplesbusinesshr_en.pdf

2 Available at: https://www.ohchr.org/EN/Issues/FreedomOpinion/Pages/ContentRegulation.aspx

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".

The UN Human Rights Council declared in its resolution 32/13 of 1 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."

Public vs private speech rules

Almost every State in the world has in place a set of national rules governing the dissemination of ideas, information, and opinions, online and offline.

Besides this, hosting providers do generally moderate content according to their own – private - rules. Content moderation consists of a series of governance mechanisms that structure participation in a community to facilitate cooperation and prevent abuse. Platforms tend to promote the civility of debates and interactions to facilitate communication among users[3]. Platforms do not only set and enforce private rules regarding the content published by their users. These rules (nowadays very much detailed and developed) will guide oversight activities within their own spaces as well as determine what content is *visible* online and what content – although published – remains hidden or less notorious than other.

In the United States there is so far clear legal consensus that intermediaries have a First Amendment right to moderate content (supported also by the provisions included in Section 230[4]. In Europe, on the other hand, the debate is more open. Platforms have been literally pushed by EU institutions to moderate content[5]. However, national courts have also considered, in some cases, the adoption of positive measures to protect certain forms of online speech when the role and presence of a specific platform within the public sphere has direct implications on the exercise of fundamental rights (notably freedom of expression or right to non-discrimination) or affects basic national and European constitutional

3 James Grimmelmann, "The Virtues of Moderation", 17 Yale J.L. & Tech (2015). Available online at: https://do https://digital-commons.law.yale.edu/yjolt/vol17/iss1/2

4 According to Section 230(c) of of the Communications Decency Act (CDA), which is included in the United States Telecommunications Act of 1996, and particularly the so-called *Good Samaritan* clause, platforms are not liable for the third-party content that they share or decide, in any circumstance, to keep available. Specifically, this means that platforms are not liable for illegal content that they fail to detect or assess. Platforms are free to set their own content policies, which may essentially be tailored to the characteristics of their users and of their commercial, social or even political interests. All this also means that platforms are actually encouraged to ban, police and remove not only presumed illegal posts, but also lawful, yet still harmful or offensive content. Section 230 particularly refers to specific categories of content including "obscene, lewd, lascivious, filthy, excessively violent, harassing" although it also refers to content "otherwise objectionable". Courts have generally seen this objectionability as a catchall notion to cover any type of content that platforms themselves consider objectionable, under their own criteria and internal standards. See Goldman, E., "Why Section 230 Is Better than the First Amendment", 2 *Notre Dame Law Review* Reflection 34 (2019).

5 See for example the Code of Conduct signed by Facebook, Microsoft, Twitter and YouTube with the European Commission in May 2016 with the objective of "countering illegal speech online", the Code of Practice on disinformation signed between the Commission and Facebook, Google and Twitter, Mozilla, as well as by advertisers and parts of the advertising industry in October 2018, with Microsoft and TikTok adhering more recently. It is also important to mention the Communication of the European Commission on tackling illegal content online of 2017 and the Recommendation on measures oriented to the same purpose of 2018.

principles (such as pluralism)[6]. This is still, of course, a very open question, that would deserve proper and consistent future elaboration from the side of national constitutional courts and the European Court of Human Rights.

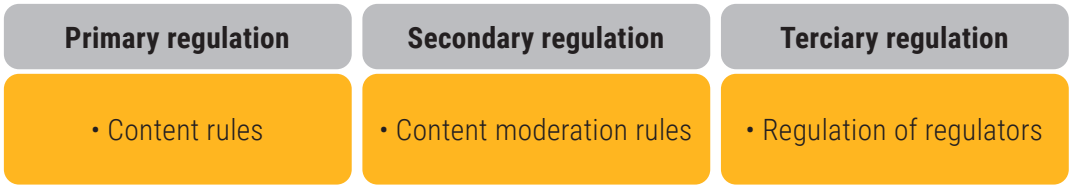
Therefore, State bodies (legislators, regulators, courts…) define the legitimate limits and conditions to the exercise of the right to freedom of expression. They also establish the parameters for the provision of certain services and the legal regime applicable to different activities. All these provisions, regulations and case law establish the criteria to differentiate between legal and illegal online content. When establishing their own private speech rules, platforms may also get inspiration from and replicate traditional freedom of expression principles and cultures, as well as aim at protecting similar values as those covered by State legislation (for example, when it comes to areas such as hate speech, harassment, protection of minors, etc). Platforms have also the power to shape and regulate online speech beyond legal and statutory content provisions in a very powerful way. The unilateral suspension of the personal account of the United States former President Donald Trump's accounts on several major social media platforms has become a very clear sign of this power. Platforms' content policies are often based on a complex mix of different principles: stimulating user engagement, respecting certain public interest values – genuinely embraced by platforms or as the result of policymakers and legislators' pressures –, or adhering to a given notion of the right to freedom of expression, as it has already been mentioned.

In the sphere of online platforms, *primary* regulation is defined as rules and sanctions aimed at platform users, specifying what they may or may not do, with what legal effect, and what the sanctions are. These rules and sanctions can be general (as it is the case of hate speech, which uses to be defined in criminal codes regarding both online and offline communication) or specific (some States, and due to its alleged wide reach, particularly criminalise the dissemination of terrorist content via online platforms, for example). *Secondary* regulation of online speech is integrated by the legal rules meant to induce providers of digital services to influence the activity of their users: the direct target of the regulation are the intermediaries, but the final target are the users. Last but not least, *tertiary* regulation and regulation of online speech incorporates rules that are meant to regulate the activity of regulators, when monitoring or regulating the activities of intermediary services providers, in cases where the latter "regulate" or moderate the activities of users[7].

6 See the very thorough analysis on these matters provided by Kettemann, M.C. and Tiedeke, A.S., 2019, "Back up: Can Users Sue Platforms to Reinstate Deleted Content? A Comparative Study of US and German Jurisprudence on 'Must Carry'", GigaNet 2019. Available at: https://www.giga-net.org/2019symposiumPapers/05_Ketteman_Back-Up-Can-Users-Sue-Platforms.pdf Just to mention a few examples, in Germany, the Higher Regional Court of Munich ruled that a comment posted by a right-wing politician was, according to Facebook, in violation of its internal content rules, nevertheless constituted an exercise of freedom of expression protected under the German constitution. In Italy a Civil Court in Rome decided to reactivate the Facebook account of the far-right party CasaPound and a fine of 800 € for each day the account had been closed. Similar to the German case, the account had allegedly violated internal Facebook's anti-hatred policies. However, the Court considered that the decision of the platform created an unacceptable exclusion or limitation of the voice of CasaPound in the Italian political debate. See a description and analysis of the cases at Columbia University Global Freedom of Expression Database: https://globalfreedomofexpression.columbia.edu/cases/casapound-v-facebook/ and https://globalfreedomofexpression.columbia.edu/cases/heike-themel-v-facebook-ireland-inc/

7 This classification is taken from Sartor, G. and Loreggia, A., The impact of algorithms for online content filtering or moderation, European Parliament's Committee on Citizens' Rights and Constitutional Affairs, 2020. Available at: https://www.europarl.europa.eu/thinktank/en/document.html?reference=IPOL_STU(2020)657101

Figura 1



These three types of legislation and regulation can be seen as modalities of direct or indirect "State action" having implications and impact on the exercise of the right to freedom of expression[8]. The two last categories present the most important level of complexity from a human rights perspective and they are present in several pieces of legislation already adopted by EU institutions, as well as incorporated into the proposal of a Digital Services Act (DSA) as it will be further analysed below.

Platform regulation in the EU and human rights: what we have so far

The decision of the Court of Justice (CJEU) regarding the so-called right to be forgotten – later incorporated into the EU legislation – probably represents the first relevant example of a public intervention on the capacity of platforms to adjudicate on content with particularly strong human rights implications. The important element in this case is that intermediaries (search engines, in particular) become legally obliged to take certain decisions under certain parameters pre-established by a public body (a court in this case). Moreover, such decision does not only have a clear impact on the exercise of important human rights (including the right to freedom of expression and freedom of information) but it also puts in the hands of private actors the responsibility to ponder the different rights at stake. The potential human rights impact of this legal construct was criticised by digital rights organizations[9], moreover the Representative on Freedom of the Media of the Organization of Security and Cooperation in Europe issued a communiqué[10] stating that the decision "might negatively affect access to information and create content and liability regimes that differ among different areas of the world, thus fragmenting the Internet and damaging its universality". The Representative also stressed that "information and personal data related to public figures and matters of public interest should always be accessible by the media and no restrictions or liability should be imposed on websites or intermediaries such as search engines. If excessive burdens and restrictions are imposed on intermediaries and content providers, the risk of soft or self-censorship immediately appears." The so-called Copyright Directive[11] contains a series of obligations vis-

8 Legal challenges related to the identification of State action beyond primary legislation are presented in detail in Keller, D. Who Do You Sue? State and Platform Hybrid Power over Online Speech, Hoover Working Group on National Security, Technology, and Law, Aegis Series Paper No. 1902 (January 29, 2019), available at https://www.lawfareblog.com/who-do-you-sue-state-and-platform-hybrid-power-over-online-speech

9 See for example the communiqué by the Electronic Frontier Foundation "Unintended Consequences, European-Style: How the New EU Data Protection Regulation will be Misused to Censor Speech", published on November 20, 2015, available at: https://www.eff.org/deeplinks/2015/11/unintended-consequences-european-style-how-new-eu-data-protection-regulation-will

10 Available at: https://www.osce.org/fom/118632

11 Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC.

à-vis providers, particularly to ensure the unavailability of certain copyright protected works (article 17). Recital 70 states that such steps "should be without prejudice to the application of exceptions or limitations to copyright, including, in particular, those which guarantee the freedom of expression of users"[12]. The Audiovisual Media Services Directive[13] encompasses a series of duties of so-called video sharing platforms (VSPs) concerning the prevention and moderation of content that constitutes hate speech and child pornography, affects children's physical and mental development, violates obligations in the area of commercial communications, or can be considered as terrorist. Besides this, national authorities (mainly independent media regulatory bodies) are given the responsibility of verifying that VSPs have adopted "appropriate measures" to properly deal with the types of content mentioned above (alongside other undesirable content). Under this scheme, overseen in last instance by public regulatory bodies, platforms do not only bear a duty to take down certain kind of content, but they may also have an obligation to leave legitimate content online[14]. In this context, platforms are broadly requested to consider "the rights and legitimate interests at stake, including those of the video-sharing platform providers and the users having created or uploaded the content as well as the general public interest." The recently adopted Regulation of the European Parliament and of the Council on addressing the dissemination of terrorist content online (TERREG) contains important obligations for hosting service providers in terms of illegal content removal and putting in place specific measures to address the dissemination of terrorist content online. The Regulation incorporates imprecise guidelines establishing that when adopting such measures providers need to take into account "the risks and level of exposure to terrorist content as well as the effects on the rights of third parties and the public interest to information" (recital 22). Designated "competent authorities" (sic) will "determine whether the measures are effective and proportionate".

The following conclusions can be derived from the EU legislation briefly presented:

a) Sector-specific legislation contains relevant provisions aiming at regulating the way platforms moderate speech (including ToS, the use of filtering mechanisms, reporting and flagging tools, handling and resolution complaints mechanisms, etc.). Some of these legal indications are considerably vague and put in the hands of platforms the primary responsibility of defining them (for example, article 5 of the TERREG regarding measures to address the dissemination of terrorist content).
b) Despite the very strong human rights impact the performance of such duties may entail, legislation is extremely vague regarding the criteria, parameters and safeguards that would

12 It is important to note that the Government of Poland has precisely requested to the CJEU the annulment of some of the provisions included in the mentioned article claiming that "the imposition on online content-sharing service providers of the obligation to make best efforts to ensure the unavailability of specific works (…) and the imposition on online content-sharing service providers of the obligation to make best efforts to prevent the future uploads of protected works (…) make it necessary for the service providers – in order to avoid liability – to carry out prior automatic verification (…) and therefore make it necessary to introduce preventive control mechanisms. Such mechanisms undermine the essence of the right to freedom of expression and information and do not comply with the requirement that limitations imposed on that right be proportional and necessary." Republic of Poland v European Parliament and Council of the European Union (Case C-401/19).

13 Directive (EU) 2018/1808 of the European Parliament and of the Council of 14 November 2018 amending Directive 2010/13/EU on the coordination of certain provisions laid down by law, regulation or administrative action in Member States concerning the provision of audiovisual media services (Audiovisual Media Services Directive) in view of changing market realities.

14 See Barata, J., Regulating content moderation in Europe beyond the AVMSD, LSE Blog (25 February 2020), available at: https://blogs.lse.ac.uk/medialse/2020/02/25/regulating-content-moderation-in-europe-beyond-the-avmsd/

need to be considered or incorporated by platforms when adopting and implementing the mentioned measures.

c) The legislation also empowers designated competent national authorities (not necessarily judicial or independent bodies) to verify ex post whether these measures are "appropriate", "effective" or "proportionate". From a general point of view, no specific and detailed mandates – neither procedural, nor substantive - regarding proper consideration and protection of human rights can be found. Public intervention appears to be mainly oriented towards guaranteeing that illegal content is effectively addressed or eliminated.

The proposed DSA, due to its general scope, could be an important tool in order to guarantee a proper protection of fundamental rights by sector specific legislation, and particularly regarding the impact that the imposition of certain duties on private platforms may have on the right to freedom of expression of their users and third parties. In addition to this, the DSA is also an unprecedented opportunity to properly define the procedure and activities of relevant public oversight bodies vis-à-vis both the adequate fulfillment of public interest objectives regarding illegal content, and the mentioned and appropriate safeguard of the right to freedom of expression.

2. The DSA, new duties for platforms and their impact on freedom of expression

Introduction

The DSA constitutes, no doubt, a very relevant and comprehensive proposal. It establishes a series of fundamental rules and principles regarding, essentially, the way intermediaries participate in the distribution of online content. It focuses especially (but not only) on content hosting and sharing platforms, such as Facebook, TikTok, Twitter, or YouTube. The DSA does not repeal the basic provisions established under the E-Commerce Directive[15], and particularly the principle of liability exemption for intermediaries. It also incorporates new important rights for users and obligations for service providers (particularly the so-called very large online platforms: VLOPs) in areas such as terms and conditions, transparency requirements, statements of reasons in cases of content removals, complaint-handling systems, and out-of-court dispute settlements among others.

This paper will focus on the impact that several provisions included in the DSA may have vis-à-vis the right to freedom of expression of users and third parties, on the basis of the conceptual framework defined in the pages above. In particular, attention will be devoted to three types of provisions: action orders from relevant authorities, notice and action mechanisms, and assessment and mitigation of systemic risks.

Article 8: orders to act against illegal content

Article 8 regulates possible orders to service providers from relevant judicial and administrative national authorities to act against a specific item of illegal content, on the basis of the applicable Union or national law, and in conformity with Union law. According to paragraph 2 of this article, such orders may not only cover the territory of several member States, but also have an extraterritorial effect beyond the European Union (and potentially a global one). According to the DSA, the scope of these orders will be determined by the competent authority "on the basis of the applicable rules of Union and national law, including the Charter, and, where relevant, general principles of international law, does not exceed what is strictly necessary to achieve its objective".

Possible extraterritorial effects of orders regarding online content hosted by platforms have been the object of two relevant decisions of the CJEU.

In the case of *Google LLC vs Commission Nationale de l'Informatique et les Libertés (CNIL) and others*[16], in relation to the right to be forgotten, the Court states that "currently, there is no obligation

15 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.
16 Judgment of 24 September 2019, case C-507-17.

under EU law, for a search engine operator who grants a request for de-referencing made by a data subject, as the case may be, following an injunction from a supervisory or judicial authority of a Member State, to carry out such a de-referencing on all the versions of its search engine". In addition to this, it notes that "numerous third States do not recognise the right to de-referencing or have a different approach to that right" and that "the balance between the right to privacy and the protection of personal data, on the one hand, and the freedom of information of internet users, on the other, is likely to vary significantly around the world"[17].

In the case of Eva Glawischnig-Piesczek vs Facebook Ireland Limited[18], the Court affirms that injunctions granted for the purpose of blocking access to or removing stored information previously declared to be illegal due to its defamatory nature, and equivalent content to that which was declared to be illegal, can have a global reach without violating the provisions of the E-Commerce Directive, provided that member States ensure that worldwide-effects measures take into account "rules applicable at international level" (sic). Putting now aside the fact that the decisions of the Court must always be interpreted in light of the specific case, this portion of the Court's decision is extremely short and has created many interpretation doubts and legal caveats. Above all, the main (unanswered) question would be: which international legal standards permit the identification and global banning of content, on the basis of the adjudication made by one single State? In other words, the Court in this case seems to (wrongly) assume that international standards do not only provide general principles and guidance for States when establishing specific legal regimes covering different areas of speech, but they may also grant global validity to decisions taken by national authorities in particular national contexts[19].

All these questions resonate when reading the mentioned paragraph included in article 8. Extra-territorial effects of speech restrictions have very strong implications in terms of international human rights protections. These effects indirectly affect the protection of freedom of expression "regardless of frontiers" afforded under article 19 of the International Covenant on Civil and Political Rights and article 10 of the European Convention on Human Rights. They also preempt the powers, duties and responsibilities of (other) national authorities (also established under international law) regarding the protection and the facilitation of the exercise of the right to freedom of expression in their respective territories. Under the mentioned paragraph national authorities (including administrative bodies) are granted a very open and almost discretionary power to unilaterally impose a specific interpretation of international freedom of expression standards to third countries. In addition to this, the mentioned provisions do not contain any specific safeguards permitting the access to and

17 It is also interesting to note, regarding de-referencing injunctions covering the territory of the European Union, that the Court acknowledges that "the EU legislature has now chosen to lay down the rules concerning data protection by way of a regulation, which is directly applicable in all the Member States". Such rules provide national supervisory authorities with "the instruments and mechanisms necessary to reconcile a data subject's rights to privacy and the protection of personal data with the interest of the whole public throughout the Member States in accessing the information in question and, accordingly, to be able to adopt, where appropriate, a de-referencing decision which covers all searches conducted from the territory of the Union on the basis of that data subject's name". In other words, contrary to what would be the case from a global perspective, an EU-wide injunction would be possible in this case inasmuch as EU legislation has undergone a process of harmonization which permits the adoption of human rights-sensitive decisions of that territorial scope.

18 Judgment of 3 October 2019, case C-18/18.

19 Svantesson, D.J.B., "Bad news for the Internet as Europe's top court opens the door for global content blocking orders", post on Linkedin (3 October 2019), available at: https://www.linkedin.com/pulse/bad-news-internet-europes-top-court-opens-door-global-svantesson/

consideration of the specific circumstances, impact and consequences that the adoption of measures against a certain piece of content may have within the context and legislation of a third country, and particularly vis-à-vis recipients of the information in question.

This would violate the international law principles of comity and reciprocity and open the door to the possibility that other countries with a more restrictive conception of the right to freedom of expression may be able to legitimately extend to the territory of the European Union (and globally) similar remedies based on their national law[20]. It is also important to note that such disparities regarding the interpretation of scope of the right to freedom of expression in light of international standards can even exist between national authorities within the European Union, as the case law of the European Court of Human Rights (unfortunately) shows.

In case the DSA aims at facilitating the interruption of the dissemination of manifestly and seriously illegal content even beyond the EU borders (in cases of child pornography, for example), such circumstance must be better defined in the context of the article in question, including also clear references to fundamental international human rights principles, particularly legality, necessity and proportionality, as well as other international legal standards that would sufficiently justify the adoption of such measures (for example, the United Nations Convention on the Rights of the Child).

Notice and action mechanisms

Article 14 of the proposal regulates notice and action mechanisms. This paper cannot provide a thorough review of the provisions included in this important article. There are however two main areas for concern in terms of implications vis-à-vis the right to freedom of expression.

Although the basis of the notice and action mechanism is the existence of a specific illegal content item, the DSA deliberately refrains from providing a definition of what would be

considered as "illegal" in this context, and in general, the context of the overall Regulation. Paragraph 2 of the mentioned article establishes that notices must contain "an explanation of the reasons why the individual or entity considers the information in question to be illegal content". This categorization would result "from Union law or from national law in accordance with Union law" (according to the explanatory memorandum).

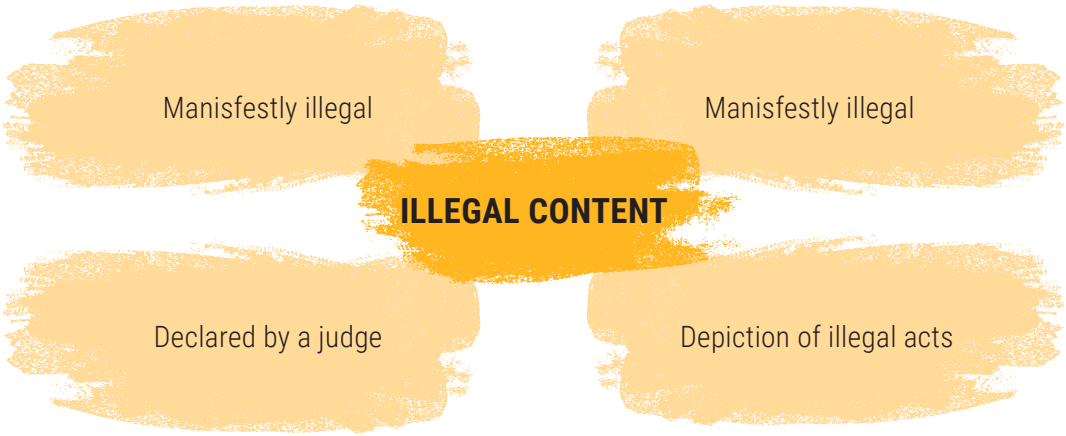
This vagueness and broadness may trigger over-removals of content and affect the right to freedom of expression of users. Illegal content as a broad category may present very diverse typologies, including manifestly illegal and criminally penalised content (child pornography), illegal content as defined by other sources of national legislation (for example, advertising

certain products), content which would only be firmly considered as illegal upon a judicial decision requested from an interested party (defamatory content), or content that depicts or

20 It is interesting to point at the arguments presented by Human Rights Watch, Article 19, Open Net (Korea), Software Freedom Law Centre and Center for Technology and Society in their intervention before the Supreme Court of Canada vis-à-vis the case of Google inc. vs Equustek Solutions Inc. Available at: https://cis-static.law.stanford.edu/cis/downloads/HRW%20Equustek.pdf

represents illegal activities taking place in the *physical* world (which could not be necessarily considered as illegal, as such).

Figure 2



It does not seem necessary and proportionate, in terms of impact on the right to freedom of expression, that all the mentioned categories entail the same consequences in terms of forcing hosting services to expeditiously adopt restrictive measures based on the mere reception of a notice. Therefore, without necessarily assuming the task of defining what content is illegal, the DSA needs to establish the obligation for notifiers to determine not only why a certain piece of content is considered to be illegal, but to properly substantiate the circumstances, context and nature of the alleged violation of the law. In this context, it is important to insist on the fact that, as it has been shown, not all kinds of illegality can be equally acknowledged by a hosting service provider as the result of the sole reception of a communication by a private third party.

Connected to this, paragraph 3 affirms that notices that include, among others, such an explanation "shall be considered to give rise to actual knowledge or awareness". The mere fact that a user argues that a certain piece of content is illegal must not necessarily create knowledge or awareness for the purposes of article 5, unless the notified content reaches a certain threshold of obviousness of illegality (in line with what has been explained in the previous paragraph). It would therefore be important to introduce an additional provision establishing that when considering notices and before adopting any decision regarding disabling access or removal, hosting providers are entitled and required to make their own good-faith assessment on the basis of the principles of legality, necessity and proportionality. In addition to this, it would also be important to spell out that in cases where the mentioned assessment is dismissed by the competent authority, this does not eliminate providers' liability exemptions.

Notifications of suspicions of criminal offences

Last but not least, and in line with what has already been presented in this section, it is also necessary to refer to the provisions included in article 21 of the proposal, obliging platforms to promptly inform the relevant national law enforcement or judicial authorities as soon as

they become aware of "any information giving rise to a suspicion that a serious criminal offence involving a threat to the life or safety of persons has taken place, is taking place or is likely to take place". Once again, this provision puts in the hands of online platforms the responsibility of making really complex and human-rights sensitive adjudications within a context where national legislations (including in EU member States) may sensibly differ, as well as complex jurisdiction problems could also arise. In any case, notions such as "serious criminal offence", "becoming aware" or threats "likely to take place" are too vague, at least, to be seen as compliant with the legality principle that, above all, must guide the drafting of this kind of provisions.

Assessment and mitigation of systemic risks

Very large online platforms (VLOPs) as defined by article 25 of the proposal, will need to assume under the DSA new duties to assess and mitigate "systemic risks".

Article 26 aims at defining such systemic risks by classifying them in three broad categories.

i. Dissemination of illegal content through VLOPs' services

This paper has already presented the problems derived from the introduction of a very broad notion of "illegal content", and in particular, those stemming from the fact that platforms are imposed the legal responsibility, under different articles of the proposal, to make their own determinations in this field.

Article 26 does not use the term "illegal content" to refer to specific pieces of information that would require the adoption of targeted measures by platforms (as in the case of notice and action mechanisms, for example). This provision understands illegal content not only as a broad category, but also as something that needs to be assessed by VLOPs in bulk. However, it does not clarify how this qualification is granted: i.e., whether it refers to content that has already been declared illegal by a relevant authority or at least has already been the object of specific measures under the provisions of the DSA, or it is rather pointing at the foreseeability that still-to-be-produced illegal information could end up being disseminated via the mentioned platforms.

The wording of the provision seems to combine both approaches and to establish that platforms may need to articulate content moderation policies particularly targeting users, accounts, pages, etc. which are proven to have become (or may foreseeably become) sources of illegal content. The most problematic aspect of this provision is the complete lack of concretion regarding the interpretation and enforcement by platforms of a series of key and, once again, freedom of expression-sensitive elements:

a) No specific categories of illegal content are specified and therefore no gradual and granular approach is recommended, on the basis of the different possible types of illegality that platforms are supposed to "assess" and mitigate.

b) There are no indications regarding the introduction of possible – and binding - safeguads aiming at avoiding unnecessary and disproportionate impacts on the exercise of the right to freedom of expression by users and third parties (neither by platforms themselves or, as

it will be shown, oversight bodies). As a matter of fact, the only applicable guidelines are contained in paragraph 2 of this article, and they merely provide a "restrictive approach" by requesting platforms to consider how their content policies favor the "potentially rapid and wide dissemination of illegal content".

c) The provision does not acknowledge the fact that the identification of illegal content is strongly dependent on different areas of not necessarily harmonized national legislation, which therefore creates important discrepancies between Member States. Moreover, horizontal categories like "hate speech", "terrorism" or "extremism" are currently given considerably divergent interpretations across the EU by national law enforcement and judicial authorities, triggering in some cases serious human rights concerns. These differences may have a clear impact not only in terms of assessment but also when it comes to establishing appropriate mitigation measures by platforms that clearly operate beyond borders. Once again, no indications are provided to solve very complex and foreseeable conundrums in this field.

ii. Any negative effects for the exercise of fundamental rights

Alinea b) of paragraph 1 of the mentioned article literally describes as a systemic risk "any negative effects for the exercise of the fundamental rights to respect for private and family life, freedom of expression and information, the prohibition of discrimination and the rights of the child", as enshrined in the Charter of Fundamental Rights.

This systemic risk is presented in a very problematic way, due to the following reasons:

a) Any violation of a fundamental right is, per se, illegal. Therefore, content that causes an illegitimate restriction to a fundamental right would already be contemplated by the general notion of "illegal content" mentioned above.

b) The provision uses the language "any negative effects", which is not appropriate in terms of human rights law. To mention just an example, reporting on matters of public interest may sometimes have a negative effect on the right to public and family life of certain public individuals, although this effect is in most cases overridden by the preeminent protections granted by the European Convention on Human Rights and the case law of the European Court of Human Rights to the right to freedom of expression and freedom of information. In a similar vein, protections granted to children vis-à-vis certain types of content (for example, pornography) are not to be of such nature that deprive adults from safely accessing the same kind of information. On the basis of this provision, national authorities could consider, for example, that heavy criticism against public authorities constitutes in fact a systemic risk to be dealt with by online platforms.

c) The reference to "any violation" to fundamental rights is made on the basis of the consideration of such rights as completely separated realities, and without considering the very frequent need to articulate an interpretation that properly ponders the presence of different "conflicting" rights. In any case, it is unrealistic to understand that platforms may be able, as part of their risk assessment duties, to articulate complex legal interpretations (which are usually performed on a case-by-case basis by national courts and the Court

in Strasbourg through an assessment stretching over several years) with regards to all pieces of content that may trigger such conflicts of rights.

iii. Intentional manipulation of the service

Probably the most problematic provision regarding the description of systemic risks is the one regarding the "intentional manipulation of their service, including by means of inauthentic use or automated exploitation of the service, with an actual or foreseeable negative effect on the protection of public health, minors, civic discourse, or actual or foreseeable effects related to electoral processes and public security". This provision also needs to be particularly connected to the already mentioned 2nd paragraph of article 26, which establishes that "(w)hen conducting risk assessments, very large online platforms shall take into account, in particular, how their content moderation systems, recommender systems and systems for selecting and displaying advertisement influence any of the systemic risks referred to in paragraph 1, including the potentially rapid and wide dissemination of illegal content and of information that is incompatible with their terms and conditions".

These provisions have serious implications vis-à-vis the right to freedom of expression on the basis of the following considerations:

a) The references to negative effects on public health, minors (which needs to be understood as something different from "rights of the child" in the previous Alinea), civic discourse, electoral processes and public security together with the mention to incompatibility, beyond the law, with terms and conditions, clearly show that platforms may face the legal responsibility (overseen by public bodies) to restrict access to lawful content (and therefore protected under the freedom of expression clause) which can be considered as "harmful" under the very vague mentioned criteria. These criteria are subjected to very open interpretations that are dependent on largely different political approaches and sensitivities within the European Union. As a consequence, it is also likely that platforms end up demoting or restricting otherwise legal "borderline content" which could be connected to the mentioned harms.

b) Once more, no specific safeguards are contemplated in order to acknowledge and properly protect the particular role that free expression plays when associated to relevant societal activities such as, precisely, civic discourse, electoral processes or public health. The analyzed provisions exclusively take a "negative" approach towards speech thus legitimizing possible State interventions which would otherwise be forbidden if applied to other content distribution or publication means. If finally adopted, these provisions would give relevant authorities a *backdoor* to introduce (via the mechanisms that will be later described) non-legally based, non-truly transparent and very difficult to account restrictions to the right to freedom of expression on the basis of a series of criteria which, in any case, violate the principles of clarity and foreseeability that are to be required in such scenario. Moreover, the way the establishment of these restrictions is defined would make it impossible for any authority or any court, ex ante or ex post, to assess their necessity and proportionality according to applicable human rights standards.

c) Last but not least, article 26 does not properly define when a risk becomes "too risky" in order to justify the adoption of mitigation measures. In other words, political, economic and social life incorporates per se many disfunctions and risks within the context of modern societies. These problems, including illegal behaviors, exist in parallel with or independently from online platforms. The key element here is to properly assess to what extent intermediaries generate "extra risks" or increase the existing ones up to an "unacceptable" level. The next big question is whether platforms can be put in the position of making such a complex analysis and deciding the best tools to deal with those negative effects. It is important to insist on the very strong human rights implications that these tasks entail. In addition to this, authorities designated by the proposal to oversee platforms' decisions in this area may have the capacity to assess the procedures and practices incorporated by platforms in the fulfillment of these "duties of care". However, can these authorities be entrusted, or better yet, do they have the legitimacy to make comprehensive judgements regarding the desirable openness and plurality of the public discourse, the fairness of the electoral process or the protection of public security? Aren't these matters at the core of our democracies and therefore, don't they require the most open and plural civic debates and institutional procedures?

All this being said, it is now necessary to refer to the different ways the mentioned risks may be mitigated, according to article 27. They include the possible adoption of a wide range of internal content moderation practices (paragraph 1), to be complemented with criteria provided by the Board and the Commission (paragraph 2) and guidelines provided by the Commission in cooperation with national regulators. Although these criteria from public bodies may seem, at first glance, to be "optional", they quickly become mandatory, as failure to follow them can lead to penalties.

According also to article 27, measures adopted to mitigate the systemic risks need to be "reasonable, proportionate and effective". However, considering the complexity of the tasks assigned to online platforms in the first instance, these general principles do not provide much clarity or foreseeability regarding the measures and practices to be implemented.

Recital 68 establishes that "risk mitigation measures (…) should be explored via self- and co-regulatory agreements" (contemplated in article 35) and in particular that "the refusal without proper explanations by an online platform of the Commission's invitation to participate in the drawing up and application of such a code of conduct could be taken into account, where relevant, when determining whether the online platform has infringed the obligations laid down by this Regulation". Such determination is particularly implemented via enhanced supervision mechanisms in the terms of article 50. Within this context, it is necessary to make the following remarks:

a) Regarding risk mitigation measures, it needs to be noted that in many cases the only possible way to deal with systemic risks and/or respect the rules established via the mentioned codes may require the use of automated filtering mechanisms. Without prejudice to the transparency obligations included in the DSA regarding the use of such mechanisms, it is important to note here that errors by automated monitoring tools can seriously and irreversibly harm users' fundamental rights to privacy, free expression and information, freedom from discrimination, and fair process. However, the DSA does not

contain any clear and binding directive to guide the design and implementation of this type of measures, particularly when it comes to human rights implications.

b) It is important to note the absence of any relevant provision establishing the need that platforms, co-regulatory mechanisms and oversight bodies properly consider the impact on human rights, and particularly freedom of expression, that the implementation of the mentioned mitigation measures may entail.

c) In the European model, the establishment of restrictions to the right to freedom of expression by non-legislative bodies is connected to the presence of an independent body not subjected to direct political scrutiny or guidance. The very important role that a non-independent body like the European Commission may play vis-à-vis the articulation and implementation of measures with a clear impact on speech is in contradiction with this model.

d) From a more general point of view, there are no specific provisions requiring that platforms' internal and independent processes and audits incorporate a clear, international law-based and thorough human rights impact perspective, particularly in the area now under consideration.

e) Last but not least, the activities and measures undertaken and adopted within the framework of articles 26 and 27 cannot in any case be seen as mere private content policies under the exclusive responsibility of online platforms. They are rather the result of a complex intervention involving public bodies/State authorities (at the national and the EU level). Such intervention takes place ex ante, via the rules included in the DSA, and ex post, due to the capacity of different public bodies to shape and constrain the different ways platforms deal with systemic risks, which entail the dissemination of and access to far more types of content than merely illegal information. Therefore, in such a context, the proper introduction and application of principles and safeguards regarding the protection of human rights as freedom of expression becomes an unavoidable requirement.

3. Conclusions

The DSA constitutes a very relevant and comprehensive proposal. It establishes a series of fundamental rules and principles regarding, essentially, the way intermediaries participate in the distribution of online content. It also incorporates new important rights for users and obligations for service providers (particularly VLOPs) in areas such as terms and conditions, transparency requirements, statements of reasons in cases of content removals, complaint-handling systems, and out-of-court dispute settlements, among others.

This being said, duties and responsibilities regarding the assessment and mitigation of systemic risks may have an unnecessary and disproportionate impact on the right to freedom of expression of users, according to what has been presented in this paper.

Regarding possible ways to properly address these matters, particularly when it comes to the wording of articles 26 and 27, it is suggested to take into account the following elements:

a) Systemic risks assessment/mitigation scheme can be considered as a very particular tool which entails a prior restraint mechanism regarding speech. According to the case law of the European Court of Human Rights, these mechanisms need to be subjected to the strictest scrutiny. Under the described regime, regulatory and executive bodies will intervene in setting up the variable criteria and parameters according to which content will be moderated and "regulated" by platforms. International and regional standards establishing that content can only be restricted/limited by State authorities on the basis of clear and or foreseeable legal provisions are difficult to articulate and apply to this context as part of the discussion.

b) Advocating for the complete elimination of articles 26 and 27 of the proposal would collide with general political and societal concerns regarding the role and risks posed by online platforms vis-à-vis, at least, the dissemination of illegal content.

c) Possible amendments to articles 26 and 27 need to promote a focus on clearly defined illegal content, rather than broader categories of so-called harmful content.

d) Rather than imposing due diligence obligations to deal with legal-but-harmful content (content regulation via delegation), the DSA would need to reinforce liability exemptions that incentivise platforms' own initiatives regarding the moderation of content, particularly in connection with properly and adequately formulated terms of service[21].

21 See the proposals contained in Barata, J., "Positive Intent Protections: Incorporating a Good Samaritan principle in the EU Digital Services Act", Center for Democracy and Technology. Available at: https://cdt.org/wp-content/uploads/2020/07/2020-07-29-Positive-Intent-Protections-Good-Samaritan-principle-EU-Digital-Services-Act-FINAL.pdf See also Barata, J.,"The Digital Services Act and the Reproduction of Old Confusions: Obligations, Liabilities and Safeguards in Content Moderation", VerfBlog, 2021/3/02. Available at: https://verfassungsblog.de/dsa-confusions/

e) Any possible due diligence obligation aimed at preventing the dissemination of illegal content need to be commercially reasonable, transparent, proportionate, more principled than prescriptive, and flexible. Such obligations should not focus on the outcomes of content moderation processes, i.e. intermediaries should not be evaluated on whether they have removed "enough" illegal content, as this creates a strong incentive towards over-removal of lawful speech. In order to facilitate the effectiveness of such measures, intermediaries might be subject to *ex ante* regulatory oversight, receive support and assistance from regulators, civil society and other major stakeholders, and engage in the adoption of codes of conduct.

f) Provisions related to the matters mentioned in the previous paragraph would need to establish the obligation to undertake solid and comprehensive human rights impact assessments. As a matter of principle, any private or regulatory decision adopted in this field would need to guarantee that human rights restrictions/conditions have been properly considered on the basis of the principles of necessity and proportionality.

g) In addition to this, the DSA would also need to establish proper mechanisms for ex post assessment of due diligence measures implementation, as well as proper appeal mechanisms for all interested parties regarding all relevant decisions adopted in this field by competent bodies.