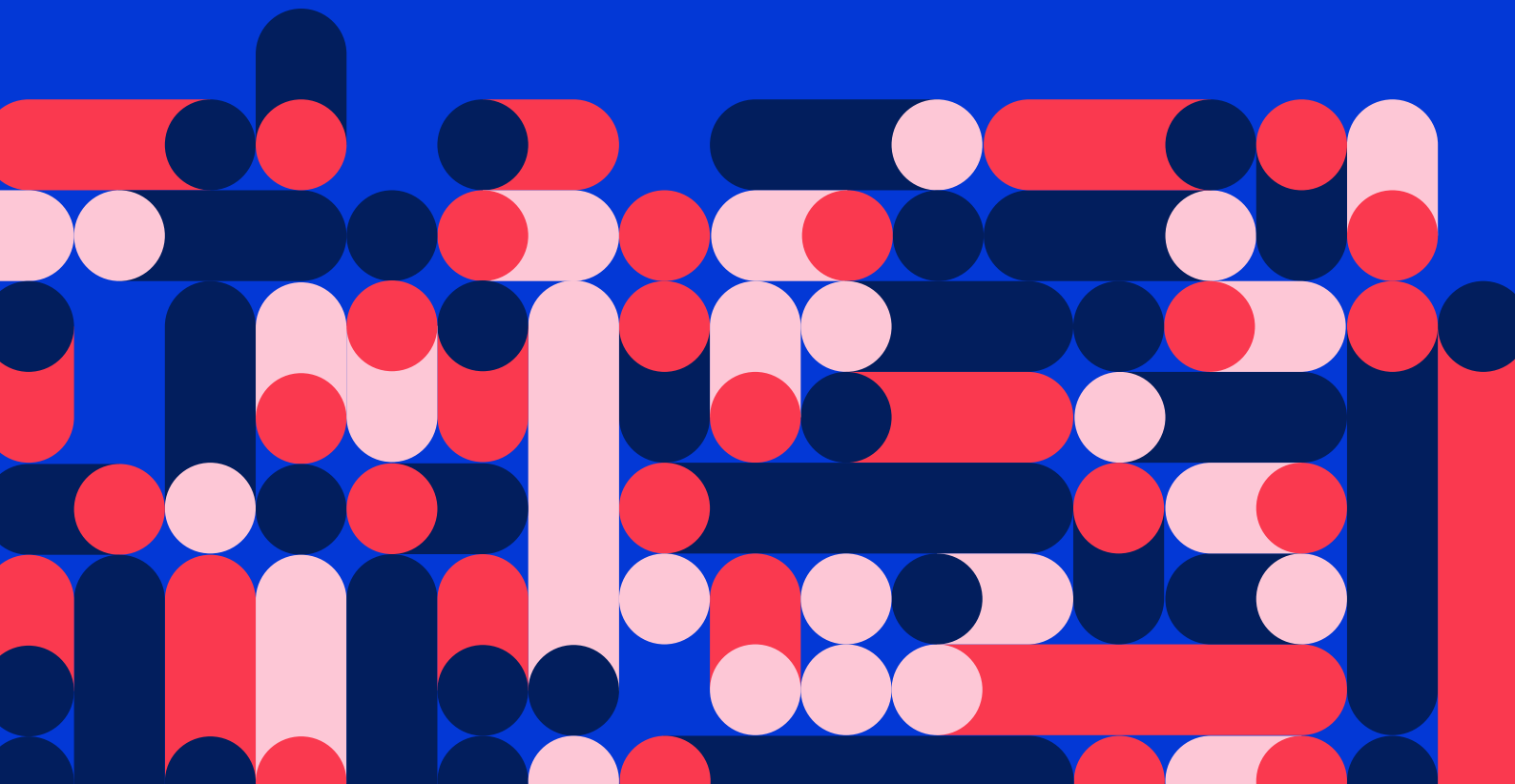
Mirando Al Sur

Towards new regional consensus on intermediary liability and content moderation on the Internet

AlSur



Mirando Al Sur. Towards new regional consensus on intermediary liability and content moderation on the Internet
April, 2021.

This document was made by **Al Sur** with the support of the INDELA Fund.
First version finalized in December 2020.

Authors: **Agustina Del Campo** and **Morena Schatzky** from the CELE of the University of Palermo; and **Laura Hernández** and **J. Carlos Lara** from Derechos Digitales.

AlSur is a consortium of organizations working in civil society and academia in Latin America that seek to strengthen human rights in the region's digital environment by working together. For more information about Al Sur and its members, visit https://www.alsur.lat/en.

Table of contents

I. Introduction

Unlike the United States or Europe, where there are laws that have adopted concrete positions to deal with intermediary liability, in Latin America the issue arrived on the public agenda relatively late and indirectly through free trade agreements (FTAs). When legislative proposals that adopted and reflected the consensus that seemed to exist at a global level began to emerge in the region, they suffered some heavy blows to their foundations. Starting in 2013, with the Snowden revelations and, more particularly, since 2016 with the election of Donald Trump as president of the United States, the Brexit, the Pizzagate, among others, harsh questioning began to emerge around the immunity of intermediaries regarding third-party content. This global questioning landed with force in the Latin American region in 2018, as a result of the presidential elections in Brazil.

Faced with these changes in global consensus and taking into account the lack of consensus in Latin America, civil society organizations have been debating the issue in different instances. Great efforts have been made to maintain a view from the regional human rights framework that can be articulated with the ongoing global debate, while being respectful of the interaction of public and private interests linked to the issue. Along these lines, Al Sur, having a common perspective on privacy and freedom of expression, promotes debates aimed at informing our positions in light of the rapid and multiple technological, regulatory and industrial developments. This document is part of a process that formally began in March 2019, with a meeting of experts organized by the CELE in Buenos Aires, where the current developments on the subject were collectively analyzed at the regional and comparative levels, agreements and disagreements in the approach to this issue were identified and different options and advocacy initiatives were evaluated[1].

The Buenos Aires meeting was held in the aftermath of the failure of what was perhaps one of the most paradigmatic legislative initiatives since Brazil's *Civil Rights Framework for the Internet*: Argentina's intermediary liability project. This project led to a parliamentary discussion that went on for more than five years, with a text being agreed between the ruling party and the opposition in the Senate, driven by the decision of the Supreme Court of Justice in the *Case Belén Rodríguez v. Google* (2014)[2]. Although the project was not perfect and experts warned about some of its provisions (particularly those referring to self-regulation), its termination generated strong frustration in both the private sector and civil society, who had invested years in the discussion. On top of this, issues that would later prove to be thorny began entering the discussion, such as the dispute between media and platforms over advertising, a claim that in Europe had led to the adoption of the *European Directive on Copyright* of 2019 (granting journalists the right to a fee for the content disseminated and shared online), or the worrying trend towards authoritarian regulations in the region regarding the liability of internet intermediaries, such as those that have emerged in countries such as Venezuela[3] or Honduras[4]. At that time,

---

1 Center for Studies on Freedom of Expression and Access to Information (CELE), *Experts' Meeting on Intermediary Liability*, Buenos Aires, Argentina, March 2019.

2 Supreme Court of Argentina, *Ruling in the case 'Rodríguez, María Belén v. Google Inc. without damages'*, R.522.XLIX, 28 Oct 2014. Available at: https://sjconsulta.csjn.gov.ar/sjconsulta/documentos/verDocumentoByIdLinksJSP.html?idDocumento=7162581 [in Spanish]

3 See, for example, RELE's Press Release R179/17 on Venezuela's "Law Against Hate". Available at: http://www.oas.org/en/iachr/expression/showarticle.asp?artID=1082&lID=1

4 See more information at: http://www.ipsnews.net/2018/08/laws-threats-curtail-freedom-expression-honduras/

in the face of regulatory proposals that lacked a clear alignment to a framework of respect for human rights, the context urged for the abandonment of an active promotion of intermediary regulation, to pursue instead a more reactive strategy.

Two years after the Buenos Aires meeting, the region is clearly going through a different moment. On the one hand, several countries are resuming strategic discussions around intermediary liability at the judicial or legislative level. In Brazil, the constitutionality of some provisions of the *Civil Rights Framework* is being challenged before the Supreme Federal Court[5]; in addition, various initiatives have emerged seeking to regulate online misinformation, such as the bill against fake news[6]. In Mexico, the issue gained relevance as a result of the recent adoption and implementation of an FTA with the United States that 'imports' their *Communications Decency Act* (CDA) and *Digital Millennium Copyright Act* (DMCA), later advanced by the proposal of Senator Monreal in 2021, which introduces an amendment to the Federal Telecommunications and Broadcasting Law aimed at defining the scope of the exercise of freedom of expression on social networks. The proposal is presented as an attempt to defend freedom of expression, but includes proposals that risk becoming mechanisms of censorship and control. Finally, in Argentina the discussion has been resumed, although not yet formally in the legislative sphere.

In February 2021 the Interamerican Commission on Human Rights (IACHR) and its Office of the Special Rapporteur for Freedom of Expression (RELE) deplored the "general deterioration of public debate"[7], highlighting the correlation between online violence and physical violence, with the possible lethal consequences of the latter. Furthermore, they stressed the challenge for its regulation and the related political processes. Through its press release, the IACHR opened an Interamerican dialogue process on these issues, under the leadership of the RELE, as an invitation to work together from a multisectoral perspective, with the aim that the results of this dialogue will serve as non-binding inputs for the IACHR to deliberate on the future of the Interamerican Standards on Freedom of Expression on the Internet. The first activity in this process was the thematic hearing *Internet content moderation and freedom of expression in the Americas* held by the IACHR on March 25, 2021, in their 179th Session[8].

At the same time, both in the United States and in Europe, potential reforms are being debated to the intermediary liability frameworks that have hitherto marked the development of the Internet as we know it. These debates were recently joined by Australia and Canada[9]. For their part, the intermediaries themselves have developed self-regulation mechanisms and proposals, aimed at dealing with the prevailing crisis while strengthening their legitimacy and avoiding less

---

5 Supreme Federal Court (Brazil), *Rapporteur believes that messaging apps cannot be forced to provide encrypted data*, STF News, 27 May 2020. Available at: http://www.stf.jus.br/portal/cms/verNoticiaDetalhe.asp?idConteudo=444265 [in Portuguese]. It is worth noting that this is still the only comprehensive regulation on intermediary liability in the region.

6 See more information at: https://www.bbc.com/portuguese/brasil-53244947 [in Portuguese]

7 See more information at: http://www.oas.org/en/iachr/jsForm/?File=/en/iachr/media_center/preleases/2021/026.asp

8 Available at: https://youtu.be/esZnTtKkx8A [in Spanish]

9 Due to time constraints, the Australian and Canadian initiatives are not analyzed in this document. It should be noted that they refer primarily to the relationship between media and platforms, and developing the subject would merit a new study specifically dedicated to it.

permissive regulations[10]. Additionally, different sectors of civil society have developed proposals for regulation, self-regulation and coregulation, which are currently under discussion[11].

The first section of this document raises the importance of the discussion for the region and how this discussion is inserted in comparative debates. The second section describes and maps the legal discussion around intermediary liability at the Interamerican regional level, addressing regulatory and self-regulatory initiatives, as well as proposals and principles promoted by civil society. The third section does the same at an international and comparative level, looking particularly at Europe and the United States. The document is limited to initiatives presented until the end of December 2020, however, being aware of the global impact that the events of January 6, 2021 (and the subsequent "deplatforming" of then President Trump) had on the discussion over the liability of Internet intermediaries, the document includes a brief reference to it.

Finally, arriving at the conclusions, we detail the coalition recommendations to push forward the dialogue in our region, both within the framework of the regional process convened by the IACHR, and in discussions at the local level. These are starting points that attempt to identify the minimum points of agreement and the decisive issues in this debate, based on the accumulated experience of a group of civil society organizations specializing in the promotion of fundamental rights in digital environments which, being knowledgeable about global developments, embrace a regional perspective anchored in the standards of the Interamerican Human Rights system.

10 See, for example, Facebook's Oversight Board: https://oversightboard.com/

11 E.g. the Social Media Councils created by ARTICLE 19: https://www.article19.org/social-media-councils/

II. Importance of the discussion

The importance of the Internet as a medium and platform for the exercise of human rights is indisputable. Since as early as 2017, some actors refer to social media, even officially, as the new *public forum* and, although the analogies with a physical public square may guide the debate to a certain extent, their limitations have become evident[12]. The growing adoption of private policies to regulate the debate in these spaces has created new tensions, in addition to the fact that the proliferation of services linked to certain platforms —especially the largest ones— has generated intense debates about which ones formally constitute intermediation services and which ones have acquired a different nature. The discussion that emerged in Argentina as a result of the lawsuit introduced by former President Cristina Kirchner against Google brought forth this tension for the first time in Latin America[13].

In a region characterized by the lack of specific regulation on intermediary liability related to third-party content across its whole spectrum, the ever increasing complexity of the debate at the global level threatens the construction of minimal consensus. Perhaps one of the most fundamental principles that both civil society and governments seem to agree on today is that any regional position regarding the liability of intermediaries must, necessarily, take into account the regional principles and standards on freedom of expression[14]. But even this consensus is currently subject to multiple and diverse interpretations and there is no shortage of attacks in light of the growing concern regarding the proliferation of hate speech, fake news, scams, slander and insults.

---

12 The *Public Forum Doctrine* was born in the United States as a result of the precedent *Hague v. CIO* (1939), where the Supreme Court ruled that banning a group of citizens from holding political meetings in a public place violated the group's freedom to assemble under the First Amendment. The case overturns the precedent that assimilated the power of public officials over public space to that of private landlords over their private property, and imposes on the Government the obligation to facilitate and protect freedom of expression in public spaces (*Hague v. Committee for Industrial Organization*, 307 U.S. 496 at 515-16). With the passage of time and the emergence of digital spaces, the Public Forum Doctrine was updated through different rulings. In *Packingham v. North Carolina* (2017), the Supreme Court refers to the application of this doctrine on the Internet, extending its application to non-traditional forums, such as social media networks (*Packingham v. North Carolina*, *582 U.S.* (2017), 137 S. Ct at 1735-36): "While in the past there may have been difficulty in identifying the most important places (in a spatial sense) for the exchange of views, today the answer is clear. It is cyberspace—the "vast democratic forums of the Internet" in general, Reno v. American Civil Liberties Union, 521 U. S. 844, 868 (1997), and social media in particular"; "(...) Social media offers "relatively unlimited, low-cost capacity for communication of all kinds." Reno, supra, at 870. On Facebook, for example, users can debate religion and politics with their friends and neighbors or share vacation photos. On LinkedIn, users can look for work, advertise for employees, or review tips on entrepreneurship. And on Twitter, users can petition their elected representatives and otherwise engage with them in a direct manner."

13 See Del Campo, Agustina, *CFK vs. GOOGLE: Theoretically interesting, factually mundane*, Legislative Observatory on Freedom of Expression, CELE, 14 Aug 2020. Available at: https://observatoriolegislativocele.com/en/cfk-c-google-theoretically-interesting-factually-mundane/

14 "Given the importance of private actors as intermediaries for access to and use of the Internet, authorities need to give them the safeguards to operate transparently before the rest of the system's actors (especially the end users), and they should create the conditions to be able to effectively serve as a vehicle for the exercise of the universal right to freedom of expression." IACHR, Annual Report 2013 of the Office of the Special Rapporteur for Freedom of Expression, Chapter IV, Freedom of Expression and the Internet, § 110, p. 510; "There is an international consensus and a commitment to the need to promote universal access to the Internet as an essential means for the effective exercise of human rights online, particularly freedom of expression; and the multi-stakeholder governance of the Internet as a guarantee for the development of technologies respectful of human rights." IACHR, Annual Report 2016 of the Office of the Special Rapporteur for Freedom of Expression, Chapter III, Standards for a Free, Open, and Inclusive Internet, § 18, p. 402.

Faced with this scenario, the debate and the construction of consensus among civil society and academia have special relevance. First, to preserve the integrity of a global network such as the Internet and to promote from the region the defense of the decentralized, inter-jurisdictional, neutral, non-discriminatory and accessible network that the Special Rapporteurs for Freedom of Expression of the OAS, UN and others have been referring to since 2001[15]. Second, to promote a regulation that respects the architecture of the Internet in its different layers, guaranteeing users the basic human rights to privacy and to freedom of expression, both at the infrastructure level and at the logical and content layers.

Any serious effort to address the thorny issue of intermediary liability at this historical moment must address, in addition to the local context, the regional and global context. Understanding what exactly is being discussed and which alternatives are being debated locally, regionally and globally, evaluating the best and worst practices that have emerged so far and distilling from there general consensus is essential.

III. The current situation and the debate in the Americas

Within the regional framework, there is no consensus on what the appropriate regulations should be regarding the liability of Internet intermediaries. In 2013 and 2016, the OAS Office of the Special Rapporteur for Freedom of Expression established regional guidelines for the regulation of intermediary liability, with the aim of protecting freedom of expression[16]. However, these guidelines have not been transferred to the internal regulatory environment of the different countries, with the exception of Brazil. In general terms, the little regulation that exists is fragmented (mainly incorporated through FTAs) and the bills to address the issue from a set of general principles did not emerge until after the jurisprudence of various courts addressed the matter[17].

The lack of a specific legal framework regarding intermediary liability means that the framework applied in our region continues to be that of the general principles of civil liability. Although most countries have ruled out the application of strict liability regimes, discussions on this issue persist and multiple and varied resolutions are found in different judicial instances for different cases. The complexity of the discussion at the international level has also undermined the few consensuses reached at the regional level on the subject.

15 Special Rapporteurs for Freedom of Expression of the IACHR, OAS, UN, OSCE, *Joint Declarations*, 2001 onwards. Available at: http://www.oas.org/en/iachr/expression/basic_documents/declarations.asp

16 IACHR, *Annual Reports of the Special Rapporteur for Freedom of Expression*, 2013 and 2017. Available at: http://www.oas.org/en/iachr/expression/reports/annual.asp; and Freedom of Expression and Internet, 2013. Available at: http://www.oas.org/en/iachr/expression/topics/internet.asp

17 For example, see *Rodríguez, María Belén v. Google Inc.* in Argentina or Gloria v. Google y El Tiempo in Colombia, among others.

a. Free Trade Agreements (FTA)

In Latin American countries, most of them characterized by a lack of specific regulation on intermediaries, the implementation of FTAs with the United States has been one of the main factors triggering the discussion on the liability of Internet intermediaries.

The FTAs include chapters dedicated to intellectual property rights, which incorporate provisions that limit the liability of Internet intermediaries for copyright infringement by their users. These regulations are a direct export of the "safe harbor" model established in the US *Digital Millennium Copyright Act* (DMCA), which shields intermediaries from liability in cases of copyright infringement, provided that they comply with a series of conditions specified in the act itself.

The DMCA includes the controversial "notice-and-takedown" provision, by means of which intermediaries may benefit from the exclusion of liability: Internet Service Providers (ISPs) must promptly remove the alleged infringing material upon notification; there must be a designated representative to receive notifications; they must adopt a policy for repeat offenders and to accommodate standard technical measures and; there shall be no obligation to actively monitor copyright infringement, among others.

However, the provisions included in these treaties have not been fully or uniformly adopted by all the countries with which the United States signed such agreements. The only countries that have fully incorporated the copyright-related intermediary liability provisions contained in the DMCA are Chile and Costa Rica[18]. More recently, Mexico has been added to this list, despite the fact that the adopted regulations are being litigated in several courts[19].

In the Chilean case, the "safe harbor" model was implemented by Law No. 20,435 of May 2010, which mostly follows the provisions set forth in the Chile-United States FTA. The law establishes that the limitations to the liability apply only when the ISP does not initiate the transmission, nor does it select the content or its recipients. Furthermore, it explicitly excludes the obligation of ISPs to actively monitor content. Hosting and caching providers must establish a policy to terminate service for repeat offenders; they shall not interfere with technological protection measures and; they shall not participate in the creation of content or the selection of its recipients. Another requirement is that ISPs must provide contact information to receive the notifications of copyright infringement.

Nonetheless, the Chilean legislation moves away from the DMCA regulation in the way in which the exclusion of joint liability is activated: while the DMCA establishes in its safe harbor model the removal of alleged infringing content through mere private notification, the Chilean law establishes that the joint liability operates only when receiving a court order enforcing it.

---

18 Lara, J. C. & Sears, A. M., *The Impact of Free Trade Agreements on Internet Intermediary Liability in Latin America*, in Frosio, G. (editor), *The Oxford Handbook of Online Intermediary Liablity*, Oxford University Press, 2020.

19 See González, Matías, *The Mexican T-MEC 2020: Practices and lessons*, Legislative Observatory on Freedom of Expression, 29 Oct 2020. Available at: https://observatoriolegislativocele.com/en/el-t-mec-mexicano-2020-practicas-y-lecciones/

The mechanism adopted by Chile led to the United States denouncing that the South American country was not complying with the obligations derived from the FTA, indicating that it was not doing enough to protect the interests of copyright holders. This caused Chile to be included, since 2011, in the "Priority Watch List" of the USTR's annual *Special 301 Report*, calling on the Chilean government to amend the law in order to guarantee effective and expeditious action against Internet piracy.

In Costa Rica, the standards on intellectual property and intermediary liability arrived through the CAFTA-United States agreement, of which Costa Rica is a signatory. The implementation of the provisions on the liability of Internet intermediaries was carried out through the approval of Executive Decree No. 36,880 of December 2011, which establishes the following requirements for Online Service Providers (OSPs): 1. Having a contract termination policy for repeat offenders; 2. Not interfering with technological protection measures, or with content, its origin or recipients; 3. Designating a recipient for notifications and; 4. Expeditiously removing content according to the decree's guidelines.

In the Costa Rican case, the way in which the exclusion of liability is activated is mixed, through judicial enforcement and through a collaborative procedure. In the first scheme, the court decides, in accordance with the copyright general rules, to impose appropriate measures such as removing an account, deleting or blocking the content identified as infringing, or other measures deemed necessary. In the second scheme, a series of notifications addressed to the user who shared the content considered infringing allow them to remove the content in good faith. If they do not agree to do so voluntarily, the OSP may cancel the offender's account or remove/block the content. Costa Rica has also been included in the "Priority Watch List" of the annual Special 301 Report, under the consideration that the notification system terms do not collaborate in the fight against online piracy.

On the other hand, the attempts to nationalize the obligations derived from the FTAs in Colombia and Peru have failed, largely due to the resistance from civil society to copying the DMCA notice-and-takedown provision instead of presenting a model with more guarantees. Consequently, these countries have been incorporated into the "Priority Watch List" and the "Watch List", respectively, of the annual Special 301 Report. Meanwhile, Panama and the rest of the CAFTA bloc countries have not adapted the obligations emerging from the agreements to their internal regulations.

The most recent case of the incorporation of intermediary liability rules through FTAs is that of Mexico with the USMCA, which has not yet been fully implemented[20]. In June 2020, the General Congress of the United Mexican States approved a series of laws on the subject, including amendments to the copyright law and the regulation dealing with intermediary liability, which are a faithful reflection of CDA's Section 230 and the DMCA. Mexican civil society organizations have launched an aggressive campaign against these measures and have managed to initiate a debate around the constitutionality of the treaty in relation to this point[21].

---

20 TN: *The Agreement between the United States of America, the United Mexican States, and Canada* (commonly referred to as *New NAFTA*) is known in the United States as USMCA, in Canada as *CUSMA* (in English) or *ACEUM* (in French) and in Mexico as *T-MEC*.

21 R3D, *The implementation of the T-MEC must include sufficient exceptions for eluding digital locks*, 23 Jun 2020. Available at: https://r3d.mx/2020/06/23/la-implementacion-del-t-mec-debe-contemplar-excepciones-suficientes-para-la-elusion-de-candados-digitales/ [in Spanish]

Undoubtedly, the FTAs brought the issue of intermediary liability to the regional agenda prior to a broad or comprehensive debate on the issue. The clauses related to intellectual property in particular had a regional penetration even beyond the texts or obligations contained in the agreements. Uruguay, for example, has recently incorporated into its legislation, through amends to its Budget Law, an administrative notice-and-takedown mechanism for copyright infringements arising from unauthorized broadcasting of pay television signals[22]. The scope of the original project was broad and ambiguous, with the complainant being able to activate the takedown mechanism by just submitting an administrative complaint, without the need for a judicial order to ensure due process guarantees. To this purpose, the *Communication Services Regulatory Unit* (URSEC) was empowered to adopt the respective sanctioning and preventive measures and order the blocking of sites that broadcast or link to the broadcast of such services. The original proposal was widely resisted by civil society, and although the article 712 that was finally approved restricts the initially proposed scope of the measure, it still maintains the nature of an administrative blockade[23].

Differences persist in the divers regimes at the regional level, with strong clashes of interests around the issue. The search for consensus has been truncated for the moment[24]. This discussion recently acquired particular relevance with the demands of the media sector for its contents to be incorporated into the intellectual property regime (accompanying a global trend marked by the adoption of the European Copyright Directive and the most recent experiences with the Australian law and the Canadian proposal[25]). These new claims and regulatory initiatives have raised the profile of the copyright discussion again, both regionally and globally.

---

22 General Assembly of Uruguay, *National Budget Law 2020-2024*. Available at: https://parlamento.gub.uy/documentosyleyes/ficha-asunto/147701 [in Spanish]

23 OBSERVACOM, *International concern for Uruguayan Law that authorizes blocking and deleting audiovisual content without a court order*, 3 Nov 2020. Available at: https://www.observacom.org/preocupacion-internacional-por-articulo-de-ley-de-presupuesto-de-uruguay-que-autoriza-notificacion-y-bajada-por-derechos-de-autor-sin-orden-judicial/ [in Spanish]

24 The pressure through inclusion in the Special 301 Report has been a constant for decades, with the FTA implementation duties being a relevant factor to maintain that pressure. This is the case of the pressure exerted to promote the protection of software in Brazil (cfr. Valente, M. G., *A construção do direito autoral no Brasil: cultura e indústria em debate legislativo*. Belo Horizonte: Editora Letramento, 2019). In contrast, the Chilean government has rejected this pressure, denouncing that it is a mechanism that does not belong in the FTA, and that it ignores the advances regarding intellectual property. See in this regard the statements issued in response by the Ministry of Foreign Affairs of Chile: *Chile: Solid institutionality regarding intellectual property*, 20 Jul 2008. Available at: https://www.subrei.gob.cl/sala-de-prensa/noticias/detalle-noticias/2008/07/21/chile-institucionalidad-solida-en-materia-de-propiedad-intelectual; and *Public Statement - Special 301 Report*, 28 Apr 2020. Available at: https://www.subrei.gob.cl/sala-de-prensa/noticias/detalle-noticias/2020/04/29/declaracion-publica-reporte-especial-301 [both in Spanish]

25 Parliament of Australia, Treasury *Laws Amendment (News Media and Digital Platforms Mandatory Bargaining Code) Act 2021*, No. 21, 2021. Available at: https://www.legislation.gov.au/Details/C2021A00021; In Canada, a project was presented in February by opposition Sen. Claude Carignan: Senate of Canada, *Bill S-225 - An Act to amend the Copyright Act (remuneration for journalistic works).* Available at: https://www.parl.ca/LegisInfo/BillDetails.aspx?Language=E&billId=11108562

b. Civil Rights Framework for the Internet (MCI)

In Brazil, prior to the enactment of the *Civil Rights Framework for the Internet* (MCI), resolutions used to hold the strict liability of Internet providers regarding third-party content infringements, based on the risk created by the services[26]. The approval of the MCI provided clarity by delimiting the liability of Internet intermediaries, introducing as a general rule a system of exclusion of liability with a model of judicial notice-and-takedown mechanism concerning third-party illegal content. This rule, contained in its article 19, admits two exceptions: for copyright matters, the provisions of the corresponding law are followed, with a notice-and-takedown model in line with the copyright rules of the DMCA; regarding the non-consensual dissemination of intimate images, a private notification suffices to order the removal of such content.

The Brazilian path has been based on the jurisprudence of the Superior Court of Justice (STJ) to strengthen the provisions contained in the MCI. The STJ has repeatedly indicated that: Internet intermediaries are not responsible for the illegal content that users generate; they cannot be forced to preemptively filter the content that users generate; they must remove the infringing content as soon as they become aware of the existence of illegal content hosted on their platforms, and must pay a compensation for damages if they fail to do so; finally, they must develop and maintain effective mechanisms to facilitate the identification of their users.

Another important general rule that seems to be derived from the jurisprudence of the STJ is that whoever seeks the removal of infringing content on the Internet must specify the URL where that content is located, in order to comply with the requirement of the MCI that the information delivered must be "clear and specific, allowing the unequivocal identification" of the location of the infringing content. On the one hand, this aims to protect app providers and, on the other, to mitigate potential negative impacts on freedom of expression.

Regarding the de-indexing of content in Internet search engines, there have been some contradictions. Before the MCI came into effect, the STJ considered that these providers were not responsible for user-indexed content; however, in a 2018 case regarding the "right to be forgotten", the same court determined by a tight majority that all content related to news that linked the plaintiff with a past case of fraud should be de-indexed whenever the search query linked the actual name of the plaintiff to the offense.

Various investigations have pointed out that the judicial mediation has guaranteed that unfounded content removal or blocking requests do not lead to the suppression of legitimate content. According to a study by the independent research center Internetlab, up to 2018, 60% of the court cases on requests to remove content were considered illegitimate, unfounded or abusive, which means that, if the platforms had immediately responded to these requests, a significant amount of legitimate content would have been removed[27]. The Brazilian case shows how the requirement of a judicial notification encourages the platforms to keep that content online, instead of preemptively removing it.

---

26 Presidency of the Republic, Civil House and Legal Affairs Subsection, *Civil Rights Framework for the Internet* (Marco Civil da Internet), Law No 12.965, April 23, 2014. Available at: http://pensando.mj.gov.br/marcocivil/wp-content/uploads/sites/2/2015/03/bill-12965.pdf

27 Oliva, Thiago, *Liability of intermediaries and the guarantee of freedom of expression on the Internet*, Internetlab, 23 Apr 2019. Available at: https://www.internetlab.org.br/pt/especial/responsabilidade-de-intermediarios-e-a-garantia-da-liberdade-de-expressao-na-rede/ [in Portuguese]

The MCI has also helped in stopping the use of extrajudicial notifications as a way of hindering freedom of expression, a ploy oftentimes used by public figures who use intimidation strategies aimed at silencing critical speech against them, often conveyed in the shape of humorous satire[28]. The intimidation strategy consists of sending extrajudicial notifications or initiating defamation proceedings, with the sole objective that the person or organization who created the content, or the hosting platform, are urged to block it. This strategy also serves to dissuade those who may be planning to produce the same type of content. The provisions on the liability exception regime prevent mere extrajudicial notifications from intimidating Internet platforms and encourage them to keep the content online as far as its removal is not ordered by a judicial authority, thus preserving the right of users to freedom of expression.

Currently, a judicial file challenging the constitutionality of article 19 of the MCI is being processed in court, its resolution still pending. If declared unconstitutional, the current intermediary liability model would be replaced by a mere private notification model, which would be a setback for the exercise of the rights of freedom of expression and access to information.

c. New front in Latin America: hate speech and disinformation

One of the most difficult problems to solve facing intermediary liability is the moderation of content on complex topics such as hate speech, disinformation or the right to honor.

Regarding the right to honor, the jurisprudential development in some Latin American countries in recent years has been consistent in exempting intermediaries from liability, despite the fact that in some cases this trend is not the rule. Both in Argentina and in Colombia, it has been jurisprudentially established that Internet intermediaries are not responsible for the content published by their users[29]. Rulings have been made based on subjective liability.

An equally complex issue is the use of social media and other Internet platforms to broadcast hateful messages and, consequently, the responsibility that OSPs may have in spreading that content. Given that there is no consensus on the definition or characterization of the term among countries, or even among platforms, it is very difficult to propose legislation that can regulate the dissemination of this content, without imposing at the same time excessive restrictions on freedom of expression[30].

---

28 *CTRL+X*, a project organized by the *Brazilian Association of Investigative Journalism* (Abraji), keeps an interactive infographic with complaints brought to justice by politicians and others against the dissemination of information and demanding content removal. Available at: https://www.ctrlx.org.br/#/infografico [in Portuguese]

29 Supreme Court of Argentina, *Ruling in the case 'Rodríguez, María Belén v. Google Inc. without damages'*, R.522.XLIX, 28 Oct 2014. The criterion was ratified in 2017 in *Gimbutas v. Google*, after the new Civil and Commercial Code came into force (Cfr. *Ruling in the case 'Gimbutas, Carolina Valeria v. Google Inc. without damages'*, CIV. 40500/2009/CS1, 12 Sep 2017. Available at: http://sjconsulta.csjn.gov.ar/sjconsulta/documentos/getDocumentosExterno.html?idAnalisis=739942 [in Spanish]); Constitutional Court of Colombia, *Ruling in the case 'Gloria v. la Casa Editorial El Tiempo'*, Dossier T-4296509, 12 May 2015. Available at: http://e.tribunalconstitucional.cl/img/Attachment/1929/T_277_15_Sentencia.pdf [in Spanish]

30 E.g., the definition of *hate speech* on Facebook is different from the definition of *hate speech* on Twitter.

One of these attempts was embodied in the *Constitutional Law Against Hatred, for Peaceful Coexistence and Tolerance* of Venezuela, which imposes on social media the obligation to remove content that incites "national hatred", among others, establishing fines and the blocking of sites for non compliance. Due to the vague nature of the definitions used in the law, as well as the lack of clarity about the procedures to be followed, there is a risk of this regulation being used to silence critical voices opposing the government of that country.

d. Fake News Law

In 2020, a bill was introduced in Brazil with the aim of regulating the massive dissemination of misinformation on various Internet platforms, such as social media or messaging services. In its original wording, the bill raised concern among activists and experts in freedom of expression, who claimed that it fostered the indirect liability of Internet intermediaries for the misinformation included in user-generated content. In June 2020, the Senate approved the final text of what became colloquially known as *Lei das Fake News*, or *Fake News Law*[31]. This bill has been strongly criticized, among other reasons, for promoting the excessive moderation of content by Internet platforms, hence violating the right to freedom of expression[32].

One of the main observations is related to the interpretation of the wording of article 12 § 2, within Section IV on Moderation Procedures, establishing the mechanisms to guarantee due process when measures based on the terms of use have been applied to a user. This guarantee includes the duty to notify the user about the grounds, the analysis process and the application of the measure, as well as about the deadlines and appealing procedures. According to § 2, this notification can be neglected in the cases established in the same subsection, among which: immediate damage that is difficult to repair; violation of the rights of children and adolescents; crimes typified in Law Nº 7,716 of January 5, 1989. This wording is interpreted by some specialists as an obligation to immediately remove content, which would imply granting the platforms extensive powers, resulting in them deciding what content may or may not circulate on the Internet.

Another observation regards the creation of the *Council for Transparency and Responsibility on the Internet*, established in Chapter IV. First, there is a fear that its independence cannot be guaranteed, since its installation and operation depend on the budget of the Brazilian Senate. Second, the competencies granted could encourage excessive content moderation by the platforms, since the Council is empowered –among other things– to dictate a code of conduct or evaluate the terms of service and moderation procedures of the platforms, all of this within the context of a bill that fails its purpose by lacking clarity when defining terms as relevant as *disinformation*. This vagueness could give rise to a dangerous breadth in the interpretation of the terms, allowing the inclusion in its scope of certain arbitrariness, which could then be transferred to the codes of conduct and affect other evaluation powers conferred on this entity trough Article 25.

---

31 Brazilian Senate, *Bill No. 2630 to establish the Brazilian Law of Freedom, Responsibility and Transparency on the Internet* (Fake News Law), 2 Jul 2020. Available at: https://static.poder360.com.br/2020/08/PL2630-comabte-fake-news.pdf [in Portuguese]

32 Dias Oliva, Thiago, The Brazilian Congress approach to disinformation and platform regulation puts freedom of expression at risk, CELE, 2020. Available at: https://observatoriolegislativocele.com/en/the-brazilian-congress-approach-to-disinformation-and-platform-regulation-puts-freedom-of-expression-at-risk/; see also: Brito Cruz, Francisco and Valente, Mariana, *Disinformation laws require more than good intentions*, Internetlab, 19 May 2020. Available at: https://www.internetlab.org.br/pt/imprensa/leis-para-desinformacao-exigem-mais-do-que-boas-intencoes/ [in Portuguese]

Finally, the debate's spotlight falls also on the application of sanctions to platforms (discussed in Chapter VI, article 31), considering them to be disproportionate and unnecessary when compared to other measures such as the advancement of independent fact-checking mechanisms or the promotion of positive actions such as digital literacy of the population, among others.

In view of the criticisms made of the approved text, a new document was informally presented, incorporating some of the observations. Even though the new document maintains some of the provisions observed in the official project, such as the Council's dependence on the Senate budget, it contains some provisions that are in line with the recommendations made by civil society initiatives[33]. Regarding transparency, for example, it specifies that the platforms must produce biannual transparency reports in Portuguese, which are to be made available on their websites, to inform about current procedures and decisions related to content moderation. According to article 7 of the bill, these reports must contain, among other information:

> "... II - total number of account and content moderation measures adopted due to compliance with the private terms of use of Internet application providers, indicating at an aggregate level, motivation and methodology used in the detection and application of the terms of use; III - total number of account and content moderation measures adopted and their reasons for compliance with court order, specifying the legal bases that justified the removal decision; IV - history and comparisons of practices, methodologies and metrics for the identification and moderation of accounts and content, including metrics on detection and response time; V - basic information about the team responsible for applying the moderation rules to accounts and content, also indicating the nature of the reviewer, whether automated or human; ... VII - total number of content identification measures and types of identification, removals or suspensions that have been reversed by the platform; VIII - mean time between the detection of irregularities and the adoption of measures in relation to the accounts and contents referred to in items II, III and IV; ... X - the criteria, methodologies and metrics used by its automated systems in monitoring and complying with its policies and terms of use; XI - the criteria, methodologies and metrics to assess the scope of the promoted content and advertising, subject to independent verification and audit; XII - information on the use and operation of automated systems, including the information, bases of operation and training of algorithms and the analysis of their impacts on circulation, availability, promotion, reduction of scope or elimination of content; and XIII - updates of the policies and terms of use made in the quarter, the date of modification and the justification for their adoption."

In turn, the initiative encourages platforms to enable a digital channel to receive complaints and establishes that when certain content or an account is subject to moderation measures, the platforms must inform the user about the basis of the measure, as well as the procedure and term to file an appeal against it.

---

33 Such as, for example, those made by OBSERVACOM, in its proposal *Standards for a democratic regulation of large platforms that guarantee freedom of expression online and a free and open Internet*, 7 Aug 2019. Available at: https://www.observacom.org/wp-content-uploads-2020-09-estandares-regulacion-grandes-plataformas-internet-pdf/ [in Spanish]

Chapter III, for its part, establishes the need to make advertised and electoral-driven content transparent (Section V, Subsection II). In this way, it requires that the account responsible for this type of content be identified, and that it is made visible to the user which content is promoted for electoral purposes.

These last points are relevant, given that one of the main criticisms being made of social media platforms (and of the original text of the bill) is that the application of their content moderation policies is far from transparent and generates doubts regarding the coherence in its implementation, which ends up affecting, mainly, the expression of vulnerable groups.

Finally, in its article 21 (one of the most controversial), the bill establishes the joint liability of platforms for the damages caused by third-party promoted content, when they did not comply with the obligation to request identity documents from the owners of such accounts. Although it may be positive to require commercial advertising content to be identified in the way regulated by the text, it should be questioned whether the same requirement is justified when it comes to regular users promoting their content. On the other hand, it would be reasonable to ask what is meant by damages, who will quantify them and how and, finally, if this joint liability does not contradict the MCI, since the assumption regulated in the proposal falls outside the scope of what the MCI establishes as grounds for intermediary liability.

IV. Regulation: comparative law. Experiences and proposals

a. The USA case: first exportation of intermediary immunity

i. The First Amendment and Section 230 of the CDA

The regulatory discussion in the United States has particularities that distinguish it from the debates in other countries, with the *First Amendment* and Section 230 of the *Communications Decency Act* (CDA) being the fundamental pillars of all action in the matter. The First Amendment, established in 1791, grants a protection to freedom of expression that, in its broadest interpretation, appears to be absolute:

> "Congress shall make no law respecting an establishment of religion, or prohibiting the free exercise thereof; or abridging the freedom of speech, or of the press; or the right of the people peaceably to assemble, and to petition the Government for a redress of grievances."[34]

The literality of the First Amendment is embedded in the American culture and reaches not only its citizens but also its companies[35]. This is particularly relevant in the discussion regarding the liability that Internet intermediaries may face, as companies claim their own constitutional right of expression when moderating (or not moderating) third-party content.

In the mid-90s, when the Internet was booming, the CDA was enacted in order to promote the decency of communications, and Section 230 was included to expressly enable the moderation of content on the Internet[36]. In summary, this norm grants the providers of interactive computer services immunity over the content created or published by their users. The providers should not be treated as publishers and, furthermore, they are granted immunity from any action taken in good faith to allow or remove third-party content from their platforms, "whether or not such material is constitutionally protected[37]." There are exceptions, that we will analyze in the following paragraphs, related to intellectual property and other federal crimes such as child pornography and sex trafficking.

---

34 *First Amendment to the United States Constitution* (Amendment I), adopted on December 15, 1791, as one of the ten amendments that constitute the Bill of Rights.

35 Bazelon, Emily, *Free speech will save our democracy*, The New York Times Magazine, 13 Oct 2020. Available at: https://www.nytimes.com/2020/10/13/magazine/free-speech.html

36 Kosseff, Jeff, *The Twenty-Six Words that Created the Internet*, Cornell Univ. Press, 2019. The title refers to the text of Subsection (c)(1) of Section 230 of the CDA.

37 47 U.S. Code § 230 – "(c) Protection for "Good Samaritan" blocking and screening of offensive material (1) Treatment of publisher or speaker: No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider. (2) Civil liability: No provider or user of an interactive computer service shall be held liable on account of— (A) any action voluntarily taken in good faith to restrict access to or availability of material that the provider or user considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally protected; or (B) any action taken to enable or make available to information content providers or others the technical means to restrict access to material described in paragraph (1)."

Section 230 of the CDA is part of the laws that Professor Eric Goldman calls "speech-enhancing statutes", which extend the protections established in the First Amendment[38]. According to him, Section 230 and the jurisprudential interpretation that has been made of it provide a broader protection of freedom of expression than the First Amendment itself, reaching not only defamatory content, but also any complaint based on third-party content[39]. Unlike the immunity provided in the First Amendment, the Section 230 immunity may apply even when the service provider becomes aware of an offense.

Before the advent of the Internet, the discussion revolved around the nature of the actors involved in the creation and circulation of content: publishers (newspapers) or distributors (book sellers). In 1959, the United States Supreme Court held that book sellers could not be held responsible for the content if they were unaware of it:

> "For, if the bookseller be criminally liable without knowledge of the contents, he will tend to restrict the books he sells to those he has inspected, and thus a restriction will be imposed by the States upon the distribution of constitutionally protected, as well as obscene, books."[40]

Broadly speaking, prior to Section 230 the discussion on intermediary liability centered on whether to apply to them the status of publisher or that of content distributor, with knowledge being the key element to determine the liability in each case. As a distributor, the platform would be liable in case of having knowledge of the crime, while the publisher could not claim ignorance given its level of control over the content.

Two cases from the early 1990s exemplify the dilemma. In *Cubby, Inc. v. CompuServe Inc.* (1991) the Southern District Court of New York understood that CompuServe was not responsible for the defamatory content of one of its subscribers, to the extent that it did not review or filter the content and simply "acted as a distributor[41]." In *Stratton Oakmont, Inc. v. Prodigy Services Co.* (1995), unlike Cubby, the court found that Prodigy (an online site similar to CompuServe) was responsible for the defamatory content of its subscribers, since the site took measures to remove certain offensive content and this, according to the court, meant Prodigy had taken an editorial role and that assimilated them to a publisher[42]. With this precedent, there was a clear incentive not to moderate content on the platforms: if they exercised any type of control or moderation, the service providers would have acquired effective knowledge and could be considered "publishers."

---

38 Goldman, Eric, *Why Section 230 Is Better Than The First Amendment*, 95 Notre Dame L. Rev. Reflection 33, 2019. Available at: https://scholarship.law.nd.edu/cgi/viewcontent.cgi?article=1074&context=ndlr_online; Eric Goldman is Professor of Law and Co-Director of the High Tech Law Institute at the Santa Clara University School of Law.

39 "If a plaintiff alleges that the defendant "knew" about tortious or criminal content, the defendant can still qualify for Section 230's immunity. The First Amendment does not require this result. For example, in Smith v. California, the Supreme Court held that the First Amendment prohibited a retail bookseller from being strictly liable for criminal obscenity. However, any scienter about obscenity could have exposed the bookseller to liability. Similarly, the First Amendment sometimes prevents strict liability for defamation, but sufficient scienter can override any First Amendment protection." Goldman, Eric, op. cit., Chapter II, § C, "Scienter Is Irrelevant to Section 230", p. 38.

40 U.S. Supreme Court, *Smith v. California*, 361 U.S. 147, pp. 152-154, 1959. Available at: https://supreme.justia.com/cases/federal/us/361/147/

41 U.S. District Court for the Southern District of New York, *Cubby, Inc. v. CompuServe Inc.*, 776 F. Supp. 135, 1991. Available at: https://law.justia.com/cases/federal/district-courts/FSupp/776/135/2340509/

42 New York Supreme Court, *Stratton Oakmont, Inc. v. Prodigy Services Co.*, 1995 WL 323710, 1995. Available at: https://h2o.law.harvard.edu/cases/4540

The Stratton Oakmont ruling did not have a good impact in the US Congress. The non-moderation of content brought with it certain problems that were concerning for society, especially, the situation regarding the safety of children online and defining which content is suitable for them. In response to this, a bipartisan initiative to regulate on the matter by amending the CDA to include Section 230 was passed in 1996 with almost unanimous vote (420-4). The initiative had two objectives: on the one hand, to promote the development of freedom of expression on the Internet and, on the other hand, to allow digital service providers to implement their own rules regarding third-party content and the promotion of children's safety online[43]. With the emergence of Section 230, OSPs obtained the immunity that allows them, today, not only to host and disseminate but also to moderate and curate the content of their users.

Some experts say that Section 230 of the CDA has paved the way for the development of the Internet as we know it today, allowing companies to grow without having to limit themselves, out of fear of being liable for the content they host on their platforms[44]. In recent years, however, Section 230 has been the target of various criticisms, and one that stands out is that, according to the interpretation of different courts, Section 230 has given an almost absolute power to platforms when it comes to content moderation (or the lack thereof), predominantly harming minorities and vulnerable groups[45].

ii. Limits to Section 230 Immunity

- **DMCA**

The Digital Millennium Copyright Act (DMCA) includes the exemption from direct and indirect liability of OSPs and other intermediaries[46]. In simple terms, for this exemption to be activated, it is sufficient for the intermediary OSP to block or remove allegedly infringing content after having received a private notification from the affected owner claiming such content is available on the provider's platform. In addition, they must comply with the requirements contained in § 512(c), which establishes as conditions that (1) they do not receive a financial benefit directly attributable to the infringing activity; (2) they are unaware of the presence of infringing material or of any fact or circumstance that could make the infringing material evident; and (3) upon receiving a notification from the copyright holders or their agents, they act expeditiously to remove, or disable access to, the material that is claimed to be infringing.

The Act also includes provisions that allow users to appeal content removal decisions. The mechanism proposed by the DMCA constitutes an exception to Section 230 of the CDA, and establishes a special and parallel mechanism for content potentially infringing copyright law.

---

43 See Electronic Frontier Foundation (EFF), *CDA 230: Legislative History*, 11 Feb 2013. Available at: https://www.eff.org/issues/cda230/legislative-history

44 Koseff, Jeff, op. cit.

45 "Unfortunately, Section 230 has also been used in court as a tool by platforms to avoid meaningfully confronting the role their products can play in furthering racial inequality; tech companies may hide behind this law to avoid discussing how their platforms affect civil rights issues, including housing, employment, and lending. In this moment of racial reckoning, the benefits of Section 230 must work in concert with platform accountability — not just for content moderation but also for civil rights." Lee, Bertram, Where the Rubber Meets the Road: Section 230 and Civil Rights, Public Knowledge, 12 Aug 2020. Available at: https://www.publicknowledge.org/blog/where-the-rubber-meets-the-road-section-230-and-civil-rights/

46 U.S. 105th Congress, *Digital Millennium Copyright Act*, PUBLIC LAW 105–304, 28 Oct 1998. Available at: https://www.govinfo.gov/content/pkg/PLAW-105publ304/pdf/PLAW-105publ304.pdf

Despite being a regulation that is limited in scope and provides for appeal mechanisms, the criticisms are many and from various fronts[47].

Since the main platforms that operate internationally are based in the United States, this rule is applied worldwide with undesired effects. In addition, as noted above, the standard has been exported to different countries and regions through FTAs that the United States has promoted with different countries over the years. Critics argue that the rule has served to over-censor content, including that of a political nature[48]. Notwithstanding this, the entertainment industry calls for stricter measures locally and globally.

- **FOSTA-SESTA**

More recently, in April 2018, a new exception to Section 230 was passed through the bills known as FOSTA-SESTA, aimed at dealing with sex trafficking online[49]. In short, it holds digital services responsible for third-party advertisements that promote sex work on their platforms. As a consequence of the norm, and in order to avoid any potential liability, some popular websites with classified ads sections like Craigslist or Reddit directly removed all their personal ads sections, including sections unrelated to sexual content. Those who oppose the norm, whose constitutionality has been challenged in court, argue that it promotes online censorship, has an impact on the removal of legal content and, importantly, does not solve the problem of sex trafficking[50].

iii. Current discussion: bills introduced in US Congress

The anger manifested by certain sectors towards social media networks, whether due to excesses or defects in the moderation of content, and added to the blame attributed to them for the alleged 2016 electoral interference, placed Section 230 in the most unstable position it has been since it was enacted in 1996. In this context, several bills were presented in the US. Congress by both blocks of the political spectrum, aiming to limit the immunity of Internet platforms. Among these bills, the *Ending Support for Internet Censorship Act*, which seeks to remove the immunity of large tech companies that have not had an external audit establish that their content moderation practices are politically neutral[51]; the *Limiting Section 230 Immunity to Good Samaritans Act*, whose objective is to provide a framework for "good faith" moderation and to compel Big Tech companies to update their terms of service accordingly in order to receive

47 Bertoni, Eduardo and Sadinsky, Sophia, *The use of the DMCA to limit freedom of expression*, in *Internet and Human Rights II: Contributions for the discussion of public policies in Latin America*, compiled by Eduardo Andrés Bertoni, CELE, Ediciones del Jinete Insomne, 2016. Available at: https://www.palermo.edu/cele/pdf/InternetyDDHHII.pdf [in Spanish]

48 Ibid.

49 U.S. 115th Congress, *Allow States and Victims to Fight Online Sex Trafficking Act* (FOSTA), H.R. 1865, 2017. Available at: http://congress.gov/bill/115th-congress/house-bill/1865; and *Stop Enabling Sex Traffickers Act* (SESTA), S.1693, 2017. Available at: http://congress.gov/bill/115th-congress/senate-bill/1693

50 Keller, Daphne, *SESTA and the Teachings of Intermediary Liability*, Center for Internet and Society (CIS), Stanford Law School, 2 Nov 2017. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3121296; see also Reynolds, Matt, *The strange story of Section 230, the obscure law that created our flawed, broken Internet*, WIRED, 24 Mar 2019. Available at: https://www.wired.co.uk/article/section-230-communications-decency-act

51 *Ending Support for Internet Censorship Act*, introduced by U.S. Senator Josh Hawley (R-MO), 19 Jun 2019. Available at: https://www.hawley.senate.gov/senator-hawley-introduces-legislation-amend-section-230-immunity-big-tech-companies

Section 230 immunity[52]; the EARN IT Act, whose main objective is to protect children from online sexual exploitation[53]; and the PACT Act, which focuses on increasing the transparency and consistency with which moderation decisions are made, and reinforcing the obligation of platforms to remove content that has been deemed illegal by a court order[54].

The discussion around Section 230 came to the fore on the US. political agenda at the hands of former President Donald Trump, who did not hide his anger against social media platforms, repeatedly accusing them of censoring him[55]. At the end of May 2020, after Twitter labeled two of his tweets on mail-in voting as misleading, Trump reacted by issuing an executive order requesting the review of "Federal spending on advertising and marketing paid to online platforms", and instructing the US Attorney General to "assess whether any online platforms are problematic vehicles for government speech due to viewpoint discrimination." Furthermore, the order accuses platforms of abusing the immunity provided in Section 230, and questions the scope of liability regarding what he sees as editorial conduct in the following terms (bolding is ours):

> "Section 230 was not intended to allow a handful of companies to grow into titans controlling vital avenues for our national discourse under the guise of promoting open forums for debate, and then to provide those behemoths blanket immunity when they use their power to censor content and silence viewpoints that they dislike. When an interactive computer service provider removes or restricts access to content and its actions do not meet the criteria of subparagraph (c)(2)(A), it is engaged in editorial conduct. It is the policy of the United States that such a provider should properly lose the limited liability shield of subparagraph (c)(2)(A) and be exposed to liability like any traditional editor and publisher that is not an online provider."[56]

Following the executive order, the DoJ presented in June 2020 its considerations on Section 230 of the CDA, as well as legislative guidelines to reform it[57]. Among them, it favors limiting immunity for "bad samaritans", as well as in specific cases where "a platform had actual knowledge or notice that the third party content at issue violated federal criminal law or where the platform was provided with a court judgment that content is unlawful in any respect."

---

52 *Limiting Section 230 Immunity to Good Samaritans Act*, introduced by U.S. Senators Marco Rubio (R-FL) and Josh Hawley (R-MO), 17 Jun 2020. The law would "prohibit Big Tech companies from receiving Section 230 immunity" if they breach their "contractual duty of good faith", by taking an action "that includes (...) the intentionally selective enforcement of the terms of service", among others. Available at: https://www.rubio.senate.gov/public/index.cfm/press-releases?ContentRecord_id=47276D77-62D6-4E04-9FA2-1CD761179B90

53 *EARN IT Act*, introduced by U.S. Representatives Sylvia Garcia (D-TX) and Ann Wagner (R-MO), 30 Sep 2020. The acronym *EARN IT* stands for "Eliminating Abusive and Rampant Neglect of Interactive Technologies." Available at: https://sylviagarcia.house.gov/media/press-releases/representatives-sylvia-garcia-and-ann-wagner-introduce-bipartisan-earn-it-act

54 *PACT Act*, introduced by U.S. Senators Brian Schatz (D-HI) and John Thune (R-SD), 24 Jun 2020. The acronym *PACT* stands for "Platform Accountability and Consumer Transparency." Available at: https://www.schatz.senate.gov/press-releases/schatz-thune-introduce-new-legislation-to-update-section-230-strengthen-rules-transparency-on-online-content-moderation-hold-internet-companies-accountable-for-moderation-practices

55 Fung, Brian, Nobles, Ryan and Liptak, Kevin, *Trump signs executive order targeting social media companies*, CNN, 29 May 2020. Available at: https://edition.cnn.com/2020/05/28/politics/trump-twitter-social-media-executive-order/index.html

56 Trump White House Archives, *Executive Order on Preventing Online Censorship*, 28 May 2020. Available at: https://trumpwhitehouse.archives.gov/presidential-actions/executive-order-preventing-online-censorship/

57 The U.S. Department of Justice Archives, *Department of Justice's review of Section 230 of the Communications Decency Act of 1996*, 17 Jun 2020. Available at: https://www.justice.gov/archives/ag/department-justice-s-review-section-230-communications-decency-act-1996

The discussion also became notorious in the judiciary. In October 2020, Justice Thomas, member of the Supreme Court of the United States, held that courts have interpreted Section 230 immunity too broadly, extending beyond what the text of the norm indicates[58]. In his arguments, Thomas analyzed the distinction between publishers and distributors and questioned the interpretation that different courts adopted when extending the immunity of Section 230 on the latter. Along these lines, a few days later, the *Federal Communications Commission* (FCC) announced its intention to clarify the meaning of Section 230[59].

Criticism has also emerged from sectors of civil society and academia. The United States is the only country that has such extensive protections for intermediaries, and the word immunity seems more appropriate in that context than in any other. The fact that platforms can rely on Section 230 and the First Amendment to protect them against any civil liability has shielded them from basic civil rights demands, such as non-discrimination. *Public Knowledge* recently argued that, while Section 230 is a great protection that must be defended, the interpretation that has been made of it requires some adjustments[60]. Among other things, they argue that platforms have used these laws to fight lawsuits related to discriminatory advertising practices in housing, loans, and other areas. The problem lies not only with large platforms but with all platforms, beyond social media or search engines. The author proposes excluding advertising from the protections of Section 230, as a mechanism that would be less harmful than excluding specific types of content such as, for example, hate speech.

In January 2021, after the deplorable events in and around the US Capitol and the decision by online platforms (both large and small) to suspend the accounts of then President Donald Trump (currently referred to as "deplatforming"), the discussion about intermediary liability regarding content moderation was reignited. Accusations of censorship from the most conservative sectors inside and outside the United States did not take long to emerge. At the same time, online platforms were criticized for their lack of transparency and consistency when applying the same rules in the rest of the world. In this context, issues such as incitement to violence, hate speech and political or public interest speech came to the center of the discussion. Issues such as the applicability of moderation to different layers of the Internet, the *common carrier* obligations and its extension to the Internet, or the market concentration that a group of technology companies has achieved and its impact on public discussion also gained notoriety. All

---

58 "Yet there are good reasons to question this interpretation. First, Congress expressly imposed distributor liability in the very same Act that included § 230. Section 502 of the Communications Decency Act makes it a crime to "knowingly . . . display" obscene material to children, even if a third party created that content. 110 Stat. 133–134 (codified at 47 U. S. C. § 223(d)). This section is enforceable by civil remedy. 47 U. S. C. § 207. It is odd to hold, as courts have, that Congress implicitly eliminated distributor liability in the very Act in which Congress explicitly imposed it." (...) "Extending § 230 immunity beyond the natural reading of the text can have serious consequences. Before giving companies immunity from civil claims for "knowingly hosting illegal child pornography," Bates, 2006 WL 3813758, *3, or for race discrimination, Sikhs for Justice, 697 Fed. Appx., at 526, we should be certain that is what the law demands." Supreme Court of the United States, *Statement of Thomas, J.*, 592 U. S. ___ (2020), § 1.A, p. 5 and § II, p. 10, 13 Oct 2020. Available at: https://www.law.cornell.edu/supremecourt/text/19-1284

59 "As elected officials consider whether to change the law, the question remains: What does Section 230 currently mean? Many advance an overly broad interpretation that in some cases shields social media companies from consumer protection laws in a way that has no basis in the text of Section 230. The Commission's General Counsel has informed me that the FCC has the legal authority to interpret Section 230. Consistent with this advice, I intend to move forward with a rulemaking to clarify its meaning." Federal Communications Commission (FCC), *Chairman Pai Statement on Section 230*, 15 Oct 2020. Available at: https://www.fcc.gov/document/chairman-pai-statement-section-230

60 Lee, Bertram, *Where the Rubber Meets the Road: Section 230 and Civil Rights*, Public Knowledge, 12 Aug 2020. Available at: https://www.publicknowledge.org/blog/where-the-rubber-meets-the-road-section-230-and-civil-rights/

these issues are currently being studied, developed and analyzed by States, academia and civil society.

Although there are many arguments supporting the modification of Section 230, experts and activists in the United States who fight for human rights, especially freedom of expression, criticize these initiatives, noting that limitating Section 230 immunity with confusing operational rules and vague wording generates an incentive towards content removal overreach, that would inevitably affect legitimate speech[61]. EFF and ACLU are just a couple of the many organizations that came out against these projects[62].

b. European Union and others

The legal framework for Internet intermediary liability in the European Union is made up of different directives, codes of conduct, recommendations and other standards. The *Directive on Electronic Commerce* of 2000 is perhaps the system's cornerstone, but its application by the Member States has not been homogeneous[63].

The push for new regulatory or semi-regulatory initiatives, including some whose application is optional, has gained strength in recent years, seeking to address some specific challenges such as copyright, personal data protection, terrorism, hate speech or disinformation[64]. In this context, it became clear that the *Directive on Electronic Commerce* deserved an update, more in line with the discussion on liability of Internet intermediaries that is currently being presented around the world.

---

61 Citron, Danielle Keats and Wittes, Benjamin, *The Internet Will Not Break: Denying Bad Samaritans Section 230 Immunity*, Fordham Law Review, Forthcoming, U of Maryland Legal Studies Research Paper No. 2017-22, 24 Jul 2017. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3007720

62 Mullin, Joe, *Urgent: EARN IT Act Introduced in House of Representatives*, Electronic Frontier Foundation (EFF), 2 Oct 2020. Available at: https://www.eff.org/deeplinks/2020/10/urgent-earn-it-act-introduced-house-representatives; For its part, ACLU stated that "By requiring platforms to broadly monitor and censor speech to which children might be exposed online, the EARN IT Act's commission may recommend best practices that disproportionately censor, among other things: sex education materials, online support systems and communities for youth who are transgender or non-binary, and all other youth who are in any way questioning their gender or sexual identity to communicate with each other and with community members, any sex-related speech, particularly the speech of sex workers and of those in the sex industry, and any communication or speech involving youth. Paradoxically, the best practices could harm children's ability to engage fully and experience the tremendous benefits to education and enrichment the internet offers." Ruane, Kate, *The EARN IT Act is a Disaster for Online Speech and Privacy, Especially for the LGBTQ and Sex Worker Communities*, American Civil Liberties Union (ACLU), 30 Jun 2020. Available at: http://aclu.org/news/free-speech/the-earn-it-act-is-a-disaster-for-online-speech-and-privacy-especially-for-the-lgbtq-and-sex-worker-communities/

63 The European Parliament and the Council of the European Union, *Directive 2000/31/EC on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market (Directive on Electronic Commerce)*, 8 Jun 2000. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32000L0031&from=EN

64 Proposals that are not enforced include the *Code of conduct on countering illegal hate speech online*, signed by the European Union, Facebook, Google, Microsoft and Twitter on 31 May 2016 (Available at: https://ec.europa.eu/info/sites/info/files/code_of_conduct_on_countering_illegal_hate_speech_online_en.pdf); and the *Proposal for a Regulation of the European Parliament and of the Council on preventing the dissemination of terrorist content online*, published by the European Commission on 12 Sep 2018 (Available at: https://eur-lex.europa.eu/resource.html?uri=cellar:dc0b5b0f-b65f-11e8-99ee-01aa75ed71a1.0001.02/DOC_1&format=PDF); Iniciatives on specific challenges include the *Audiovisual Media Services Directive of 10 March 2010*, consolidated version 18 Dec 2018 (Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32010L0013&from=EN) and the *Regulation on promoting fairness and transparency for business users of online intermediation services of 20 June 2019*, of the European Parliament and the Council of the European Union (Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32019R1150&from=EN)

With the aim of adapting it to the complexities of today, in early 2020 the *European Commission* announced its intention to work on a set of standards focused on two pillars: first, establishing a legal framework for digital services with clear liabilities and rules, which considers the risks assumed by users and protects their rights through a monitoring system; second, establish *ex ante* rules for large online platforms who act as "gatekeepers", in order to ensure that they behave fairly, allow competition, and encourage innovation. By the end of the year, this set of standards became the *Digital Services Act package*[65].

In this section, we will briefly review each of these regulations and initiatives, in order to identify the focus of the regulatory discussion that is particular to the European region.

i. e-Commerce Directive

The *Directive on Electronic Commerce*, in force since the year 2000, includes principles on intermediary liability and the prohibition of implementing "a general obligation on providers to monitor the information that they transmit or store[66]." The directive establishes that "Member States shall ensure that the service provider is not liable" for third-party content, on certain conditions, among which: the ISP must keep a passive position regarding that content, and; they must act expeditiously to remove or to disable access to illegal content, after "a court or an administrative authority has ordered such removal or disablement[67]." The directive does not define ISPs, but rather mentions the three types of activities that could be conditionally exempted from liability: *mere conduit*, *caching* and *hosting*[68].

Although it does not include an explicit *notice-and-takedown* provision, platforms can resort to these mechanisms to avoid falling into non-compliance. In the same sense, the prohibition to implement a general monitoring obligation on content does not prevent platforms from carrying out certain monitoring, either to ensure effective compliance with their terms of service or to avoid liability risks. On the other hand, unlike what happens in the United States, the European directive does not include the "good samaritan" CDA 230 protection for Internet platforms regarding the moderation of third-party generated content. Therefore, platforms must exercise caution when moderating, to avoid losing their neutral position and incurring liability.

Regarding illegal content, it is established that the service provider will not be liable for the stored data unless it has *actual knowledge* of illegal activity or information, or is aware of "facts or circumstances from which the illegal activity or information is apparent." Upon becoming aware, they must act expeditiously to remove or disable access to the information concerned. The directive's lack of clear provisions regarding the (implicit) notice-and-takedown mechanism resulted, in practice, in its implementation varying from one State to another.

In 2018, the European Commission published a report on the implementation of what they call the *notice-and-action procedures* within the Member States, from which it appears that some

65 European Commission, *The Digital Services Act package*, 15 Dec 2020; comprised of the *Digital Services Act* (DSA) and the *Digital Markets Act* (DMA). Available at: https://ec.europa.eu/digital-single-market/en/digital-services-act-package

66 EU Directive on Electronic Commerce, Art. 15 § 1.

67 Ibid. Arts. 12 § 1 and 13 § 1(e).

68 Ibid. Arts. 12-14.

of them did adopt specific rules on the type of illegal material reached and the procedural aspects, including what type of information should the notice include[69]. Furthermore, the report indicates that, with some exceptions, the States consider that judicial and administrative orders are necessary to obtain "actual knowledge" of the illegal content. Regarding notifications, they can trigger actual knowledge "when the material is manifestly illegal, but this must be evaluated on a case-by-case basis as there is no common guidance emerging from national case law[70]."

Notwithstanding their radical importance for setting a regulatory framework in the European sphere, the different regulatory, coregulatory and voluntary initiatives promoted by the European Union have generated numerous exceptions as well as alternative, complementary and even supplementary interpretations of the Directive on Electronic Commerce. In the midst of this diversification of standards, some of them possibly contradictory, the proposal for the *Digital Services Act* emerged, a new framework law to address the comprehensive regulation of digital communication services in Europe.

ii. Digital Services Act

As early as 2016, the European Commission expressed the need to harmonize the regulatory framework of the digital economy in the European Union, arguing that:

> "... in order for Europe to reap the full benefits from the platform economy and stimulate growth in European platform start-ups, self-evidently, there cannot be 28 different sets of rules for online platforms in a single market. Differing national or even local rules for online platforms create uncertainty for economic operators, limit the availability of digital services, and generate confusion for users and businesses."[71]

Since then, the Commission launched the preparatory work for the *Digital Services Act* (DSA), an initiative that comprehensively updates the regulation of the digital ecosystem as part of the European *Digital Single Market* strategy.

Between June and September 2020, amid the COVID-19 pandemic, a public consultation process was carried out aimed at citizens and organizations inside and outside Europe, in order to gather the information and evidence that would help finalize the outlining of the regulation[72]. In that same period, three bodies within the European Parliament presented reports with their

---

69 European Commission, *Overview of the legal framework of notice-and-action procedures in Member States*, SMART 2016/0039, July 2018. Available at: https://op.europa.eu/en/publication-detail/-/publication/c5fc48ac-2441-11e9-8d04-01aa75ed71a1/language-en/format-PDF/source-102736628

70 Ibid., § 3.1.1, p. 2.

71 European Commission, *Online Platforms and the Digital Single Market. Opportunities and Challenges for Europe*, Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, COM(2016) 288 final, 25 May 2016, § 4, p. 4.

72 Experts, academics and civil society organizations from around the world presented their comments. See, among others, *Access Now's Position on the Digital Services Act Package*, Position Paper Series, September 2020. Available at: http://accessnow.org/cms/assets/uploads/2020/10/Access-Nows-Position-on-the-Digital-Services-Act-Package.pdf; Even the UN High Commissioner for Human Rights, Michelle Bachelet, made public a Letter addressed to the President of the European Commission on 7 Sep 2020 with her comments. Available at: https://europe.ohchr.org/EN/Stories/Documents/2020 09 07 Letter HC to EC President.pdf

own considerations[73]. The three reports agreed on some relevant points, such as maintaining the principle of immunity established in the *Directive on Electronic Commerce* (also known as "safe harbor"), and maintaining the prohibition of imposing a general monitoring obligation[74]. Additionally, they agreed on implementing a new notice-and-action procedure regarding illegal content shared by users, reinforcing transparency and reparation obligations (including an independent mechanism for dispute resolution), and suggested the creation of a specialized supervisory body within the European Union.

In December 2020, after several months of deliberation, the Commission presented a package of proposals including two bills: the *Digital Services Act* (DSA) and the *Digital Markets Act* (DMA).

The DSA, according to the Parliament, promotes innovation, development and competitiveness, in addition to facilitating the growth of smaller platforms and startups, with its main focus on the rights of European citizens.. An article on the Commission's official website states that in this project "(t)he responsibilities of users, platforms, and public authorities are rebalanced according to European values, placing citizens at the center[75]."

The proposal broadly maintains the liability rules, as well as the prohibition of imposing a general monitoring obligation, that were set out in the e-Commerce Directive. Its provisions reach different types of digital service providers and the obligations vary according to the nature of their services, their size and their impact. There are basic obligations for providers of intermediary services, hosting services, online platforms and very large online platforms, regarding transparency reports (which are now mandatory), terms of service (which must take into account fundamental rights), cooperation with national authorities, notice-and-action mechanisms, and others.

Along these lines, the DSA aims to improve the mechanisms for removing illegal content and the protection of fundamental rights online, as well as creating a system of public supervision over platforms, particularly the larger ones.

The DMA, on the other hand, is focused on the competitive behavior of large platforms, promoting that those that behave as "gatekeepers" do so fairly. The regulation aims to guarantee that those who depend on gatekeepers to operate may offer their services in a fairer environment, thus promoting innovation and the development of tech start-ups, allowing consumers to access more and better services, and granting them the opportunity to choose the provider they prefer. As with the DSA, the DMA is a good starting point, although some experts have already raised their voices on provisions that need improvement[76].

---

73 I.e., the European Parliament committees on the *Internal Market and Consumer Protection* (IMCO), on *Legal Affairs* (JURI) and on *Civil Liberties, Justice and Home Affairs* (LIBE).

74 Existing both in the *Directive on electronic commerce* (see f.n. 55) and in Art. 17 of the *Directive on Copyright and Related Rights in the Digital Single Market*, Directive (EU) 2019/790 of the European Parliament and of the Council, 17 April 2019. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32019L0790&from=EN

75 European Commission, *The Digital Services Act: ensuring a safe and accountable online environment*, 15 Dec 2020. Available at: https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/digital-services-act-ensuring-safe-and-accountable-online-environment_en

76 See comments by Cory Doctorow and Christoph Schmon: *The EU's Digital Markets Act: There Is A Lot To Like, but Room for Improvement*, 15 Dec 2020. Available at: https://www.eff.org/deeplinks/2020/12/eus-digital-markets-act-there-lot-room-improvement

The DSA and DMA proposals primarily address the regulation of key aspects such as transparency, due process and accountability. From this point of view, the proposals were well received by civil society, academia and States, not only in Europe but also globally[77]. The truth is that the European legislation process is, in general terms, a more open and less reactive process than many others seen in other parts of the world, and therefore a good setting for the development of this debate. Despite these conditions, the proposal has not been exempt from criticism from civil society, which points to the insufficiency of the substantive and procedural safeguards proposed to ensure the protection of the fundamental rights of users and harmonize procedures that allow them to effectively defend their rights in cases of abuse by private actors[78]. The results of this process, of course, are not guaranteed.

iii. Directive on Copyright and Related Rights in the Digital Single Market

The *Directive on Copyright and Related Rights in the Digital Single Market*, which came into force on June 2019, is intended to modernize Copyright regulations in the digital age[79], as well as to address some of the problems mentioned earlier in this document, namely dealing with some of the claims brought forward by traditional media. The Directive is considered controversial, largely due to its articles 15 and 17[80].

Article 15 deals with the "protection of press publications concerning online uses." Shortly, it establishes that service providers must recognize and pay the rights to the publications of press publishers, excluding private or non-commercial uses, hyperlinking and the use of individual words or very short extracts of a press publication. In addition, it commands Member States to provide that "authors of works incorporated in a press publication receive an appropriate share of the revenues that press publishers receive for the use of their press publications by information society service providers." The incorporation of this article responds to an ongoing discussion worldwide on how to handle the costs of the press industry for the production of information that later circulates online, and how to deal with search engines benefiting from a large share of the advertising that was formerly controlled by the media itself.

The Internet industry argued that the role of search engines in the dissemination of media content and the redirection of users to the media websites is beneficial to them, and that creating a licensing system would only discourage their sharing of information and press articles from

---

77 See comments by Christoph Schmon and Karen Gullo: *European Commission's Proposed Digital Services Act Got Several Things Right, But Improvements Are Necessary to Put Users in Control*, EFF, 15 Dec 2020. Available at: http://eff.org/deeplinks/2020/12/european-commissions-proposed-regulations-require-platforms-let-users-appeal; and by Access Now: DSA: European Commission delivers on the first step toward systemic rules for online platforms, 15 Dec 2020. Available at: https://www.accessnow.org/dsa-systemic-rules-for-online-platforms/

78 See: Berthélémy, Chloé and Penfrat, Jan, *Platform Regulation Done Right (EDRi Position Paper on the EU Digital Services Act)*, EDRi, 9 Apr 2020. Available at: https://edri.org/wp-content/uploads/2020/04/DSA_EDRiPositionPaper.pdf

79 Along with its new provisions, the directive amends the previous *Database Directive* (Directive 96/9/EC of 11 March 1996 on the legal protection of databases) and the *Information Society Directive* (Directive 2001/29/EC of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society). Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32019L0790&from=EN

80 Proof of its controversial nature was its parliamentary treatment, passing with 348 votes in favor, 274 votes against and 24 abstentions.

the media[81]. Google was one of the first platforms to manifest their opposition to this provision, by reducing the length of their snippets in France and removing the Google News platform in Spain, among other measures[82].

From civil society, the incorporation of this article received criticism both in Europe and worldwide. The reproduction and diffusion of news through different channels should be not only allowed, but encouraged. Framing the discussion about the sustainability of the media in an area as traditionally conservative and proprietary as copyright implies to some extent ignoring the role that the media plays in a democracy[83].

With respect to Article 17 of the Directive, on the "use of protected content by online content-sharing service providers", it establishes the procedure that these service providers must follow when giving access to copyright-protected content uploaded by their users, in order to avoid liability for "unauthorized acts of communication to the public." To achieve this, service providers fulfilling the definition of online content-sharing service provider given in recital § 62 and article 2(6) of the directive, must demonstrate that they have:

> "(a) made best efforts to obtain an authorization, and
>
> (b) made, in accordance with high industry standards of professional diligence, best efforts to ensure the unavailability of specific works and other subject matter for which the rightholders have provided the service providers with the relevant and necessary information; and in any event
>
> (c) acted expeditiously, upon receiving a sufficiently substantiated notice from the rightholders, to disable access to, or to remove from their websites, the notified works or other subject matter, and made best efforts to prevent their future uploads in accordance with point (b)."[84]

The norm does not impose an obligation on platforms to implement upload filters but, as a result of its vague wording (e.g., it does not elaborate on the exact meaning of "best efforts"), companies dealing with content on a large scale could be pushed to resort to these types of

---

81 An interesting study on news consumption and online behavior: Athey, Susan; Mobius, Mark and Pal, Jeno, *The Impact of Aggregators on Internet News Consumption*, Stanford Business, 11 Jan 2017; In Germany, a few years earlier, a local copyright rule was enacted which, like the Copyright Directive, allowed publishers of press releases to charge Google for the snippet views it shared. However, after seeing how Google's decision to cut down snippet fragments was dramatically dropping the traffic to their websites, the publishers allowed Google to display them at no cost. Available at: https://www.gsb.stanford.edu/gsb-cmis/gsb-cmis-download-auth/406636

82 Richard Gingras, Vice President of News at Google Inc, *New copyright rules in France: our compliance with the law*, Official blog of Google France, 25 Sep 2019. Available at: https://france.googleblog.com/2019/09/comment-nous-respectons-le-droit-dauteur.html [in French]

83 See *The European Directive on Copyright and its impact on users in Latin America and the Caribbean: a perspective from civil society organizations*, Al Sur, 2019: "Articles 15 and 17 are there to balance, theoretically, the profits of big platforms that depend on user-generated content (UGC), earned to the detriment of the authors and the media copyrights. However, they will affect the entire Internet ecosystem. Large companies will have the resources to implement these mechanisms (at the expense of users' freedom of expression), but other small or new services will probably see an increase in their costs that could directly affect their services and even their economic survival. In other words, with the alleged intention of punishing the big and dominant players, this measure will end up making them stronger while hurting new and small players." Available at: https://www.alsur.lat/sites/default/files/2020-04/Version Final _ La Directiva Europea de Derecho de Autor y su impacto en los usuarios de America Latina y el Caribe.pdf [in Spanish]

84 *European Directive on Copyright and Related Rights in the Digital Single Market*, Art. 17 § 4.

mechanisms in order to avoid liability. Added to this, the complexity and high cost of these technologies suggest that only the largest companies will be able to comply with the provision, discouraging the emergence of smaller companies and further reducing the level of competition in the Internet ecosystem.

Like the e-Commerce Directive, the Copyright Directive establishes that its application "shall not lead to any general monitoring obligation[85]." In practice, however, this prohibition seems to contradict the provisions of the previous paragraphs.

This model has been criticized, mainly, for two reasons: first, because the only way for a user to request the restitution of allegedly infringing content that was removed is subsequently presenting a claim before the judicial, a procedure that, due to its onerous nature, may limit the exercise of freedom of expression; secondly, because in practice the large number of content removal requests, many times made in an abusive way, has led to the automation of the process, causing in some cases the removal of content that does not infringe copyright.

In line with the aforementioned criticisms, David Kaye, former UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, had warned about the scope of the norm even before it was introduced in parliament. He expressly raised concerns about possible prior censorship, the ineffective recourse mechanisms and the disproportionate burden on NGOs and small companies[86].

c. Recent national regulations: NetzDG in Germany, Avia Law in France, the Online Harms White Paper in the UK, the Australian case and the Austrian proposal

The pressure on Internet companies to remove problematic content has been felt strongly in democratic societies around the world, especially in recent years. In Germany, the federal parliament passed in 2017 the *Network Enforcement Act* (NetzDG, also known as the *Facebook Act*), which forces platforms to remove content considered unlawful by local legislation in periods as short as "within 24 hours of receiving the complaint", and establishes substantial regulatory fines as sanction for non-compliance[87]. Unlike other legislations, the German law creates a Committee of Experts to handle questions by the companies regarding the definition of certain terms or the application of the norm in problematic cases.

In line with the German law, in 2020 the French National Assembly adopted the so-called *Avia Law* (named after its main promoter, deputy Laetitia Avia), which forces companies to remove within 24 hours manifestly illegal content and content that incites hatred or violence[88]. In the

---

85 Ibid., Art. 17 § 8.

86 UN Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, *Internal Communication Clearance Form*, 13 Jun 2018. Available at: https://www.ohchr.org/Documents/Issues/Opinion/Legislation/OL-OTH-41-2018.pdf

87 Bundestag (German Federal Parliament), Act to Improve Enforcement of the Law in Social Networks (Network Enforcement Act), Jun 2017. Available at: https://www.bmjv.de/SharedDocs/Gesetzgebungsverfahren/Dokumente/NetzDG_engl.html

88 French National Assembly, Bill aimed at combating hateful content on the internet (Avia Law), adopted on 13 May 2020. Available at: https://www.assemblee-nationale.fr/dyn/15/textes/l15t0419_texte-adopte-seance.pdf; Cfr. Law No. 2020-766 aimed at combating hateful content on the internet, enacted by President Emmanuel Macron on 24 Jun 2020. Available at: https://www.legifrance.gouv.fr/download/file/CP05NSqcPl5lPNu3MsP2PSu1fmt64dDetDQxhvJZNMc=/JOE_TEXTE [both in French]

case of terrorist propaganda, the content must be removed or made inaccessible within one hour of the notification. A month later, the French Constitutional Council considered that the text was largely contrary to the Constitution, in particular because it disproportionately infringes freedom of expression by not meeting the requirements of necessity and proportionality to the objective pursued. As a result, the bill was modified and purged of its provisions deemed unconstitutional, including the one mentioned above.

Similarly concerned about the dissemination of harmful and illegal content on the Internet, in April 2019 the British government published the *Online Harms White Paper*, in which it presented a plan to develop a new regulatory framework for tech companies[89]. According to the document, the future regulatory framework will enforce on Internet companies certain *codes of practice* in the digital space, with the government establishing a new statutory *duty of care*[90]. In addition, it stipulates that companies may be liable for *online harms*, presenting a list that is "by design, neither exhaustive nor fixed", which includes illegal content or activities and other content or activities considered harmful but not necessarily illegal. The list is organized in three categories and presents vague and ambiguous definitions, without further explanation[91]. Although the white paper states that companies cannot be required to carry out general monitoring of all content, it does consider that specific monitoring should be required for "tightly defined categories of illegal content", namely, threats to national security, physical safety of children and terrorism[92]. The proposal has aroused severe criticism for establishing a monitoring obligation and creating an administrative unit that could define which content could be blocked or removed[93].

There have been strong efforts to limit or control hate speech and illegal content on the Internet in Australia as well. In April 2019, the Australian Congress approved an amendment to its criminal code, by which it introduced the criminal liability of Internet platform executives who fail to remove what it calls *abhorrent violent material*[94]. It is particularly concerning that the project has been processed and approved without due public consultation and in an expeditious manner.

Lastly, the Austrian parliament is currently debating the Bill *Communication Platforms Act* (KoPl-G), which contains provisions similar to the German regulation discussed above, such as the 24-hour period to remove illegal content, with some modifications regarding transparency

---

89 HM Government, *Online Harms White Paper*, Presented to Parliament by the Secretary of State for Digital, Culture, Media & Sport and the Secretary of State for the Home Department by Command of Her Majesty, April 2019. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/793360/Online_Harms_White_Paper.pdf

90 Courtney Radsch, of the Committee to Protect Journalists (CPJ), makes an interesting point about this: "In the U.K., the government is developing a new regulatory framework to mitigate "online harms" by imposing a "duty of care" on platforms based on the perceived failure of voluntary initiatives. If a voluntary database that could identify terrorist content exists and is not used, it would not be far-fetched to expect that this could be seen as abrogating a platform's duty of care." Radsch, Courtney, *GIFCT: Possibly the Most Important Acronym You've Never Heard Of (Hint: It Involves the Internet and Terrorism)*, Just Security, New York University School of Law, 30 Sep 2020. Available at: http://justsecurity.org/72603/gifct-possibly-the-most-important-acronym-youve-never-heard-of/

91 *Online Harms White Paper*, Table 1: Online harms in scope, Part 1 § 2, p. 31.

92 Ibid., Part 2 § 3.12, p. 43.

93 For some of the criticism about the U.K. Online Harms White Paper, see for example: Goldman, Eric, *The U.K. Online Harms White Paper and the Internet's Cable-ized Future* (2019). Ohio State Tech. L.J., Forthcoming. Available at: https://ssrn.com/abstract=3438530 or http://dx.doi.org/10.2139/ssrn.3438530

94 The Parliament of the Commonwealth of Australia, *Criminal Code Amendment (Sharing of Abhorrent Violent Material) Bill 2019*, A Bill for an Act to amend the Criminal Code Act 1995, and for related purposes, 4 Apr 2019. Available at: https://parlinfo.aph.gov.au/parlInfo/download/legislation/bills/s1201_aspassed/toc_pdf/1908121.pdf

obligations[95]. Some civil society organizations have raised their voices against the bill for considering that the proposal "significantly interferes with the right to freedom of expression" by delegating censorship powers to private companies[96].

d. Usual criticisms of these regulatory systems

The recent regulatory initiatives have been criticized primarily for their adverse effects on human rights, with a particular focus on the right to freedom of expression. The threat of liability, including significant financial fines or, even worse, prison sentences in the case of the Australian amendment, added to the pressure to resolve issues in extremely short terms, create an incentive for excessive removal of content, a practice commonly known as "private censorship". The concern is that platforms will likely respond to this pressure by eliminating not only content that is allegedly or manifestly illegal but also legal content, thus violating the protection of the right to freedom of expression, recognized in countless international norms.

Notice-and-takedown mechanisms ("notice-and-action" in the European terminology) will remain a cause for concern as long as they do not include clear rules to guide their implementation, do not provide the necessary guarantees to prevent legal content from being removed, and do not establish appeal mechanisms and adequate remedies when mistakes have been made. The Copyright Directive, for example, enables a mechanism that could be abused not only by companies (by setting upload filters) but also, indirectly, by governments seeking to censor hostile political content. Furthermore, the directive incorporates some obligations that may be impossible to fulfill in practice, such as the one provided in its article 17 (bolding and underlining are ours):

> "Complaints submitted under the mechanism provided for in the first subparagraph shall be processed without undue delay, and decisions to disable access to or remove uploaded content shall be subject to human review."[97]

In large-scale content moderation, requiring companies to act expeditiously (or "without undue delay") prevents them from carrying out an adequate analysis of each case and indirectly pressures them to implement automation mechanisms to assist in the rapid detection of illegal or problematic content[98]. Among other examples, this requirement is present in the German *Network Enforcement Act* and in the EU's *Proposal for a Regulation on preventing the dissemination of terrorist content online*[99]. The automation of moderation has its own problems, ranging

---

95 Austrian National Council, *Federal Act on measures to protect users on communication platforms* (Communication Platforms Act), 15 Sep 2020. Available at: https://ec.europa.eu/growth/tools-databases/tris/en/search/?trisaction=search.detail&year=2020&num=544

96 ARTICLE 19, *Austria: the draft Communication Platforms' Act fails to protect freedom of expression*, 16 Oct 2020. Available at: http://article19.org/resources/austria-draft-communication-platforms-act-fails-freedom-of-expression/

97 *Directive on Copyright and Related Rights in the Digital Single Market*, Art. 17 § 9.

98 Such implementation is operationally and financially costly, increasing in proportion to the complexity of the moderation decisions that must be made or facilitated. This affects in turn the development of the platforms themselves, as well as the growth of moderation as an outsourced service. See Cambridge Consultants, *Use of AI in Online Content Moderation*, 22 Jul 2019. Available at: https://www.ofcom.org.uk/__data/assets/pdf_file/0028/157249/cambridge-consultants-ai-content-moderation.pdf

99 European Commission, *Proposal for a Regulation of the European Parliament and of the Council on preventing the dissemination of terrorist content online, A contribution from the European Commission to the Leaders' meeting in Salzburg on 19-20 September 2018*, COM/2018/640 final, 12 Sep 2018. Available at: https://eur-lex.europa.eu/resource.html?uri=cellar:dc0b5b0f-b65f-11e8-99ee-01aa75ed71a1.0001.02/DOC_1&format=PDF

from content misinterpretation or lack of context to inherent biases that enhance discrimination against vulnerable groups. That is why, although investment in content moderation technologies and complaint analysis is necessary for large-scale content moderation, the algorithms and activities involved must be carried out appropriately so as to ensure that they do not affect the users' right to freedom of expression.

On the other hand, the fact that Internet platforms became responsible for determining the legality or illegality of a user's expression is a problem in and of itself, since it implies a delegation of government powers to private actors. Giving private companies the power to decide without due transparency, accountability, or effective appeal and/or reparation procedures which expression is legitimate and which one is not violates democratic principles. In the words of David Kaye:

> "Complex questions of fact and law should generally be adjudicated by public institutions, not private actors whose current processes may be inconsistent with due process standards and whose motives are principally economic."[100]

100 Kaye, David, UN Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, *Online content regulation*, Annual Thematic Report 2018 addressed to the UN HRC (A/HRC/38/35), 6 Apr 2018, § 17. Available at: https://undocs.org/A/HRC/38/35

V. Other proposals: self-regulation and coregulation

The regulation of content within private platforms is not only a matter of legislative approach. In addition to the regulatory efforts described above, in the last decade a series of initiatives and recommendations have emerged, developed by different groups and coalitions of diverse integration, with proposals of regulatory, self-regulatory and coregulatory nature. These initiatives are a key tool for modeling the behavior of the private companies that control online platforms[101].

a. Non-binding initiatives in the European Union

In recent years, various non-binding agreements, recommendations and initiatives addressing Internet platforms have been published. Some of these initiatives were jointly developed by private companies with state support, mostly in the European Union.

The first initiatives of this kind were linked to the protection of children and the regulation of the content to which they are exposed online. In 2009, the European Commission published the *Safer Social Networking Principles*, an agreement among major web companies that sought to ensure the implementation of appropriate measures to improve the safety of children and adolescents[102]. Over the years, the issue has become more complex, with governments, companies and society as a whole pushing for action against hate speech, terrorism, fake news, and other worrisome topics.

In May 2016, largely in response to the Brussels bombings of 22 March, some major IT companies, together with the European Commission, announced a *Code of Conduct on illegal online hate speech*[103]. In the agreement, the companies express their commitment to tackle hate speech online, putting in place "clear and effective processes to review notifications regarding illegal hate speech on their services so they can remove or disable access to such content," and reviewing "the majority of valid notifications for removal of illegal hate speech in less than 24 hours." Furthermore, the IT Companies and the European Commission agreed to "assess the public commitments in this code of conduct on a regular basis, including their impact[104]."

---

101 Gorwa, Robert, *The platform governance triangle: conceptualising the informal regulation of online content*, Internet Policy Review, 8(2), DOI: 10.14763/2019.2.1407, 30 Jun 2019. Available at: https://policyreview.info/pdf/policyreview-2019-2-1407.pdf

102 European Commission, *Safer Social Networking Principles for the EU*, developed by SNS providers in consultation with the European Commission, 10 Feb 2009. Available at: https://ec.europa.eu/digital-single-market/sites/digital-agenda/files/sn_principles.pdf

103 European Commission and IT Companies, *Announcement from the European Commission*, Press release IP/16/1937. Available at: https://ec.europa.eu/commission/presscorner/detail/en/IP_16_1937; and *Code Of Conduct On Countering Illegal Hate Speech Online*, originally signed by Facebook, Microsoft, Twitter, YouTube, and later joined by Google, Snapchat, Dailymotion, TikTok, and others, 31 May 2016. Available at: https://ec.europa.eu/info/sites/info/files/code_of_conduct_on_countering_illegal_hate_speech_online_en.pdf

104 For some comments and criticisms around the agreement, see Hughes, Owen, *Twitter, Facebook and YouTube sign EU code of conduct to help combat online hate speech*, International Business Times, 2 Jun 2016. Available at: http://www.ibtimes.co.uk/twitter-facebook-youtube-sign-eu-code-conduct-help-combat-online-hate-speech-1563163

The European Commission, concerned about the persistence of illegal content on the Internet, adopted in March 2018 the *Recommendation on measures to effectively tackle illegal content online*[105]. In this document, the Commission suggests that the mechanisms to report unlawful content "should allow for and encourage the submission of notices which are sufficiently precise and adequately substantiated to enable the hosting provider concerned to take an informed and diligent decision in respect of the content to which the notice relates", adding that those mechanisms should "facilitate the provision of notices that contain an explanation of the reasons why the notice provider considers that content to be illegal content[106]." In cases where the hosting service provider removes or disables access to any content because they consider it to be illegal content, the recommendation states that "the content provider should, without undue delay, be informed in a proportionate manner of that decision and of reasons for taking it, as well as of the possibility to contest that decision." However, this last provision "should not apply where it is manifest that the content concerned is illegal content[107]."

Perhaps the most innovative contribution made by the recommendation is that it includes the provision that content service providers "should be encouraged to take, where appropriate, proportionate and specific proactive measures in respect of illegal content." These measures "could involve the use of automated means for the detection of illegal content only where appropriate and proportionate and subject to effective and appropriate safeguards," so as to avoid removing lawful content[108].

Additionally, it contains specific provisions regarding terrorist content, an issue that is currently being discussed in depth in the trilogue process concerning the *Proposal for a Regulation on preventing the dissemination of terrorist content online*, a controversial proposal since it contains alarming provisions regarding fundamental human rights, in particular freedom of expression and privacy[109]. Among other provisions, it establishes the obligation to implement automated filters and to remove broadly defined terrorist content within one hour of receiving the notice[110].

---

105 European Commission, *Commission Recommendation on measures to effectively tackle illegal content online*, (EU) 2018/334, 1 Mar 2018. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32018H0334&from=EN

106 Ibid., § 6.

107 Ibid., § 9 and 10.

108 Ibid., § 18; see § 19 on Safeguards: "In order to avoid removal of content which is not illegal content, without prejudice to the possibility for hosting service providers to set and enforce their terms of service in accordance with Union law and the laws of the Member States, there should be effective and appropriate safeguards to ensure that hosting service providers act in a diligent and proportionate manner in respect of content that they store, in particular when processing notices and counter-notices and when deciding on the possible removal of or disabling of access to content considered to be illegal content."

109 In European terminology, a *Formal trilogue meeting* ("trilogue" meaning a conversation with three parties) is a EU legislative process involving the European Commission, the Council of the European Union and the European Parliament.

110 It should be noted that, since 2017, there is a Directive that calls for member states to "take the necessary measures to ensure the prompt removal of online content constituting a public provocation to commit a terrorist offense, (...) that is hosted in their territory"; The European Parliament and the Council of the European Union, *Directive on combating terrorism*, (EU) 2017/541, 15 Mar 2017, Art. 21 § 1. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32017L0541&from=EN

In October 2018, representatives of IT companies, social media platforms and advertisers signed the *EU Code of Practice on Disinformation*, published by the European Commission[111]. Just like the *Code Of Conduct On Countering Illegal Hate Speech Online*, the Code of Practice is a voluntary agreement of a self-regulatory nature, this time aimed at combating disinformation and fake news. The document, much more extensive and detailed than the previous one, includes a list of best practices and added in 2020 a *COVID-19 Monitoring Programme*, describing their actions to counter disinformation during the pandemic[112].

All these initiatives have certain limitations, such as the fact that their scope is limited to the specific framework of EU laws and the political supervision of the European Commission, or that they do not establish clear regulatory consequences for those who fail to comply with their provisions. Likewise, even when there is a variable level of detail in the required commitments, compliance seems to necessarily demand a certain operational capacity that only large platforms are capable of assuming, regarding both the development and the observance of policies that seek to protect the interests of the European community.

Transferring the responsibility for content control tools to Internet platforms (and sometimes advertisers) seems to be a common trait of these initiatives, without the counterbalance of fixed legal obligations to deal with the content that is allegedly harmful or infringes some highly abstract European standards. Despite tackling at the institutional level a range of valid concerns on topics like the dangers of social media regarding the rights of children and adolescents, or the spread of misinformation in its different forms, or the online presence of groups that call for violence or disseminate discriminatory hate speech, this appears to be a sort of politically supported self-regulation model that precludes the formulation of rigid legal norms in an entire region[113].

On the one hand, within the European Union these agreements are exported as European regulations are imposed on each Member State, disregarding the existing local regulations on freedom of expression that vary from one country to another; on the other hand, the adoption of these standards within the terms of service of Internet companies often implies the imposition of the European criterion to all other jurisdictions.

---

111 European Commission, *EU Code of Practice on Disinformation*, signed by Facebook, Google, Twitter, Mozilla, Microsoft, TikTok, and others, Oct 2018. Available at: https://ec.europa.eu/digital-single-market/en/code-practice-disinformation

112 See European Commission, *Disinformation: EU assesses the Code of Practice and publishes platform reports on coronavirus related disinformation*, Press release, 10 Sep 2020. Available at: https://ec.europa.eu/commission/presscorner/detail/en/ip_20_1568; and *Fifth set of reports – Fighting COVID-19 disinformation Monitoring Programme*, Report/Study, 18 Jan 2021. Available at: https://ec.europa.eu/digital-single-market/en/news/fifth-set-reports-fighting-covid-19-disinformation-monitoring-programme

113 Keller, Daphne, *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, 29 Jan 2019. Available at: http://lawfareblog.com/who-do-you-sue-state-and-platform-hybrid-power-over-online-speech

b. GIFCT

Since 2016, the phenomenon of voluntary partnerships between governments and IT companies has been expanding rapidly, and one recent example is *The Global Internet Forum to Counter Terrorism* (GIFCT), created in 2017 to foster technical collaboration among member companies and promote cross-industry efforts to counter the spread of terrorist and violent extremist content online[114]. With the mission of preventing terrorists and violent extremists from exploiting digital platforms, the GIFCT seeks to consolidate itself as the entity that centralizes the moderation of this type of content for its member companies. To this end, they have created a shared database of hashes (unique digital "fingerprints") of known violent terrorist imagery and videos, called *Hash-Sharing Consortium*[115]. Any copy of an image or video would have the same hash as the original file, allowing companies to quickly scan their platforms for a matching file and then take appropriate steps to limit or prevent its circulation.

Notwithstanding its good intentions (the consensus around the need to control and block terrorist content is indisputable), experts and civil society organizations have expressed some serious criticism of this initiative[116]. Overall, it is criticized for its lack of legitimacy, lack of supervision mechanisms and lack of transparency. Outside the GIFCT member companies, there is no knowledge about what content is included in the shared database, there are no external audits and there is no appeal mechanism to protest the database content. This lack of transparency implies that certain forms of expression are being altogether excluded with a total lack of contextualization or the likelihood of complaining about the decision made.

Furthermore, platforms are basing content removal decisions on unpublished private agreements, instead of directives or other binding standards, making it impossible for users to challenge their interpretations.

---

114 The *Global Internet Forum to Counter Terrorism* (GIFCT) was founded by Facebook, Microsoft, Twitter and YouTube in 2017, later joined by Discord, Instagram, WhatsApp, Amazon, Mega, LinkedIn, and others. Available at: https://gifct.org/

115 As of January 2021, the *Hash-Sharing Consortium* consists of 13 companies: Microsoft, Facebook, Twitter, YouTube, Ask.fm, Cloudinary, Instagram, JustPaste.it, LinkedIn, Verizon Media, Reddit, Snap, and Yellow. Available at: https://gifct.org/joint-tech-innovation/#row-hash

116 A significant criticism lies in the nature of the issue being discussed: the technology of "hashes" was originally implemented to combat child sexual abuse content online but, unlike these contents, determining the illegality of terrorist or extreme violence material can be much more complex, ambiguous or unclear. Because of this, there is the risk of over-including in these databases content that is legal and protected against any censorship. Furthermore, there is concern that some governments may abuse the databases, using them as a means to silence opposition voices by characterizing them as "terrorists". It should be added that the Christchurch Call seems to want to expand, in practice, the use of this shared database, thus extending the application of a highly problematic content control system. The combination of these elements could not only prevent the production and dissemination of certain legitimate content, but also lead to silencing reports on actual crimes and human rights violations, suppressing truth and investigation efforts. For more on this, see Douek, Evelyn, *The Rise of Content Cartels Urging transparency and accountability in industry-wide content removal decisions*, Knight First Amendment Institute at Columbia University, 11 Feb 2020. Available at: https://knightcolumbia.org/content/the-rise-of-content-cartels; Citron, Danielle Keats, *Extremist Speech, Compelled Conformity, and Censorship Creep*, 93 Notre Dame Law Review 1035, U of Maryland Legal Studies Research Paper No. 2017-12, 2018. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2941880; Llansó, Emma, *Platforms Want Centralized Censorship. That Should Scare You*, WIRED, 18 Apr 2019; and Radsch, Courtney, op. cit. Available at: http://wired.com/story/platforms-centralized-censorship/

c. The Christchurch Call

Within this same category of public–private partnerships, we find *The Christchurch Call*, a non-binding agreement emerging from a political summit that took place on 15 May 2019 in Paris, initiated by New Zealand Prime Minister Jacinda Ardern, and co-chaired by Ardern and French President Emmanuel Macron[117]. The resulting *Call to Action To Eliminate Terrorist and Violent Extremist Content Online* was initially signed by 17 countries and 8 large IT companies, with 31 countries joining on 24 September, including Argentina, Chile, Colombia, Costa Rica and Mexico. The document, establishing non-binding commitments for Governments and OSPs, came out as a hasty response to the Christchurch mosque shootings, a terrorist attack against the Muslim community of Christchurch, New Zealand, which resulted in the death of 51 people and was live-streamed on Facebook two months prior to the summit.

The Christchurch Call incorporates commitments for governments in broad and ambiguous terms, including considering the possibility of new regulatory measures to prevent the dissemination of terrorist and violent content. Regarding OSPs, the call encourages the adoption of "transparent, specific measures seeking to prevent the upload of terrorist and violent extremist content and to prevent its dissemination on social media and similar content-sharing services." These measures may include "technology development, the expansion and use of shared databases of hashes and URLs, and effective notice-and-takedown procedures." Other provisions that OSPs commit to, include implementing "immediate, effective measures to mitigate the specific risk that terrorist and violent extremist content is disseminated through livestreaming" and reviewing the operation of algorithms that may amplify this kind of content. The agreement also calls for the implementation of regular and transparent public reporting, the enforcement of their terms of service "in a manner consistent with human rights and fundamental freedoms", and the sharing of knowledge and expertise "to ensure cross-industry efforts are coordinated and robust."

This entire range of commitments is written in high-level language, with little precision and with equally vague commitments to respect international human rights norms and standards[118]. Suitably, the elaboration of these commitments with a possible negative impact on online expression led to a strong reaction from a significant amount of civil society organizations concerned with digital rights. These organizations denounced the lack of transparency and the exclusion of relevant voices, the introduction of vague definitions, the flaws in the proposed commitments, and the diffuse attribution of roles among States and private companies regarding the control of potentially protected expressions[119]. The reaction to these criticisms was the development of an "Advisory Network", made up of an overwhelming majority of institutions from the global north[120]. Consequently, the Christchurch Call has become the basis for the

---

117 Christchurch Call to Action Summit, *The Christchurch Call to Action To Eliminate Terrorist and Violent Extremist Content Online*, 15 May 2019. The agreement has been signed by 48 countries, the European Commission, UNESCO, the Council of Europe and major OSPs, including GIFCT members. Available at: http://christchurchcall.com/christchurch-call.pdf

118 Douek, Evelyn, *Two Calls for Tech Regulation: The French Government Report and the Christchurch Call*, Lawfare, 18 May 2019. Available at: https://www.lawfareblog.com/two-calls-tech-regulation-french-government-report-and-christchurch-call

119 *Civil Society Positions on Christchurch Call Pledge*, document prepared for the Civil Society leaders' Voices for Action meeting with New Zealand Prime Minister Jacinda Ardern, 14 May 2019. Available at: https://www.eff.org/files/2019/05/16/community_input_on_christchurch_call.pdf

120 The *Christchurch Call Advisory Network* list of members, available at: https://www.christchurchcall.com/advisory-network.html

adoption of new measures by Internet companies, including changes in the technologies for the detection and blocking of certain content on their platforms, or the acceleration of automated filtering processes.

d. Initiatives from Tech Companies

The initiatives proposed by companies to deal with content moderation do not strictly address the issue of intermediary liability. However, as Dawn Nunziato highlighted already in 2012 in her introduction to *Towards an Internet Free of Censorship*, the self-regulatory activity of companies could only be justified to the extent that criteria of transparency and due process are established, to guarantee the company's fairness and "neutrality" regarding not its rules but how those rules are applied[121]. In many cases, these measures and initiatives were a response to growing pressure from users and governments demanding greater legitimacy in the design and implementation of their terms and conditions of service. Essentially, they emerged from the companies themselves in the face of specialized criticism and the growing difficulty posed by some issues around content moderation[122].

In a different group of initiatives, in this case stemming directly from the companies, it is worth mentioning the Twitter *Trust and Safety Council*, a group of independent organizations advising the company in different areas of focus related to its policies and practices[123]. The Council is not a tool to establish specific content policies, but rather a mechanism of consultation of a company about itself[124].

For its part, the *Oversight Board*, first proposed by Facebook in September 2018, intends to serve as a body that makes content moderation decisions for the platform, receiving and reviewing appeals from users of Facebook and Instagram to "determine if decisions were made in accordance with Facebook's stated values and policies," as well as deciding "what to take down,

---

121 "In the absence of regulation, Internet service providers will have discretionary powers to discriminate content or applications, and citizens will not have guarantees of access to a multiplicity of opinions from diverse and antagonistic sources, exempt from censorship, which is necessary for them to have a relevant participation in a democratic government." Nunziato, Dawn C, *Preserving Internet Freedom in the Americas*, in *Towards an Internet Free of Censorship: Proposals for Latin America*, compiled by Eduardo Andrés Bertoni, CELE, UP, 2012. Available at: https://www.palermo.edu/cele/pdf/internet_libre_de_censura_libro.pdf [in Spanish]

122 In November 2018, Facebook CEO Mark Zuckerberg announced, in his second note about the most important issues facing the company, his intention to create an independent review body, which would later become Facebook's Oversight Board: "I believe the world is better when more people have a voice to share their experiences, and when traditional gatekeepers like governments and media companies don't control what ideas can be expressed. At the same time, we have a responsibility to keep people safe on our services -- whether from terrorism, bullying, or other threats. We also have a broader social responsibility to help bring people closer together -- against polarization and extremism. The past two years have shown that without sufficient safeguards, people will misuse these tools to interfere in elections, spread misinformation, and incite violence. One of the most painful lessons I've learned is that when you connect two billion people, you will see all the beauty and ugliness of humanity." Mark Zuckerberg, *A Blueprint for Content Governance and Enforcement*, 15 Nov 2018. Available at: https://www.facebook.com/notes/751449002072082/

123 Twitter, *Trust and Safety Council*, announced 9 Feb 2016. Al Sur is a member of the *Online Safety and Harassment Prevention* group, which "engages Twitter on priority challenges and policy issues including but not limited to online harassment and abuse." Available at: https://about.twitter.com/en/our-priorities/healthy-conversations/trust-and-safety-council

124 "The Twitter Trust and Safety Council is a group of independent expert organizations from around the world. Together, they advocate for safety and advise us as we develop our products, programs, and rules. (...) Membership is voluntary and doesn't imply endorsement of any decisions we make. Members also don't speak on Twitter's behalf."

what to leave up, and why[125]." According to Facebook, the Board will be committed to support people's right to freedom of expression and its decisions will be binding. Given the high profile of the company behind this project, it did not take long for criticism to arrive, mainly related to the scope of its work, since the Council does not have the power to establish Facebook's policies or its procedures for standards compliance regarding content moderation[126].

The two cases described above are examples of companies broadening the opportunities for organizations and people from outside the company to influence in different ways the decisions they make regarding the content of their users. These instances involve the inclusion of diverse voices in the self-regulation process of companies, with different levels of incidence depending on the structure and design. The various efforts have been welcomed broadly by the user community and with some skepticism by civil society and academia. The main criticism points to highlighting that it is nothing more than an attempt to detach part of the responsibility from the company in particularly complex cases. The number of individual cases reviewed by these bodies is minimal compared to the number of interventions that companies make on a daily basis on the content hosted on their platforms, and their impact in the design of content policies in self-regulation is doubtful.

e. Recommendations from Civil Society

Those efforts in which civil society, in a broad sense, has a more central role, should be studied separately, including the contributions of independent experts, academic entities and members of technical communities not affiliated with industries or governments. Within this group of proposals, we find the *Manila Principles on Intermediary Liability*, elaborated by a coalition of civil society organizations, and where standards are set for the adoption of policies and practices on intermediary liability that are respectful of human rights[127]. The principles include, among others, duties of transparency and due process in the control of content that may violate laws or terms of use, as well as the possibility of incurring legal liability under certain requirements. Although it is a set of widely recognized principles, its formulation still restricts its use to contexts of reduced content control, not large-scale moderation, in addition to being hard to assimilate it to the expectations that different governments have expressed on the proactive control of content on platforms.

Also within this group, we find the work of the international NGO ARTICLE 19, a central contributor in the process of elaboration of the Manila Principles. Shortly before those principles were formulated, ARTICLE 19 published their policy document *Internet intermediaries: Dilemma of Liability*, a first set of recommendations for governments and companies on intermediary liability[128]. In it, they advocate that: there should be no liability for intermediaries who store third-party content (broad immunity model); a notice model should be established before per-

125 Facebook's *Oversight Board* began operations on 22 Oct 2020. More information available at: https://oversightboard.com/

126 See, for example, Al Sur, *Facebook Oversight Board: A Perspective from Latin America and the Caribbean*, 7 Jun 2019. Available at: https://www.alsur.lat/sites/default/files/2020-04/COMENTARIOS GENERALES AL DRAFT DE OVERSIGHT BOARD FACEBOOK_ALSUR.pdf [in Spanish]

127 *Manila Principles on Intermediary Liability, Best Practices Guidelines for Limiting Intermediary Liability for Content to Promote Freedom of Expression and Innovation*, Global Civil Society Initiative, 24 Mar 2015. Available at: https://www.manilaprinciples.org/

128 ARTICLE 19, *Internet intermediaries: Dilemma of Liability*, 2013. Available at: https://www.article19.org/data/files/Intermediaries_ENGLISH.pdf

forming any restriction or withdrawal action (notice-to-notice procedures); particular rules should be established in cases of alleged serious criminality (model for specific cases). These recommendations were echoed and further developed in the Manila Principles, keeping international human rights law as the basis for the recommendations.

Several years later, in 2018, ARTICLE 19 developed a new set of recommendations, this time aimed directly at companies, assuming the practical reality of self-regulation and focusing on the problem of hate speech on social media[129]. The document raises numerous criticisms of the self-regulatory and coregulatory models promoted within the European Union, including the aforementioned *Code of Conduct* and *Code of Practice*, due to their apparent incompatibility with international human rights law. However, certain operational recommendations are established for self-regulatory attempts by companies, in addition to re-examining the role of governments in promoting and protecting the right to freedom of expression of Internet users, including through regulatory channels.

A later proposal by ARTICLE 19 went further, suggesting the creation of a body outside the companies that would be in charge of reviewing decisions, with an even more stressed independence than the Oversight Board created by Facebook. In 2019, ARTICLE 19 promoted the *Social Media Councils* (SMCs) with the aim of providing an accountability mechanism for online content moderation that would be independent, multisectoral, open, transparent, and based on international human rights standards[130].

In 2018, a group of academics and representatives of American civil society promoted the *Santa Clara Principles On Transparency and Accountability in Content Moderation*, a set of three principles to better ensure that the platform's enforcement of their content guidelines is "fair, unbiased, proportional, and respectful of users' rights[131]." Their recommendations, specifically directed at companies, focus on three aspects of content moderation: notice to users, appeal of decisions and transparency. The scope of these principles is limited, since it only covers certain aspirations regarding the operation of companies and not their regulation or their liability for failing to meet the standards. Moreover, although the recommendations are procedural rather than substantive regarding the contents of moderation rules, they still leave out some systemic aspects, such as the contextual categorization of content types[132]. The document also received multiple criticisms for having been originally promoted by a small group of academics and civil society in the United States and later circulated for signature and promotion in other parts of the globe. On the positive side, a long period of consultation carried out during 2020 opened the possibility of improvements in these recommendations, despite the fact that they will probably maintain their limited scope.

---

129 ARTICLE 19, *Self-regulation and 'hate speech' on social media platforms*, 2018. Available at: https://www.article19.org/wp-content/uploads/2018/03/Self-regulation-and-'hate-speech'-on-social-media-platforms_March2018.pdf

130 More information at: https://www.article19.org/resources/social-media-councils

131 EFF, OTI, ACLU Foundation and others, *The Santa Clara Principles: On Transparency and Accountability in Content Moderation*, 7 May 2018. Available at: https://santaclaraprinciples.org/

132 The contextual categorization of content types in operational terms plays a central role in, for example. addressing the enormous amounts of spam that are detected and blocked on a daily basis on large platforms.

At the regional level, in Latin America a group of civil society organizations recently promoted a document with standards and recommendations entitled *Standards for a democratic regulation of large platforms that guarantees freedom of expression online and a free and open Internet*, developed by various organizations under the coordination of OBSERVACOM[133]. The initiative includes several of the discussions on regulation already outlined, analyzed here from a Latin American approach. In turn, it seeks to expressly address the dominant nature of certain Internet platforms, with an emphasis on the concentration of power in private hands and considering the effect that this power has on the ability of users to effectively exercise their freedom of expression.

The document also aims to introduce a form of regulation that is different from the apparent extremes represented by self-regulation and "abusive state regulation". The proposal follows other initiatives in rejecting the strict liability of platforms and the delegation to them of the decisions on the legality, veracity or decorum of the expressions of their users. In this regard, the promotion of a coregulation system stands out, where the standards to be met must be different for "large content platforms", designated by a national regulator according to the characteristics described in the document. It also considers recommendations for community standards, due process and moderation transparency, among others with a variable level of specificity.

Regarding coregulation, it advocates the existence of a specialized state regulatory body to supervise the implementation of the principles that the document itself wields. At the same time, it focuses on social networks and, specifically, on the issue of freedom of expression as a distinct and separate issue from others potentially subject to different regulation, such as economic, public finance, competition, consumer rights issues, etc.

This initiative is not exempt from criticism either, insofar as it subjects corporate self-regulation to state control. On the other hand, although there is consensus on many of the principles, the document lacks the specificity and granularity necessary to address issues such as the inter-jurisdictional nature of the internet and the platforms that operate on the internet, the substantive differences in the definition and extension of certain criteria and human rights standards on which there is consensus but there are also significant margins of discretion for their interpretation and implementation by the States. For their part, those who promote the document advocate that this is the starting point of the conversation and not necessarily the point of arrival meant to be literally translated into a bill.

133 OBSERVACOM et al., *Standards for a democratic regulation of large platforms that guarantees freedom of expression online and a free and open Internet*, A Latin American perspective to achieve private content moderation processes that are compatible with international human rights law, Jul 2020. Available at: https://www.observacom.org/wp-content/uploads/2020/09/Estandares-para-una-regulacion-democratica-de-las-grandes-plataformas.pdf [in Spanish]

VI. Conclusions

Undoubtedly, the complexity of the subject emerges from the document and the initiatives described. The regime of liability of intermediaries for the content of third parties has historically had a direct impact on the moderation of content. From the first standards at a comparative level in the world to the few existing standards in this regard in our region.

The regional and comparative legislation, and the trends that we see in the existing projects to regulate the liability of intermediaries show us, in general terms, an appetite for restricting expression on the Internet, particularly on social media networks. In Latin America, the little State regulation on intermediary liability that exists establishes the differentiated responsibility of companies for third-party content when this content may allegedly violate copyright laws. Only recently, and in an absolutely reactive manner, have initiatives to regulate intermediary liability for the circulation of content on other issues (such as disinformation or hate speech) emerged in the region's congresses. These issues are undoubtedly concerning and have also emerged with more force on the global public agenda in recent years. The initiatives, however, in most cases, lack a prior period of study or gathering of evidence, or a discussion accompanied by a broad, participatory and plural debate; and its normative proposals, for the most part, also contravene regional and international human rights standards. We see in Latin America a "reactive" trend over a "proactive" approach to address the problems of the moment, with a deficient participation of civil society in legislative debates.

In all cases, these laws have a direct impact on the regime of intermediary liability, which until now lacks specific regulations and continues to be under the orbit of ordinary civil liability. This lack of clarification on the general framework of intermediary liability perpetuates discussions around the applicability of strict liability in the region, which has not only been discarded by the rest of the world, but has also been clearly indicated as incompatible with the regional and global standards of freedom of expression.

On a comparative level, the regulatory scenario is not particularly clear either. The United States and the European Union have framework laws that guarantee the non-liability of intermediaries for third-party content, having recently emerged some mere improvements to transparency, due process and accountability of the existing regime, in the form of limited exceptions to the general rule. This document tries to some extent to highlight the assumptions of this type of regulatory framework and its temporal development, in an attempt to nourish the discussions that are still going on in Latin America about the general principles.

The spirit of exemption from liability regarding third-party content appears to be intact from what emerges from the proposals to amend CDA 230 and the proposals embodied by the DSA and the DMA. However, inaccuracies abound in these regulatory efforts, particularly regarding the content that seems most problematic, such as disinformation, hate speech, terrorism, and radicalized speech. In addition to these, there are the traditional exceptions granted to the general frameworks of exemption from liability, reflected in the intellectual property laws to which journalistic content has also recently been added.

Technological development has undoubtedly also radically impacted the discussion around intermediary liability, by intertwining it with the consequences of content moderation for the exercise of users' rights. The internal content monitoring capabilities of platforms are currently much greater than they were a decade ago. The development of increasingly complex and obscure algorithms for sorting content and disseminating it to the community is also radically different and poses new challenges. The growth and the centralization of power among just a few actors at a global level have generated the need to look at and think about intermediaries not only because of their functions or the layers in which they function (as we did in the past) but also because of their size, social function, public impact at the individual and collective level. Hence, the initiatives that have emerged in recent years, both from States and civil society, are aimed precisely at designing more complex transparency and accountability mechanisms, different from those we thought about a decade ago.

On the other hand, these technological developments from recent years force us to think about the responsibility of intermediaries for their own actions, in addition to the responsibility they could have for the content of third parties. The prioritization of content, the development of algorithms, the fulfillment of adequate criteria to ensure non-discrimination and due process, are part of the actions that are the object of this examination of their responsibility, beyond the effect on people that would arise from the harmful content that is withdrawn from circulation. The increasing complexity of those actions can make it difficult to regulate the responsibility for their own actions, which makes the involvement of different interested parties and expert voices necessary for any regulatory debate. It implies, in turn, a regulation that manages to establish international standards and principles without being associated with specific platforms or technologies, but that can be adapted to other unknown forms of intervention by the platforms.

Notwithstanding the fact that the global legislative outlook does not seem particularly encouraging, the Latin American region is in a unique position to learn from comparative experiences, both in Europe and in the United States, where, as we have seen, good and bad practices abound. In recent years, the regulatory processes and their respective debates have demonstrated the need to generate broad discussions and serious legislative processes that incorporate impact studies of the proposed legislation on the local and/or regional ecosystems. These processes enjoy further debate and study, which allows foreseeing difficulties in implementation, reactions and errors. In Europe in particular there is, to date, an interesting accumulation of experiences regarding the implementation of the agreements and regulations established in recent years. This experience could be particularly rich to avoid common mistakes. It is the responsibility of the States, civil society and academia to promote appropriate dialogue instances, designed to identify good and bad practices, with the aim of achieving mutual benefit.

After a broad and detailed reflection process, we at AI Sur extract from comparative legislation, the practice of companies and the experiences of the 11 organizations that are part of this consortium, the following guidelines for any regulatory debate regarding intermediary liability:

- Any regulation within our regional framework must comply with Interamerican standards regarding freedom of expression, privacy, access to information, rights of association and assembly, political rights, and the right to the truth, among others;

- Whether in a framework of regulation, self-regulation or coregulation, the obligations of equality and non-discrimination both towards people and content must be promoted, together with the principles established in OC 5/85, which establishes that it will be incompatible with the Convention any regulation or private self-regulation that a priori excludes groups or voices from the debate;

- There is an urgent need to discuss a general framework regarding intermediary liability for third-party content, and clearly establish as a general principle the exemption from liability of intermediaries, except when there is a reliable judicial notification;

- Respecting the distinctions that exist between our region and Europe, establishing a legal obligation for a third party to eliminate or block harmful but not illegal information or ideas without judicial intervention, constitutes prior censorship and, therefore, not only should it not be encouraged, but it must be combated;

- If the legislation proposes obligations for intermediaries with respect to certain contents, they must be clearly and exhaustively established in the law, and subject to sufficient supervision and transparency mechanisms;

- A clear distinction should be promoted between the liability of intermediaries for their own acts, and their liability for the content or acts of third parties. The definition of "intervention" is problematic and complex, particularly as technology evolves and the role of certain intermediaries changes. This distinction for legal purposes should be subject to broad debate, incorporating the voices of academics, civil society and attending to legislate based on evidence. The adoption and promotion of laws/regulations specific to the technology or behavior of a certain platform carries the inevitable risk of becoming obsolete;

- Flexible rules should be promoted, capable of establishing a clear distinction between the obligations that correspond to different categories of intermediaries, that fulfill different characteristics in their role of intermediation within the Internet, taking into consideration the different sizes and operating business models. Debates around regulation, self-regulation and coregulation must attend to the necessary distinctions between intermediaries due to their functions, their size and their possible impact on the public debate, according to the proposed regulation (for example, if the regulation affects transparency or due process obligations, or removal and blocking periods for certain content). Promote a differentiated criterion for intermediaries according to criteria of necessity and proportionality, taking into account the impact on the intermediary's structure, the costs and its influence on the subject. The NetzDG is, in this regard, an example to follow insofar as it distinguishes intermediaries to whom specific terms and deadlines apply;

- Mechanisms for judicial review should be promoted in the adjudication of rights, both online and offline. Legal processes must be easily accessible and quick to resolve on issues related to intermediary liability, in order to guarantee the fundamental rights of users;

- Human rights criteria should be promoted around the development of algorithms, artificial intelligence (AI) and their regulation, as opposed to ethical criteria. Understanding the term AI in a broad sense, according to the definition used by David Kaye, former UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression in his report on AI and Human Rights (2019);

- Understand the regulation of intermediary liability and content moderation as a regulatory field that interacts with others and, therefore, systemically evaluate the opportunities where some of the issues presented as current shortcomings of the regulatory regime of intermediaries may be resolved through other specific regulations, such as regulation of economic concentration and commercial competition, consumer rights, protection of personal data or electoral regulation;

- Regarding the regional strategy to deepen and accelerate the discussion around the regulation of intermediaries, as a first measure, it is necessary to prioritize and advance on the areas in which today there is the strongest consensus, where we trust that we can obtain best results: the adoption of the general principle of immunity over the content of third parties, regulations regarding transparency, due process, accountability, reparation and compatibility with human rights. The process leaded by the IACHR currently underway should be aimed at providing content from a regional perspective to each of these aspects;

- It is essential to promote the participation of the American States in the different multisectoral governance spaces on various aspects of the digital economy that proliferate globally. It is clear from the information contained in the document that public-private initiatives, especially in matters of terrorism, radicalization, rights of children and adolescents, and even disinformation, have a direct impact on the practices of companies with global impact, and our region is underrepresented in these spaces, both at the civil society level and at the State level.

www.alsur.lat

AlSur