A COMPARATIVE LAW APPROACH TO THE REGULATION OF SOCIAL NETWORKING PLATFORMS

ABSTRACT

Social networking platforms have brought tremendous benefits in multiple aspects. However, they also open the door to new threats to democracies and fundamental rights, such as cybercrimes, disinformation, content censorship or mass surveillance. In the following pages we will analyse some of the most relevant regulatory approaches adopted worldwide, in order to cope with this paradigm shift. Finally, we will briefly mention the main characteristics of the Proposal for the Digital Services Act: a new Regulation which, in the following years, will presumably transform the regulation guidelines of social platforms in the European Union and in the rest of the world.

Keywords: platforms, social networks, democracy, free speech, censorship, China, United States, European Union, information society services, Digital Services Act.
RESUMEN

Las plataformas de redes sociales han aportado enormes beneficios en múltiples aspectos. Sin embargo, al mismo tiempo, también abren la puerta a nuevas amenazas para las democracias y los derechos fundamentales, tales como como los ciberdelitos, la desinformación, la censura de contenidos o la vigilancia masiva. En las siguientes páginas analizaremos algunos de los enfoques regulatorios más significativos adoptados a lo largo del planeta para hacer frente a este cambio de paradigma. Finalmente, mencionaremos brevemente las principales características de la Propuesta de Ley de Servicios Digitales: un nuevo Reglamento que, en los próximos años, presumiblemente transformará las pautas de regulación de las redes sociales en la Unión Europea, y en el resto del mundo.

Palabras clave: plataformas, redes sociales, democracia, libertad de expresión, censura, China, Estados Unidos, Unión Europea, Servicios de la Sociedad de la Información, Ley de Servicios Digitales.

I. INTRODUCTION

In November 3rd 2003, a young student at Harvard was notified that he would be investigated by the university’s Administrative Board for inappropriately using online posted pictures of students in order to create the web page facesmash.com, where students could rank each other’s pictures, similar to the (at that time) popular site “Am I Hot or Not?”¹. The following day, the page was removed. One year later, from the ashes of that project, young Mark Zuckerberg launched thefacebook.com. While access was initially restricted to Harvard students, its popularity expanded, and in 2009, after overtaking MySpace, it became the most popular social network worldwide². In April 10th 2018, the Congress of the United States required Zuckerberg to appear before them to explain how Cambridge Analytica, a political consulting firm could have harvested data from approximately 87 million Facebook users for 2016 elections. This was the first serious attempt by public powers to demand accountability to social networking platforms while acknowledging their unprecedented potential to harm open societies as we know them: A shift in the paradigm, through which the internet would be no longer a land without law, alien to the intervention of the State. Now the question is not if states have to regulate the cyber-dominion, but how, and to what extent, in order to preserve demo-liberals societies fundamental rights and freedoms.

1. Challenges posed by social networking platforms

Whereas it is true that all economic disruptions imply some sort of social impact, it is not easy to think about any historical example of elements transforming the society in such a profound way and in such a short period of time like social networking platforms have done. In this sense, growth of social networking platforms has deep implications on how we work, how we learn, how we develop our relationship. In fact, the net has changed the way we understand the world and the way we understand ourselves.

² Edosomwan, Seymour, Watson, Kouame, & Prakasan, 2011, p. 4.
However, while these changes open the door to new possibilities of progress, knowledge, involvement and development, they also generate new threatens to be faced by free societies, such as cybercrimes, disinformation, and digital suppression of fundamental rights. All of them are deeply related with rule of law and protection of fundamental rights and freedoms, and all of them implies a brand new approach in the regulatory activity of states in order to cope with this new paradigm.

Firstly, cybercrimes, or crimes committed through the internet, are those offences in which communication systems are used in the commission or facilitation of criminal activity. Cybercrimes are challenging mainly because of its transnational nature, which escapes classic structure of criminal procedural law, built over the concept of sovereignty as the supreme authority within a territory. One of the most dangerous activities in that sense are the attacks on critical infrastructures, as Colonial and SolarWinds cyberattacks made clear. In that case, cybercrime is clearly related with geopolitical conflicts and states sponsored activities. This situation needs to be addressed by international agreements on limits to those offensive activities. When the importance of boundaries falls down in favour of a ubiquitous internet, we require international instruments to enhance coordination among states, such as the Budapest Convention on Cybercrime (ETS No. 185). Messages constituting criminal offences such as copyright infringements or fraud can dramatically increase their chances of succeeding (and their negative effects) thanks to social networking platforms, as they open the door to a huge new potential audience.

Secondly, we shall discuss disinformation. Disinformation can be defined as “false, inaccurate, or misleading information designed, presented and promoted to intentionally cause public harm”\(^3\). Disinformation is different to misinformation: while both imply spreading false information, the latter is the result of an error, while the former is performed deliberately. On the last few years, a tendency of social networking platforms being used as a mean to be informed has grown globally. Pew Research Center collects how in the United States in 2016, 62% of population used social media as a source of information.

\(^3\) European Commission, 2018, p. 11.
Two years later, the proportion increased to 68%\textsuperscript{4}. In 2020, 72% of respondents admitted to get news from social media, at least rarely\textsuperscript{5}. According to Reuters Institute, in Spain social media evolved from representing 28% of total news sources in 2013 to 53% in 2021\textsuperscript{6}.

In this context, social networking platforms can be a perfect window for stakeholders willing to exercise their influence over public opinion. Multiple experts have provided evidence about the spread in the use of disinformation campaigns during the last years, especially in the context of electoral processes\textsuperscript{7}. A more recent approach to the relevance of the question is the misinformation spread on COVID 19 vaccines. In this regard, countries like Russia or China usually have engaged in the so-called “hybrid wars” or “non-linear wars”, which refer to the use of non-military means, such as campaigns based in strategically spreading disinformation, to achieve its goals. On the national level, populist movements have also benefited from strategic dissemination of disinformation messages through social networking platforms, with outstanding examples such as the roles displayed by Facebook or Whatsapp during the Brexit referendum\textsuperscript{8} or Brazilian 2018 elections, respectively. These campaigns normally use technologies that increase their credibility, such as bots, which artificially enhance the spread of a message in social networking platforms, or deep fakes, which allow the creation of ultra-realistic fake videos without the necessity of possessing high technical skills.

So far, we have mentioned cybercrimes and disinformation as two of the main problems originating in social networking platforms. A third one, directly derived from them, is the level of discretion that is granted to the social network itself to moderate problematic content. Whereas in practice all legal systems will allow (and sometimes force) social networking platforms to eliminate illegal content, they may use their discretion to moderate content in an arbitrary way. In this

\textsuperscript{4} Shearer & Matsa, 2018.
\textsuperscript{5} In its 2020 report, Pew Research Center changed the formulation of its survey’s questions and, thus, its results are not directly comparable to previous years. However, we still consider that this piece of data could be interesting for the reader (Shearer & Mitchell, 2021).
\textsuperscript{6} Reuters Institute, 2021, p. 103.
\textsuperscript{7} Oxford Internet Institute, 2019, p. 6.
\textsuperscript{8} United Kingdom Parliament, 2019.
sense, social networking platforms like Twitter have been accused of moderating and managing content with an ideological bias.\(^9\)

Which is the role of social platforms within this panorama? While it seems evident that they make the primary decisions about their moderation policies, they also may play an active role on the way disinformation is distributed. It could be thought that social networking platforms are passive intermediaries, acting merely like “public squares” through which users communicate with each other. However, the vast amount of information that they host and the need to filter it provokes that decisions on how information is filtered and ranked utterly condition which information is visible at all. Reputed voices from different fields\(^10\) start demanding responsibility to social networking platforms for the way they perform content curation, partially blaming them for favouring the spread of populism and radicalization of public discourse. In this sense, we are seeing the first cases of online platforms facing liability for the way they filter content.\(^11\)

Through the following lines we will study the present and discuss on the future of social networking platforms’ responsibility. We will focus on social networking platforms “secondary liability” (i.e. liability arising vicariously from the illegal behaviour of users). We will begin by examining some of the main regulatory approaches that have been proposed globally to solve this problem: starting from the American approach, set up over section 230 of the Communication Decency Act, we will review the interventionist Chinese model, based on the principle of cyber-sovereignty. After discussing the European system, we will briefly mention main characteristics of the Proposal for the Digital Services Act: a new Regulation which, in the following years, will presumably transform the paradigm about regulation of social platforms in Europe, and worldwide.

---

\(^9\) See, for example McMillan & Horwitz, 2020.
\(^11\) European Commission decision against Google Shopping (Case AT.39740 – Google Search (Shopping)), in the context of competition law.
II. THE UNITED STATES AND SECTION 230 OF THE CDA

The First Amendment of the United States Constitution protects freedom of speech and press and limits the ways through which Congress can restrict this freedom. Although not all types of speech receive the same protection, in general, restrictions on free speech have to satisfy a highly demanding standard\(^\text{12}\). In *Reno v. American Civil Liberties Union*, United States Supreme Court recognized that this constitutional protection extends also to communications through the cyberspace. If we wonder about the specific regulation on social networking platforms’ liability, it is found in section 230 of the Communications Decency Act (CDA) (*47 U.S.C. §230*), a federal law passed in 1996.

In order to fully understand section 230, we need to consider the historical context where it was enacted. In 1995, a New York court decision threatened the future growth of primitive communication services providers: in *Stratton Oakmont, Inc. v. Prodigy Servs. Co*, a trial court opened the door to the possibility that, when engaging in moderation of content posted through them, platforms could be considered “publishers” of information posted by users. In common law, publishers (or speakers) of illegal content may be found “strictly liable” for harms caused through that content, as it is considered that they can exercise “editorial control” over it. This would mean that they would face the same liability than a newspaper publisher for possible defamation contained in it, as long as they actively engaged in content moderation.

As this situation run the risk of disincentivizing platforms to moderate any kind of content posted by third parties, United States Congress Representatives designed section 230 as a way to protect platforms. It is necessary to bear in mind that by doing this the US was clearly foreseeing the future in order to protect a brand new digital economy in a way that has provided the country with the world leadership in digital business. Section 230 c(2) precludes platforms’ legal responsibility when they blocked offensive material in a good faith (“Good Samaritan”) effort, as a response to an increasing public

\(^{12}\) Goldman, 2020, p. 3.
concern on how to protect minors from obscene material on the internet. Thus, platforms do not have to face liability for good faith editorial decisions made to moderate, edit or delete content that they find “obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable (...).” Section 230 c(1) also precluded the possibility that platforms were regarded as “publishers” of any defaming information hosted or transmitted by them as long as it was provided by third parties, in order to promote free exchange of ideas through the internet.

It shall be noted that U.S. Supreme Court has not established uniform case law on section 230, which provokes discrepancies among the interpretations given by different regional courts. However, a common note when interpreting section 230 is that courts have granted internet service providers a much wider protection than the original writing of the law. In this sense, although section 230 only protects internet platforms from liability as “publishers”, in Zeran v. America Online, Inc., Court of Appeals for the Fourth Circuit interpreted the norm in a way that it also precluded “distributor” liability (in common law, distributors are liable if they know the content they are sharing is illegal); meaning that platforms may also be protected from sharing illegal third party content even though they are conscious of its illegal nature; other cases show how internet service providers are protected when they exercise any type of discretion in decisions over which content to eliminate, beyond the “good faith” requirement. Platforms are also protected from liability arising from illegal content which is provided only in part by third parties and in part by the same online platform, as long as platform’s contribution is not “material”.

Beyond defamation, section 230 has been interpreted as protecting online platforms in situations where it seems reasonable to think that individuals or firms participating in other economic activities may face liability. In Jane Doe No. 1 v. Backpage.com, LLC the defendant deliberately used its platform as a structure to facilitate human trafficking. In Force v. Facebook, Inc., the defendant recommended content created by terrorists. In Gentry v. eBay, Inc., a platform failed to prevent consumer fraud, and in Hinton v. Amazon.

13 MalwareBytes Inc. v. Enigma Software Group USA, LLC, pp. 7,8.
14 MalwareBytes Inc. v. Enigma Software Group USA, LLC, pp. 5,6.
com.dedc, LLC, it was found that the defendants sold dangerous products subject to a recall. In all these cases, online platforms remained in impunity.

Partially, as a response to these judiciary interpretations, in 2018 the Congress passed the “Allow States and Victims to Fight Online Sex Trafficking Act” (FOSTA)\(^\text{15}\), through which section 230 protection was removed (“carved out”) to online platforms which fostered or failed to eliminate with “reckless disregard” sexual services’ advertisement.

This expansive interpretation given by courts has received plenty of criticism. Professor Tushnet\(^\text{16}\) points to the unbalance among the big power given to internet service providers and the little responsibility asked in return. At the same time, professor Citron assesses that construction given to section 230 minimizes “incentives for better behaviour by those in the best position to minimize harm”\(^\text{17}\).

Debate on section 230 has even reached the political arena. In May 2020, due to an alleged anti-conservative bias displayed by some online platforms, former United States President Donald Trump issued an Executive Order\(^\text{18}\) through which, among other measures, it required the Federal Communications Commission (FCC) to issue regulations to clarify section 230 in a way that narrows its broad scope of application as a first step towards a radical reform of section 230. Even though before the elections President Joe Biden showed his disconformity with current state of section 230\(^\text{19}\), last May he issued an Executive Order derogating the previous Executive Order issued by Trump, and thus stopping any process of reforming the CDA. Besides that, the official attitude of Biden Administration towards section 230 has been, by now, neutral. We shall briefly mention how


\(^{16}\) *Ibid* at pp. 10-12.

\(^{17}\) “(B)usinesses that are not merely failing to take ‘Good Samaritan’ steps to protect users from online indecency but are actually being ‘Bad Samaritans’” (Citron & Wittes, The Internet Will Not Break: Denying Bad Samaritans § 230 Immunity, 2017, p. 409).


\(^{19}\) Kelly, 2020.
several bills have been introduced to amend section 230, without achieving success. Among the most interesting ones we could mention the “EARN IT Act”, that would condition section 230 protection to online platforms in cases of child pornography to the fulfilling of a list of “best practices”, such as eliminating end-to-end encryption, or the “Ending Support for Internet Censorship Act” that would regard section 230 protection as a privilege that would be obtained through the achievement of a certification expedited by the Federal Trade Commission if the online platform is able to show that it moderates content on a neutral, unbiased way.

III. CHINA AND THE PRINCIPLE OF CYBER-SOVEREIGNTY

China understands the whole Internet in a different way to Western countries. While for the latter, the internet is a global, “free and open place”, for China the internet is no different to any other space falling within its territory and, as such, it shall be able to exercise its “cyber-sovereignty” (wangluo zhuquan) over it. We know, however, that the concept of sovereignty presupposes physical boundaries separating one country from another, within which power is exercised. China, then, translate that power to the ubiquitous cyberspace.

China has created a huge filtering system, commonly known as “the Great Firewall”, which, in practice, isolates Chinese citizens from foreign online content. Also, China has favoured those actors, data, infrastructure, required to provide internet services that are located within its territory: for example, favouring national suppliers of technology over foreign ones or requiring that DNS servers registered in China are located within its territory. This way China ensures that, despite not being able to control the cyberspace per se, it can substantially influence it by controlling physical actors and process whereby internet operates.

Regarding Chinese legal regime, all types of internet service providers must comply with abundant regulation (and are subject to big liability risks) when operating in China. E-commerce platforms,

---

20 Creemers, 2020, p. 108.
21 Ibid at 116-121.
for example, may be held jointly and severally liable for damages caused by products sold by them if the platform fails to retire the link in a timely manner\textsuperscript{22}.

Focusing now on social networking platforms, they are subject to demanding \textit{ex ante} and \textit{ex post} controls. \textit{Ex ante}, online operators engaging in providing information must obtain a license in order to provide such services\textsuperscript{23}. These licenses are granted by the Cyberspace Administration of China (CAC), an official institution designed to direct and oversee online content. Social networking platforms are obligated to record and, if asked for it, grant to the authorities users’ data about their address, name, or time of connection\textsuperscript{24}.

Contrary to what happens in the United States through section 230, social networking platforms may be held liable for the content that they host. In this sense, they are required to comply with China’s censorship policy, through banning content which is regarded as “politically harmful”\textsuperscript{25}. This broad concept encompasses, \textit{inter alia}, pornographic, defamatory and discriminatory content, but also includes messages opposed to constitutional principles, rumours and other content which may subvert the public order\textsuperscript{26}. Together with engaging in censorships, platforms are required to exercise additional obligations in terms of collecting and verifying data about users’ real identities and grade them accordingly to whether they have breached certain rules. Users seriously breaching law may get their accounts closed\textsuperscript{27}.

Referring to the sanctions, users publishing or disseminating “politically harmful” content may be subject to fines, or even criminal liability\textsuperscript{28}. Chinese Cybersecurity Law also opens the door to the possibility of extending liability to social networking platforms for

\begin{itemize}
\item \textsuperscript{22} Xia, 2019.
\item \textsuperscript{23} See https://www.cecc.gov/international-agreements-and-domestic-legislation-affecting-freedom-of-expression. Actually, there are cases where Chinese government is given a special share, or veto power inside private companies dedicated to re-publishing, which grants it direct controlling power over the firms’ decisions (Wang, 2020, p. 6).
\item \textsuperscript{24} Congressional Executive Commission on China.
\item \textsuperscript{25} Wang, 2020, pp. 2-4.
\item \textsuperscript{26} Cyberspace Administration of China, 2019.
\item \textsuperscript{27} Cyberspace Administration of China, 2017.
\item \textsuperscript{28} Zhang, 2019.
\end{itemize}
hosting these contents. Sanctions may include fines, obligation to restore the damage, or even permanent closure of the platform or app²⁹.

Notwithstanding, a good example of how Chinese authorities understand the scope of powers of the state in relation with Big Tech companies can be seen in recent regulation and sanctions activities on June and July 2021, when China governmental agencies started deep consequences actions against powerful digital companies³⁰.

IV. EUROPEAN UNION AND THE E-COMMERCE DIRECTIVE

1. E-commerce Directive

The pinnacle of European legal order on social networking platforms’ liability is the Directive 2000/31/EC³¹, commonly known as the “e-commerce Directive”. E-commerce directive applies to “information society services”, which are defined as any service normally provided for remuneration, at a distance, by electronic means and at the individual request of a recipient of services³². This concept encompasses different types of services, such as online press, e-healthcare, search engines and, also, social networking platforms. The material scope of the Directive comprises different aspects related to electronic commerce activities, among which we find the system of “Liability of intermediary service providers” for third parties-generated content.

Under the e-commerce Directive, social networking platforms are protected from liability arising from illegal activities performed by their users through their platform. However, unlike section 230, this protection is not absolute: the e-commerce Directive establishes “safe harbours” for three specific types of services. These safe harbours will be conditioned to more and more requisites depending on the degree of control that each type of service provider has over the information.

The first services, described in article 12, are “mere conduit”, or transmissions of information which are not initiated, nor modified (and in which the recipient is not selected) by the service provider. If

²⁹ Wang, 2020, p. 4.
³⁰ The Economist, 2021.
these services fulfil the requisites described in the Directive, service provider is regarded as lacking control or knowledge about the content, and thus it does not face liability.

Article 13 protects services regarded as “catching”, which means the transmission of information requiring “automatic, intermediate and temporary storage of that information”, with the sole purpose of increasing efficiency, and without modifying the information. This type of service requires that information is stored during a longer period of time than “mere conduit” services, and thus extends service provider’s responsibility over it, by subjecting its safe harbour to more demanding standards. For example, service provider shall disable access to a specific information if it has “actual knowledge of the fact that the information at the initial source of the transmission has been removed from the network, (...) or that a court or an administrative authority has ordered such removal or disablement”. Such an obligation does not exist for “mere conduit”.

Finally, article 14 of the Directive protects mere “hosting” of illegal information\(^{33}\), without any further need to transmit it, as long as the service provided lacks knowledge of the illegal nature of the content. As this service implies a closer relation among the service provider and the content, the former will be protected from the illicit nature of the latter only if he ignores it, or if he acts expeditiously when he notices. Note that, contrary to “catching” the illegality of the content shall be appreciated by the service provider itself, and not by an outsider court or an administrative authority, confirming that obligations increase as so does control over content.

This requisite has received criticism by scholars\(^{34}\) for favouring service providers who do not engage in any type of moderation over those who regularly delete illicit content and do not remove a punctual comment, de facto incentivizing service provider not to engage in content monitoring, similar to what we saw with Stratton Oakmont, Inc within the American context (what some experts call the “moderators dilemma”). However, the European Commission

---

\(^{33}\) Note that, whereas the Directive does not contain a definition of “illegal content”, it is likely that it encompasses content which is illegal under EU or national law (European Commission, 2012, p. 25).

\(^{34}\) Lodder, 2017, pp. 32,33.
clarified in its communication on Tackling Content Online that “In particular, the taking of such measures need not imply that the online platform concerned plays an active role which would no longer allow it to benefit from that exemption”\(^{35}\).

Apparently, in all these cases, e-commerce Directive requires, through its recital 42, that the activities have a “mere technical, automatic and passive nature implying that the service provider has neither knowledge of nor control over the information”, so the service provider does not make any material (as opposed to technical) contribution to the data. Court of Justice of the European Union (CJEU)’s case law has shifted its view on whether these requirements apply also to “hosting” services, however, in \(L’\text{Oréal} v. \text{eBay}\), the CJEU stated that, in order to benefit from safe harbour in art. 14, service providers cannot have knowledge about the illegal nature of the information\(^{36}\), neither because of the general way in which the service is designed, nor because it has specific knowledge about the particular illegal content. Even without having such knowledge, safe harbours will not be applicable if a “diligent economic operator” would have identified such illegality\(^{37}\). Through this standard, the Directive goes beyond establishing merely an obligation to react when the illicit content is detected, requiring a positive duty of reasonable care\(^{38}\).

Besides that, Recital 43 requires that services are “no way involved with the information transmitted”. However, this does not apply for “hosting” services.

The Directive also protects internet service providers by preventing national authorities from imposing general monitoring obligations over them, through its article 15, as this would be equivalent to exercising ex ante censorship. According to recital 45, this does not preclude the possibility that Member States issue prohibitory injunctions tending to terminate or prevent a specific infringement. These injunctions, however, cannot impose “general filtering sys-

\(^{35}\) European Commission, 2017.

\(^{36}\) In its previous case law (see joined cases C-236/08, C-237/08 and C-238/08, \(\text{Google France and Others v. Louis Vuitton Malletier}\) CJEU had demanded a higher requisite of neutrality to the service provider (Valcke, Kuczerawy, & Ombelet, 2016, p. 8).

\(^{37}\) C-324/09, \(L’\text{Oréal} and Others v. \text{eBay International AG}\) para. 112-120.

\(^{38}\) Valcke, Kuczerawy, & Ombelet, 2016, p. 10.
tems”, understood by the CJEU as systems applying to all customers, as a preventive measure, at the service provider’s expense and indefinitely\textsuperscript{39}. According to recital 48, however, national authorities are also allowed to impose reasonable duties of care to detect certain illegal activities. Some scholars have criticized the vagueness of its wording and pointed to the contradiction among recital 48 and article 15\textsuperscript{40}. Whereas the exact scope of these types of prohibitions remains unclear, in \textit{Eva Glawischnig-Piesczek v Facebook Ireland Limited in.}, the CJEU declared admissible that a Member States orders the removal of a specific information worldwide\textsuperscript{41}.

In addition to the above, the European landscape of obligations for online platforms goes beyond the e-commerce Directive. Different sector-specific rules enhance the obligations of internet service providers in fields like fight against hate speech and violence\textsuperscript{42}, copyright infringements\textsuperscript{43} or terrorism\textsuperscript{44}. However, for the sake of simplicity, we will limit our discussion to the general rules contained in the e-commerce Directive.

Finally, together with hard law, EU uses soft regulation as a faster and more flexible way to fight specific types of illegal content. EU encourages and coordinates the creation of codes of conduct, fora or memoranda. Find bellow a table containing some of the most relevant initiatives taken in this regard.

\begin{table}[h]
\centering
\begin{tabular}{|l|}
\hline
\textbf{Table 1: Relevant Initiatives} \\
\hline
\end{tabular}
\end{table}

\textsuperscript{39} See C-70/10 - \textit{Scarlet Extended v. SABAM} and C-360/10 - \textit{SABAM v Netlog NV} (Valeke, Kuczerawy, & Ombelet, 2016, p. 8).
\textsuperscript{40} Barceló & Koelman, 2000.
\textsuperscript{41} \textit{Eva Glawischnig-Piesczek v Facebook Ireland Limited in.} Case C-18/18.
\textsuperscript{44} Regulation (EU) 2021/784 of the European Parliament and of the Council of 29 April 2021 on addressing the dissemination of terrorist content online.
De Streel, Buiten, & Peitz, 2018, p. 32.

2. Implementation of the Directive

Regarding the implementation of the Directive, in 2007, the European Commission published the results of a report prepared by an external consultant detailing how the e-commerce was being transposed at national level.

Whereas there may be certain degree of divergence among the scope of the concept of information services, liability safe harbours’ transpositions has been generally homogeneous\textsuperscript{45}, maybe with the exception of hosting services, as their requisites can vary across some Member States\textsuperscript{46}. Finally, art. 15 has been source of big litigation. In this sense, plenty of case law has tried to make compatible the previous apparent contradiction among this article (prohibiting the imposition of general monitoring obligations) and injunctions imposing specific monitoring obligations\textsuperscript{47}.

\textsuperscript{45} Kastberg & al, 2007, pp. 32-34.

\textsuperscript{46} For example, some countries differentiate among “actual knowledge” (subjected to criminal liability) and “knowledge of elements from which illegal nature is apparent” (subjected to civil liability). Similar divergences can be found with respect to when a content is “manifestly illegal”, the category of “expeditiousness”, \textit{inter alia} (\textit{Ibid} at 32-37).

\textsuperscript{47} \textit{Ibid} at 52.53.

<table>
<thead>
<tr>
<th>Type of illegal content</th>
<th>Hard-law</th>
<th>Soft-law</th>
<th>Co/self-regulation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BASELINE</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All types of illegal content online</td>
<td>- Dir. 2000/31 e-commerce</td>
<td>- Communication 2017 Illegal content online - Rec. 2018/334 Illegal content online</td>
<td></td>
</tr>
<tr>
<td>Terrorist content</td>
<td>- Dir. 2017/541 Terrorism</td>
<td>- Rec. 2018/334 Illegal content online</td>
<td>- EU Internet Forum</td>
</tr>
<tr>
<td>Hate speech</td>
<td>- Dir. AVMS in case of video-sharing platforms</td>
<td>- CoC Illegal hate speech online (2016)</td>
<td></td>
</tr>
<tr>
<td>IP violation – copyrighted content</td>
<td>- Prop Dir. Copyright DSM</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IP violation – counterfeit goods</td>
<td></td>
<td>- MoU Counterfeit goods online (2011-2016)</td>
<td></td>
</tr>
</tbody>
</table>
3. **Fight against disinformation**

EU acknowledges that disinformation is a threat affecting fundamental rights such as opinion or expression, and therefore tries to fight it using different tools. Currently, it does so through the collaboration with different stakeholders involved in the problem.

Among these initiatives we can begin by mentioning the Code of Practice on disinformation (CPd) presented in 2018, *“the first worldwide self-regulatory set of standards to fight disinformation voluntarily signed by platforms”*⁴⁸. Codes of conduct can be a more effective tool than hard regulation to tackle online threats due to their higher degree of flexibility. The CPd relies on a self-regulatory approach. It has been signed by different types of stakeholders that compromise themselves to perform a series of best practices in fields such as revising their advertising policies to reduce monetization to content and users that contribute to diffusion of information, increasing transparency of political advertising, limiting the use of techniques such as spam of bots or empowering consumers⁴⁹.

Currently, the CPd has been signed by important platforms such as Facebook, Microsoft or Twitter. Regarding its effectiveness, while it is true that the first results show how these actors have acted in some aspects, such as adopting measures against false accounts, increasing transparency of political advertising or rising visibility of reliable sources, the Commission stresses the need to improve in other areas, like ensuring a more uniform application of the CPd across countries, the establishment of a clear, objective monitoring system, or a closer collaboration with fact-checkers, among other points⁵⁰. The COVID 19 monitoring programme, put in place by the Commission to regularly control platforms’ efforts to fight COVID disinformation, confirmed some of these shortcomings⁵¹.

In order to tackle some of the mentioned limitations, the Commission recently issued additional guidance for the signatories of the CPd⁵². Additionally, we shall mention that voluntary codes of conduct,

---

⁴⁹ See the annex to the CPd (European Commission, 2018).
including the CPd, will become especially relevant for large platforms after the future Digital Services Act is passed, as the adoption of such standards can be regarded a “risk mitigating measure”, in the light of its article 27\textsuperscript{53}. Time will show whether this actions will be enough to make the CPd a more effective instrument to coordinate stakeholders to fight disinformation.

We shall also mention how, in order to fight strategic disinformation campaigns, the East Strategic Communication (StratCom) Task Force together with other Task Forces were created to work closely with Member States and EU institutions to fight against disinformation campaigns. They do not engage in counterpropaganda, but rather focus in finding and pointing specific disinformation messages out. Since its creation in 2015, East StratCom Task Force has identified more than 4500 pieces of disinformation originated in Russia\textsuperscript{54}.

See below a graphic illustrating some of the actions adopted by the EU to fight against disinformation.

\textbf{OVERVIEW OF EU JOINT AND COORDINATED ACTION AGAINST DISINFORMATION}

- Launch of the EEAS East StratCom Task Force: 03/2015
- Communication on tackling online disinformation: a European approach: 04/2016
- Code of practice against disinformation: Spring 2018
- Joint Framework on countering hybrid threats: 09/2018
- Package of measures securing free and fair European elections: 11/2018
- Launch of the Observatory for Social Media Analysis (SOMA): 05/2019
- Action Plan against disinformation: 12/2018
- Reporting by online platforms: 03/2019
- Set up of the Raid Alert System: 03/2019
- Inaugural meeting of the European cooperation network for elections: 01/2019
- European Media Literacy Week: January-May 2019
- European Commission’s Contribution to the informal EU27 leaders’ meeting in Sibiu: 05/2019


\textsuperscript{53} See infra chapter on the Digital Services Act.

\textsuperscript{54} European Commission, 2018, pp. 4,5.
4. Germany and France: fighting disinformation at national level

Besides efforts displayed by the European Union (EU), some Member States have developed their own initiatives to fight disinformation campaigns compromising state interests. In this sense, two of the most innovative regulatory approaches come from two norms approved in Germany and France.

In 2015, German social networking platforms were being shaken by hate speech messages, partially due to the social upheaval caused by the huge number of refugees arriving from Syria. Failure to remove illegal content by platforms led to the adoption of the "Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzen" ("Network Enforcement Act" or "NetzDG") on June 30, 2017. This norm applies to social networking platforms with more than 2 million users in Germany, and imposes them obligations like publishing transparency reports, or removing hosted content constituting offences like disseminating propaganda of unconstitutional organizations or defaming the main institutions of the State. Social networking platforms must provide its users with a transparent and effective complaint mechanisms that allows them to report content constituting any of the previously mentioned offences, which, in principle shall be deleted within 7 days since the reception of the user’s report. In case of breaching obligations under NetzDG, social networking platforms can be subject to fines up to 50 million euros. In this sense, in July 2019, Facebook was fined with 2 million euros for failing to fulfil its transparency duties.

Regarding the second case, the "Loi du 29 juillet 1881 sur la liberté de la presse" ("Law of 29 July 1881 on the freedom of press") had been in force for more than 130 in France. This law punished bad faith creation or spreading of fake news, previewing fines up to 135,000 euros. If those messages negatively affect the result of an election, French Electoral Code regulates, on its article L97, sanctions up to 15,000 euros and one year in jail. Despite these norms, fear of disinformation campaigns increased after the leaking of large amounts of false emails allegedly belonging to 2017 presidential candidate

---

55 Heldt, 2019.
Macron\textsuperscript{57}, prompting additional measures to fight disinformation. As a result, in 2018, the new \textit{Loi n° 2018-1202 relative à la lutte contre la manipulation de l’information} (“Law against the manipulation of information”) was enacted. This norm imposes a series of duties to online platforms including social networking platforms, related to increasing their levels of transparency and actively collaborating to fight disinformation. Among its most remarkable innovations we could mention the creation of a duty to establish clear systems allowing users to report false information, an increase in the transparency requirements related to algorithms and paid advertisements relating politically controverted material, or the requirement of actively acting against massive disseminators of information, \textit{inter alia}.

Transparency obligations become stricter over the course of the three months previous to an electoral process, during which platforms shall provide information about their advertising-generated income and the way they are utilizing user-generated data. Breach of these latter obligations goes in hand with sanctions up to 75,000 euros and a year in prison\textsuperscript{58}. During this pre-electoral period, certain subjects (public prosecutors, political candidates…) are given the possibility to denounce to a judge (“\textit{juge des référés}”) the online spread of a piece of disinformation that may affect the result of the election, which will adopt necessary and proportionate measures to prevent its dissemination within 48 hours. Also, the law strengthens capacities of the \textit{Conseil Supérieur de l’Audiovisuel} (CSA), even recognizing it the capacity to remove the license of an operator controlled by a foreign state, under certain conditions.

V. \textbf{The Digital Services Act: a shift in the paradigm?}

After a public consultation ended in September, 2020, the European Commission announced its proposal for the Digital Services Act Package as an additional step towards regulating how the internet works. This initiative includes two Regulations: the \textit{Digital Markets Act} (DMA) and the \textit{Digital Services Act} (DSA). DMA will limit the power of big online intermediaries regarded as “gatekeepers”,

\footnotesize{\textsuperscript{57} Boring, 2019.  \\
\textsuperscript{58} \textit{Ibid.}}
mainly through the prohibition of certain unfair practices, in order to reach a more open and equal digital market. DSA will build on the e-commerce Directive, to create a more extensive system of liabilities for online service providers.

Although the DSA still has to go through a long procedural journey before being passed, we can already note some remarkable aspects from the text of its proposal. The DSA will contribute, but not substitute, existing system of liabilities of online services providers. In this sense, the text of the e-commerce Directive, together with sector-specific Regulations and Directives will remain valid with one exception: exemptions from liability contained in articles 12-15 of the e-commerce Directive are deleted, and are incorporated in articles 2-5 of the DSA. The consequence of this is that safe harbours, as interpreted by CJEU case law, will have the rank of a EU Regulation, and its application will need to be uniform across Member States.

The DSA contains important innovations that may strongly affect the current system of social networking platforms’ secondary liability. The Regulation introduces a whole new set of obligations for all intermediary services providers. However, these obligations are not equal for all providers, but they work on an incremental way: there are a common set of obligations, comprised in section 1 of Chapter III (such as designing a single point of contact, or publishing annual reports about their moderation practices) which apply to all types of providers of intermediary services, regardless of whether they are big social networking platforms or DNS registrars. However, the rest of the obligations will apply, or not, depending on the type of online service provider affected and its respective size. Section 2 contains obligations for hosting service providers and platforms; section 3 introduces additional obligations only for platforms, and if those platforms have more than 45 million users in the European Union, they have to comply with section 4.

As we mentioned, providers of hosting services and online platforms are also affected by section 2, which includes the obligation to put in place Notice and Actions mechanisms, allowing users to easily notify the service provider about the presence of illegal content on its service. According to art. 14.3 DSA, these notices can be regarded as giving rise to “actual knowledge” of the illegality of
the content, for the purpose of the safe harbour for hosting services (previously included in art. 14 e-commerce Directive). Unfortunately, many relevant elements about the process through which this actual knowledge is obtained remains unclear, such as whether the notice gives rise to actual knowledge from the moment it is sent, or whether the platform enjoys some reasonable time to examine it before being regarded as possessing knowledge. Together with this, in case a piece of information uploaded by a user is removed or blocked by the service provider, section 2 also imposes the obligation to provide the user with a clear statement of reasons justifying the decision.

Together with the previously mentioned obligations, online platforms will also have to comply with section 3. Under this section, users will be able to challenge decisions made by social networking platforms which negatively affect them (such as deleting content uploaded by them, or even their accounts) before a free, user-friendly internal complaint-handling mechanism that must be previewed by the platform, and before an independent settlement body. According to art. 21, platforms will also assume the duty to promptly inform the authorities of a Member State in case they “become(s) aware of any information giving rise to a suspicion that a serious criminal offence involving a threat to the life or safety of persons” has occurred or it is likely to occur. Other obligations contained in section 3 are the obligation to give priority to information sent by institutions recognized as “trusted flaggers” in the context of Notice and Action mechanisms, to suspend the accounts of users which frequently provide manifestly illegal content, and to increase their level of transparency. More specifically, platforms shall make sure that users can know, in real time, why a piece of advertisement was particularly targeted to him.

Section 4 includes obligations for very large online platforms (VLOPs), which are those with a number of users exceeding 10% of the European Union population. Exhibit 54 describes how these VLOPs can generate “societal risk” due to their huge influence. In order to balance such power, article 26 includes the obligation to conduct annual assessments about the “systemic risks” stemming from their services in the Union. Within this broad category, article 26 enumerates different types of risks: first subparagraph relates to risk of “dissemination of illegal content through their services”;
second, “any negative effects for the exercise of (...) fundamental rights”; finally, “intentional manipulation of their service (...) with an actual or foreseeable negative effect on the protection of public health, minors, civic discourse or actual or foreseeable effects related to electoral processes and public security. Once identified, article 27 imposes the obligation to put in place mitigating measures specifically tailored to the identified risks, including the modification of their moderation or content curation practices.

A recent article pointed out how these two regulations are problematic in several ways. The concept of risks of “negative effects” over fundamental rights, or other public goods imposes the obligation over VLOPs to act against content which is not itself illegal (as otherwise, it would be comprised within first subparagraph, previously mentioned). Beyond this, the concept of “negative effects” seems excessively broad. No one doubts that news about a scandal involving a country’s prime minister generates negative effects over his right to private life. However, this does not mean that such publications must be removed, as these negative effects are proportionate to citizens’ right to freedom of information. In addition, the problem is not only that the exact scope of this “negative effects” concept is not clear, but that this interpretation will be left in the hands of the VLOPs, which will presumably tend to delete “borderline” content as a consequence of this uncertainty.

Besides articles 26 and 27, section 4 includes additional transparency obligation. For example, VLOPs will have to make available the main criteria used by their recommender systems to curate content. They will also have to publish a repository including information about advertisements displayed in the platform in the previous year, an annual report containing the main risks identified, and measures adopted pursuant to articles 26 and 27. Finally, in order to ensure compliance with the aforementioned obligations, VLOPs will have to appoint a compliance officer and be subject to annual independent audits.

Additionally, section 5 fosters the adoption of voluntary self or co-regulatory agreements by the service providers themselves,

---

59 Barata, 2021.
coordinated by the Commission and other public bodies, in order to complement obligations set in the DSA. These agreements can take the form of industry standards, codes of conduct or crisis protocols. Although the writing of section 5 leads to the conclusion that all these measures are voluntary, according to recital 68, adherence to codes of conduct by VLOPs can be considered a risk mitigating measure, in the sense of article 27, and refusal without proper explanation to adhere can be a determinant factor when assessing the platform’s compliance with DSA. Under this recital, lines separating voluntary and mandatory measures become partially blurred.

In case of breaching those obligations, digital service providers can be fined by the Digital Service Coordinator, designated at national level, with fines up to 6% of their annual turnover. Under some circumstances, if the breach is performed by a VLOP, the Commission can choose to initiate the proceedings. In case the Commission finds a breach, fines cannot be bigger than 6% of that VLOP’s total turnover of the previous financial year.

VI. CONCLUSION

Secondary liability over social networking platforms can be a useful tool for legislators when trying to constrain users’ conduct and deter them from carrying out certain activities deemed undesirable. However, just like any other norm tending to shape human behaviour, rules imposing secondary liability involve a trade-off between freedom and security: increasing security through the direct or indirect punishment of undesirable content goes in hand with reducing individual freedom, and vice versa. Divergent answers to the question about what levels of freedom and security are optimum are at the root of the differences between diverse legal systems.

United States legal system, for example, prioritizes the protection of individual freedoms, even though this implies that national authorities will lack legal tools to prosecute certain online intermediaries whose conduct was reprehensible. They can remain in impunity. China chose an opposite option, by imposing extensive legal duties over platforms, as a way to minimize problematic content. By doing so, however, online freedom is strongly repressed.
The European Union designed an intermediate system: online service providers are protected, as long as they comply with some requisites that show that they are (at least partially) alien to the illicit nature of content transmitted or hosted by them. However, in recent years, as a response to emerging threatens online, we can perceive a shift in the European paradigm, towards a more interventionist approach, even though it has implied intervention on individual freedoms. Some remarkable examples at national level could be France and Germany.

The recent proposal for the DSA published by the Commission is the last step taken in this direction. This Regulation will require that online services providers, in general, and VLOPs, in particular, assume positive duties to collaborate with states when fighting threatens like illegal content online and disinformation. Time will show whether this will be an adequate and proportionate measure, or whether it will imply the destruction of the internet as we know it now, built over the principles of freedom and openness, free from extensive state intervention.

What we have to underline here is that, if the Commission wants to increase regulatory pressure, it shall do so with rules using clear and straightforward legal categories to minimize uncertainty. As we mentioned before, elements such as the process through which “actual knowledge” is acquired, according to art. 14.3, or concepts such as the “negative effects” described in art. 26 leave the final interpretation in the hands of the platforms themselves. In this sense, it would be desirable for the European institutions to issue clear guidance to minimize legal uncertainty, which is the only way to ensure that, when trying to increase online security, restrictions over individual freedom are minimized.

Notwithstanding, it is clear that digital revolution will continue to change reality and it is very difficult for regulators, due to the speed of change, to keep pace with this incremental evolving landscape. Automation and AI will have an important role in order to set up systems that will allow to eliminate harmful content. However, it will sometimes be difficult to establish a clear distinction between illegal content or to consider information to be false or true. In democracies, this will sometimes be a hard task because democracy is based on free public opinion. Without that, there is no democracy. Other political
systems are not constrained by these questions and the truth is always decided by the political powers.

We are living a process of disruption, a moment of change in which the driving force is the digital revolution. It is affecting the way we live and, of course, the way states regulate. The fact that it is complex to regulate with the tools we have now cannot be an excuse to do nothing or to fail in trying to solve the problem. What is not legal in the real world cannot be legal in the digital world. Democracies have created a system of government whose foundations are the rights and freedoms of individuals. This is the very core of the system and its base of legitimation. In defence of this system, it cannot be tolerated that, due to the difficulties inherent in the regulation of digital reality, states and international organizations quit from the most important function entrusted to them: the protection of rights and freedoms. This is the model that the EU seeks to protect through the successive regulations it has approved. An approach based on the rule of law. Previous regulatory efforts such as the one carried out by the EU in the area of data protection have taught us that this regulatory activity is a very versatile soft power tool. In relation to the regulation of disinformation in the net, the EU has opted for a legal approach that can be the worldwide reference in order to regulate the protection of rights and freedoms. Some opinions suggest that this EU concern for rule of law in a digital context, will left behind western democracies in the race of digital revolution. This position may seem to be a disadvantage in the new digital society we are shaping, but it is necessary to think about the cost of non-regulation in defence of the very foundations of our democracies. Because this issue is not another regulatory question, as usual, it is directly related to the kind of future political systems and societies we decide to establish. Liberal democracy is under attack everywhere. Cyberspace, a new scenario that has been out of law’s scope until now, is the arena where the future of nations will be set up mostly. Unless we give up the idea of the rule of law, we need to regulate that reality in order to protect the same political system that created this outstanding and game-changing new reality called Internet.
BIBLIOGRAPHY


