CREDIBILITY-ENHANCING REGULATORY MODELS TO COUNTER FAKE NEWS: RISKS OF A NON-HARMONIZED INTERMEDIARY LIABILITY PARADIGM SHIFT

Teresa Rodríguez de las Heras Ballell*

Safe harbor provisions for electronic intermediary service providers represent a key common policy in worldwide Internet regulation. Although there are disparities in scope, applicable conditions, and effects, intermediary liability exemptions have been extensively incorporated into most jurisdictions and are the backbone of electronic commerce and information society services (in the EU terminology) legal framework. To date, it has been a rather undisputed assumption that the intermediary (non-) liability paradigm has accelerated the expansion and consolidation of digital activities. Safe harbors do rightly allocate incentives to reach a compromise between the free provision of intermediary services that are arguably critical for the survival and development of the digital society, and the reasonable protection of rights. However, today's digital scene has changed considerably, so as to challenge the sustainability of intermediary liability paradigm and put into question the continuation of intermediary liability in its current form. The proliferation of fake news and alarming use of disinformation campaigns based on the dissemination of deliberately false information have precipitated the debate on the actual and prospective role of digital intermediaries and the suitability of current liability rules to enhance trust and counter misinformation. Intermediaries are a determining component of misinformation machinery. Although fake news is typically user-fabricated content, intermediaries provide them with the features needed to gain impact: accessibility, visibility, virality, and, as a consequence, perceived credibility. Therefore, because the original source of the disinformation can neither be easily located nor effectively combated,

* Professor of Commercial Law, Universidad Carlos III de Madrid. teresa.rodriguezdelasheras@uc3m.es and *2017-2018 Chair of Excellence*, Oxford University, Harris Manchester College, Commercial Law Center (UC3M-Santander Chairs of Excellence Program). Member of the Expert group to the EU Observatory on the Online Platform Economy.

regulators turn their attention toward intermediaries as they are more accessible, in an attempt to control this growing information challenge. If accessibility, visibility or virality were contained, the effects of misinformation would be significantly restrained. The policy options that should be adopted to achieve such a positive outcome are difficult to pinpoint. Approaches differ, and such disparities contribute to continue debilitating credibility and foster jurisdictional arbitrage and "platform shopping" as a new version of forum shopping. In such a context, the aim of this Article is to dive into the global debate about the need for a paradigm shift in the liability policy towards an increasing involvement of digital intermediaries and platform operators to enhance credibility and counter misinformation. To that end, the Article will analyze and compare regulatory models and contrast their implications.

TABLE OF CONTENTS

I. THE LAYERS OF DIGITAL INTERMEDIATION: ACCESSIBILITY, VISIBILITY, AND CREDIBILITY

Digital intermediaries play a critical role in our digital society. Business transactions, economic activities, social interaction, educational and cultural environments, and other varied dimensions of digital economy are widely facilitated, enabled, and encouraged by intermediaries.[1] Essentially, digital intermediaries are key facilitators of digital activities by providing accessibility and visibility of digital content, data, and information, and generate trust by enhancing credibility in digital interactions. Digital intermediary activities constitute the backbone of the digital living.

To that end, digital intermediation evolves and transforms to progressively satisfy new needs, repair failures, and face novel challenges of a changing and dynamic digital society. Therefore, in tracing its evolution over the last decades, several superimposed layers of digital intermediation[2] can be discovered. Such a digital archeological initiative reveals how digital intermediaries have successively addressed and fulfilled the most urging need of digital communities at each stage of evolution. First, accessibility: intermediaries have focused on providing the most basic need for digital users: readily accessible digital content and services. Accessibility would be then the first and primitive layer. Second, visibility: as the vast informative exuberance of our overinformed digital society incremented, real accessibility and attention-capturing-and-retaining capacity dramatically decreased. To ensure effective access to pertinent, convenient, and sought information, visibility-providing strategies are imperative. Accordingly, intermediaries are necessary to provide and enhance visibility. Third, once extensive accessibility and high visibility are assured, credibility becomes the scarcest value in the digital scene. Trust generation constitutes the most critical factor for the sustainability and the growth of the digital society. Not surprisingly, intermediaries make efforts to generate confidence and create trustworthy environments as trusted third parties. Metaphorically, these "layers of digital intermediation" conform today's digital geology.

---

[1] Bailey, J. The Emergence of Electronic Intermediaries, Proceedings of the 17th ICIS, Cleveland, OH, 391-99, (1996); Bailey, J. and Bakos, Y. An Exploratory Study of the Emerging Role of Electronic Intermediaries, International Journal of Electronic Commerce, 7-20 (1996); Bakos, Y., The Emerging Role of Electronic Marketplaces on the Internet, Communications of the ACM, 35-42 (1998); P.K. Kannan et al., The Internet Information Market: The Emerging Role of Intermediaries, Handbook on Electronic Commerce, 569-90 (2000).

[2] *See* Teresa Rodríguez de las Heras Ballell, Intermediación electrónica y generación de confianza en la Red: escenarios de riesgos y responsabilidad, Revista Española de Seguros, núm. 153-54, 43-68 (2000).

These fundamental roles of intermediaries in digital society do, however, constitute their greatest vulnerability, as they exacerbate their exposure to risk. In fact, insofar as intermediaries provide accessibility, and visibility to user-generated content, where such content is illegal, harmful, or false, it might be easily argued that they do indeed facilitate the infringement, enable the causation of damage, or even amplify the impact by providing tools to disseminate the information. Likewise, to the extent that intermediaries act voluntarily, or are involuntarily treated – on grounds of the reasonable expectation of users – as trusted third parties, they might arguably be endorsing or supporting the content they transmit, store, search, link, or make available. Accordingly, their exposure to liability increases greatly.

This Article is based on the premise of the above-described two-faced role of intermediaries to devise possible strategies to counter fake news. Intermediaries are a determining component of misinformation machinery. Although fake news is typically user-fabricated content, intermediaries provide them with the needed features to gain impact: accessibility, visibility, virality, and, as a consequence, perceived credibility. Therefore, because the original source of the disinformation can neither be easily located nor effectively combated, regulators turn their attention toward intermediaries as they are more accessible, in an attempt to control this growing information challenge. If accessibility, visibility or virality were contained, the effects of misinformation would be significantly restrained. The policy options that should be adopted to achieve such a positive outcome are difficult to pinpoint. Where the intermediary liability regime ("safe harbor" provisions) was clearly designed in the appreciation of the positive role of intermediaries as providers of accessibility, visibility, and credibility, and with the aim to preserve it, how to face their contributory role from a legal perspective in the misinformation machinery is still undefined and rather uncertain. The new challenges might require a paradigm shift on liability. There are some signs that point in this direction that are being noticed.

In the European Union, how to respond to this problem is not yet defined. Recent debates at Parliament reveal a lack of agreement on how to best counter fake news. Accordingly, all regulatory alternatives are under consideration.[3] Some voices are more inclined to support a paradigm shift on liability to incentive intermediaries to act expeditiously to remove illegal

[3] Divergences among Member States to combat disinformation are also revealed by the Report of the Presidency to the European Council on June 20-21, 2019, on countering disinformation and the lessons learnt from the European elections sent to Delegates by the Presidency of the Council of the European Union at https://www.euractiv.com/wp-content/uploads/sites/2/2019/06/imfname_10910650.pdf.

(false) content, resort to fact-checkers, implement effective notice and complaint systems, or even assume a general duty to monitor in order to detect obviously false information. Other positions however, seem keener on preserving the current liability system or to rely on user-controlled monitoring schemes.[4] No EU-wide regulatory action has been adopted yet, but a High-Level Group ("the HLEG") is being set up by the European Commission to advise on policy initiatives to counter fake news and the online spread of disinformation.[5] Concurrently, Member States are individually adopting or considering the adoption of domestic initiatives. The relatively recent German *Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken (Netzwerkdurchsetzungsgesetz - NetzDG*[6]*)* – Act to Improve Enforcement of the Law in Social Networks – that entered into force on October 1, 2017, and the adoption in France of controversial legislative initiatives[7] to combat during electoral periods false information and propaganda are illustrative examples of such domestic actions.

Behind such uncoordinated response, there is a profound unfinished debate about the most effective ways to counter misinformation. As the fake-news phenomenon has aroused social alarm and political concerns, some positions defend that political action is essential. Other stances, however, tend to rely more on liability-based strategies to better allocate incentives and risks among participants. Under this approach, diverse regulatory models can be devised, such as voluntary self-regulation models, administrative sanctioning systems, or civil liability regimes. In this

---

[4] See the details on actions adopted at domestic and to be considered at EU level as described in the report *The legal framework to address "fake news": possible policy actions at the EU level*, Policy Department for Economic, Scientific and Quality of Life Policies (Author: Andrea Renda (CEPS - Centre for European Policy Studies and College of Europe), Directorate-General for Internal Policies, PE 619.013- June 2018, from p. 18 in particular – at http://www.europarl.europa.eu/RegData/etudes/IDAN/2018/619013/IPOL_IDA(2018)619013_EN.pdf.

[5] European Commission, *A multi-dimensional approach to disinformation.* Report of the independent High level Group on fake news and online disinformation, March 2018, https://ec.europa.eu/digital-single-market/en/news/final-report-high-level-expert-group-fake-news-and-online-disinformation.

[6] *Gesetz zur Verbesserung der Rechtsdurchsetzung in Sozialen Netzwerken* (Network Enforcement Act) [NetzDG] [Act to Improve Enforcement of the Law in Social Networks], Sept. 1, 2017, BGBl at 3352 (Ger.).

[7] *Loi organique n° 2018-1201 du 22 décembre 2018 relative à la lutte contre la manipulation de l'information*, https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000037847556&dateTexte=20190715 (Organic Law Against Manipulation of Information) and *Loi n° 2018-1202 du 22 décembre 2018 relative à la lutte contre la manipulation de l'information*, https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000037847559&dateTexte=20190715 (Act on the Fight Against the Manipulation of Information).

context, perceptible signs of a possible paradigm shift regarding intermediary liability regime could be indicating a choice for the latter regulatory option. A liability-based action would appear to be a very effective deterring and controlling strategy, less political, and more neutral, but in practice, it requires the transfer of power to private entities to manage the creation of opinion in the digital society. The consequences of such a model cannot be ignored.

Yet in absence of a harmonized single action, approaches differ, and such disparities contribute to the continuation of debilitating credibility, fostering jurisdictional arbitrage and "platform shopping"[8] as a new version of forum shopping.[9]

The aim of this Article is to dive into the global debate about the need for a paradigm shift in the liability policy toward an increasing involvement of digital intermediaries and platform operators to enhance credibility and counter misinformation. To that end, the Article will analyze and compare regulatory models and contrast their implications. With such goals, the analysis will be structured as follows.

First, a legal concept to embrace fake news phenomenon must be defined (*infra* Part II). Such a defining effort is conclusive to properly ponder regulatory models. My proposal is that fake news impact has two dimensions: the factual one that determines its veracity, and the social one that is based on perception. Whereas the former requires an objective test and needs a credibility reference endorsed by a trusted third party, the latter is diffuse and subjective and depends on community perception.

Second, upon the previous demarcation of the scope, alternative regulatory models will be compared (*infra* Part IV). Before diving into the different regulatory models and policy options, Part III explains the current liability regime for intermediaries to contextualize the further debate and traces the signs of change revealing a perceptible liability paradigm shift. Against such a backdrop, policy options to enhance credibility and counter misinformation could be basically implemented under the two following regulatory models: centralized credibility-enhancing models based on the

---

[8] The expression has been coined by the author and it is further described in Teresa Rodríguez de las Heras Ballell, *Rules for a Platform Economy: A Case for Harmonization to Counter «Platform shopping» in the Digital Economy*, en Ilaria Pretelli (ed.), *Conflict of Laws in the Maze of Digital Platforms - Le droit international privé dans le labyrinthe des plateformes digitales - Actes de la 30e Journée de droit international privé du 28 juin 2018 à Lausanne,* Zurich: Shulthess, 2018, pp. 55-79.

[9] *See generally* Teresa Rodríguez de las Heras Ballell, *Rules for Electronic Platforms: The Role of Platforms and Intermediaries in Digital Economy A Case for Harmonization*, UNCITRAL (Jun. 09, 2017), http://www.uncitral.org/pdf/english/congress/Papers_for_Programme/139-RODRIGUEZ-Rules_for_Electronic_Platforms.pdf.

trust-generating role of a trusted third party; and decentralized credibility models based on distributed-trust schemes and community-managed monitoring. Both models can be ably combined and coordinated. But legal rules have to decide which are the triggers to action and which are the consequences. If a policy option leading to an increasing involvement of intermediaries and platforms in detection, prevention and enforcement is chosen, the formulation of intermediary and platform duties is imperative. What kinds of duties? Is a general duty to monitor under consideration? Should automatic monitoring be deemed a general supervision? Would "best efforts" duties suffice? Would third-party fact-checkers be more effective than user-triggered flagging? An array of consequences arising from the different regulatory scenarios must be carefully considered. A legislative action aimed to intensify liability exposure could cause a retraction of intermediaries, endanger neutrality, and threaten freedom of expression under the phantom of censorship. Contrariwise, a soft-law option for promoting the adoption of code of conducts and standards could fragment the market and motivate "platform shopping."

Third, it is concluded that any action to counter fake news should be widely coordinated and harmonized at an international level. In fact, no change in liability paradigm should be conducted on a local or regional basis. Risks of a paradigm shift in intermediary liability are high, but risks of a non-harmonized action in this issue are immense. Fragmentation, discrepancies among jurisdictions, legal arbitrage and "platform shopping" would exacerbate the perception of misinformation and lack of credibility in the digital scene. This Article makes a case for international harmonization on intermediary liability.

II. DEFINING THE PROBLEM: CONCEPTUALIZING "FAKE NEWS"

The term "fake news" has become extraordinarily popular to describe many different contexts of misinformation and disinformation, but also to denote pure illegal content, defamation, parody, or simply offensive content. As a consequence, "fake news" is useful to direct attention toward a well-identified social problem, although the concept is vague, imprecise, and to a certain extent, confusing to employ in legal analysis. On the one hand, "fake news" phenomenon certainly comprises more than news. It encompasses any visual, graphical, or textual content produced and disseminated on a digital format that is likely to misinform. On the other hand, the term "fake news" is used to tag a wide array of mis- and disinformation types, including manipulated content, false content,

misleading content or fabricated content.[10] With such impreciseness, the term is unsuitable for delimiting the scope of application of any regulatory action.

Aware of the complexity of the phenomenon and the difficulties to formulate a univocal legal concept of "fake news," a purpose-specific definition and the identification of relevant factors are proposed. As the ultimate aim of this Article is to assess the feasibility and gauge the effectiveness of liability-based regulatory strategies to counter misinformation and ponder their repercussions, the definition of fake content must be formulated to achieve those purposes.

If the delimitation of the scope is approached from the perspective of intermediary liability, a categorization based on types of potential harm deriving from the content at stake becomes relevant. Precisely, harm caused by digital content can be varied in nature (moral, reputational, patrimonial, or even indirectly physical or personal) and may differ in extent. Where some digital content is likely to cause damages to identified individual persons (either natural persons or moral ones), other content simply generates collective harm. In the latter case, despite the severity of the harm and the ampleness of the negative impact, no specific victims can be singled out. Proper "fake news" in a strict definition does very frequently fall under this last category. The spread of manipulated, false, fabricated, or misleading content has a demolishing impact on collective trust, and on the ability of a society to create a common dialogue on shared accurate facts. It undermines the value of objective facts, delegitimizes experts' voices and authoritative institutions, and radicalizes confronting stances in a context of chaos and confusion.[11] "Fake news" would then be representing a variety of mis and disinformation vehicles. Repercussions are alarming, but specific quantifiable damage might not be proven, and identifiable injured persons might not be located.

The above-stressed characteristics of misinformation and disinformation vehicles have a very relevant effect on the legal analysis and a direct impact on the components of the liability machinery. If the damage is diffused, it will be questionable who is entitled to claim compensation, if

---

[10] Claire Wardle, *Fake News. It's Complicated*, FIRST DRAFT, https://medium.com/1st-draft/fake-news-its-complicated-d0f773766c79 (last visited Sept. 12, 2019).

[11] *See* U.N. Special Rapporteur on Freedom of Opinion and Expression, *Joint Declaration on Freedom of Expression and "Fake News", Disinformation and Propaganda*, http://www.ohchr.org/SP/NewsEvents/Pages/DisplayNews.aspx?NewsID=21287&LangID=E (last visited Sept. 12, 2019). The authors point out propaganda in legacy and social media is fueled by both States and non-State actors, and the various harms to which they may be a contributing factor.

any. If the harm is a devaluation of collective trust, it might be difficult to quantify damages. Damage to public interest is probably the most feared and destabilizing impact of the spread of falsity, but it may not be compensable under the coordinates of the civil liability regime. If the liability system is founded on a notice-based scheme, it might be discussed who is expected to report and allege legitimate interests to act. Consequently, should fake-news-combating response be addressed to intermediaries and articulated by a liability-oriented discourse, all these considerations must be taken into account to devise the model.

Given the previous analysis, it can be sustained that intermediaries could have to face three categories of content: illegal content, harmful content, and false content. Although in certain cases these categories can coincide, they must be treated and approached as distinct and separate ones. Illegal content and false news might not produce actual damage, whereas harmful content could be entirely accurate and truthful, and might be fully licit and legitimate. Therefore, illegality, harmfulness, and falsity constitute different factual spheres that require appropraite responses. Hence, preventive measures, reparation and compensation mechanisms contrived to combat the effects of illegality and harmfulness are not equally effective to counter falsity. False content adds intricacies in the detecting and assessing phase and in the ascertaining of damages. The incontrollable spread of "fake news", the penetrating impact of misinformation in society's stability, and the devastating effects on trust has crudely revealed such a gap, the lack of preventive and protective measures against falsity.

Yet, unlike illegal and harmful content, setting a fair balance of conflicting rights and interests at stake in case of false content is more complex and unstable. As the contours of false content are blurred, and the potential harm is – albeit severe and massive – highly diffuse, freedom of expression becomes especially vulnerable to any ill-advised restrictive or banning decision.[12]

Consequently, this Article is exclusively focused on the role of intermediaries in the sphere of falsity and the advisability of a liability-based regulatory strategy to counter misinformation to that extent. European bodies have claimed a higher responsibility of intermediaries and platforms in tackling illegal and harmful content.[13] Likewise, the perceived paradigm shift of intermediaries' liability regime, as further analyzed in this Article (*infra* Part III), would work for and extend essentially over illegal

[12] *Id*.

[13] *See* European Parliament 2016/2274 (INI), 15 June 2017, P8_TA(2017)0272.

and harmful content.[14] However, both policy proposals – higher responsibility and civil liability – have to be tested within a regulatory strategical context to counter "fake news." Implications, consequences, and intricacies will undoubtedly be different.

Within such a phenomenal delimitation, my proposal is that fake news has two dimensions: the factual one that determines its veracity, and the social one that is based on perception. Falsity perception and misleading effect may be determined by the substance or by the form. As a matter of fact, true content can be presented in a way likely to mislead, confuse, or trigger misinterpretation. Whereas the former requires an objective test and needs a credibility reference endorsed by a trusted third party, the latter is diffuse and subjective and depends on community perception.

In regard to the factual dimension, "fake news" necessarily embraces a degree of falsity. Whereas veracity presents only a face, falsity ranges a wide spectrum of inveracities. Falsity in any degree is assessed on an objective basis. Although intent is relevant to distinguish disinformation, as a deliberate act from misinformation, as an inadvertent omission or unintentional sharing of false information, as well as to determine the illegality of the act or even the compensational damages, it will be ignored for the purposes of defining "false content" in a liability scheme for intermediaries. Whether digital intermediaries and platforms decide to implement proactive mechanisms to detect false content and remove it, the intentional factor in the origination or in the dissemination should not be incorporated in the process, as it is essentially irrelevant for the limited purpose of the detection. Nonetheless, intermediaries may use objective factors as a proxy for intentionality such as repetitive dissemination of false content, volume of spread "fake news," or other circumstances revealing an organized and systematic structure to misinform. Likewise, intermediaries may calibrate the severity of penalties laid down in the platform's internal policy to rigorously respond to intentional massive spread of false content with the most radical sanction of expulsion from users' community, closing of account, or disabling of access.[15]

---

[14] This self-regulation strategy relies on voluntary cooperation of the biggest digital platforms to combat the spread of illegal hate speech in Europe. European Commission Press Release IP/16/1937, European Commission and IT Companies Announce Code of Conduct on Illegal Online Hate Speech (May 31, 2016). But see European Parliament, stating: "The liability rules for online platforms should allow the tackling of issues related to illegal content and goods in an efficient manner, for instance by applying due diligence while maintaining a balanced and innovation-friendly approach." European Parliament, *supra* note 13, at ¶ 34.

[15] *See* Teresa Rodríguez de las Heras Ballell**,** *The Legal Anatomy of Electronic Platforms: A Prior Study to Assess the Need of a Law of Platforms in the EU*, 3 No. 1 ITALIAN L.J. 149, 149-76 (2017).

Strategies implemented by global platforms and intermediaries based on the reliance upon fact-checkers, the verification by authoritative sources, and even the devising of report systems[16] are indeed directed to reinforce the veracity dimension.

The second dimension of "fake news" is a social one. The gravity of the problem created by systematic misinformation is not only caused by the falsity of the content, but also principally exacerbated by its incontrollable penetration, and its pervasive expansion producing a deafening "noise," silencing authoritative voices, and concealing fact-checked content. The risk of "fake news" is that it becomes widely credible. Factors other than the veracity of facts are able to generate a perception of credibility. Misallocated or wrongly placed trust might have a more negative effect than distrust.[17] To attenuate this wrong perception of credibility, fact checking is frequently ineffective, as content is infused by other credibility indicia based on popularity. Compared to the widely shared misinformation, fact-checking response might not gain sufficient relevance and even, perceived as a minority opinion, it dilutes its credence.

Popularity – number of likes, retweets, followers – as proxy for credibility, veracity or relevance is the expression of a deeper vulnerability of our society: the tyranny of quantification. Lists, ratings, rankings, priority orders, numbers offer today a safer way to understand an uncertain and complex world. Certainly, quantification helps decision-making. The digital revolution has drastically reduced the cost to count, quantify, rank, and rate.[18] This obsessive wish to measure every aspect of human behavior along with a blinded confidence in the value of quantification to order the world, to represent quality, to quantify credibility, to objectivize every attribute, and to beat any threat of subjectivism, lead to a "omnimetric society." Quantification suggests objectivity, evokes neutrality, and enables comparability under a very simple successive order. As a consequence, it is extremely ineffective to combat popularity-based credibility without exploiting the same power of numbers. It is irrelevant whether the rectification content is reliable, well founded, and factually objective, insofar as it is unable to gain the merits and the credence that a massive spread provides. For a simplistic understanding of an omnimetric society, minority means irrelevance, popularity means credibility, majority means veracity.

---

[16] "[T]he importance of taking action against the dissemination of fake news; calls on the online platforms provide users with tools to denounce fake news in such a way that other users can be informed that the veracity of the content …." European Parliament, *supra* note 13, at ¶ 35.

[17] *See generally* Russell Hardin, *Distrust*, 81 B.U. L. REV. 495 (2001).

[18] *See* Bruno S. Frey, *Omnimetrics and Awards*, 2017 CESifo Working Papers, 1, 3.

The omnimetric nature of modern society is aggravated by another sociological component: the proliferation of peer-based structures. Bourgeoning sharing economy, collective creation, crowdfunding, or reputational rating mechanisms are rooted in that community-based approach. Trust also relies on peers. The social consequences of that perspective directly impact the dimensions of the "fake news" phenomenon and make its containment more difficult. Peer-determined "truth" is prioritized over traditional authoritative sources that become less visible or even less credible.

Therefore, these two features of modern society, intensified by digitalization, exacerbate the intricacies of the "fake news" problem and debilitate the effectiveness of any fighting strategy against it. Apparently, the only objective truth is that which can be quantified, and the only trust is that which is shared.

The role of intermediaries is critical in this second dimension of "fake news." Intermediaries and platforms fuel credibility perception by providing accessibility, visibility, and virality mechanisms to user-generated/distributed content. From that perspective, intermediaries and platforms represent a critical component in the misinformation machinery. It is undeniable that intermediaries and platform provide the infrastructure for the dissemination, create an environment suited to ignite credibility perception, and exacerbate the massive effects of false news. Nevertheless, it is highly questionable that such an infrastructural contribution should lead to any level of liability. More interestingly, it is even more uncertain how platforms should act to contain virality, counter popularity-measured credibility, and combat with objectivity and fact checking oversized perception of trustfulness. A regulatory model that happens to dislocate incentives may trigger an overly cautious reaction of intermediaries and platforms, for fear of the liability consequences, likely to distort the free flow of ideas in the digital world, to encroach upon freedom of expression, and to fragment the information scene into biased "ideological silos."

III. INTERMEDIARY LIABILITY PARADIGM IN CONTEXT

The crucial role of digital intermediaries for a well-functioning digital market and a flourishing digital society was clearly perceived at the very early stage by national and regional regulators, in particular, the United States of America and European Union legislators. The need to ensure a proper and effective performance of intermediary activities became soon an imperative policy concern. To that end, an allocation of risks and incentives should be achieved. The formulation of intermediary liability-exempting rules ("safe harbor" provisions) has been the widespread common

regulatory response to articulate that fundamental policy. As a matter of fact, safe harbor provisions for electronic intermediary service providers represent a key common policy in worldwide internet regulation.[19] Although there are disparities in scope, applicable conditions, and effects, intermediary liability rules have been extensively incorporated into most jurisdictions.[20] Inspired by the U.S. legal precedent (essentially, Section 512 of the Digital Millennium Copyright Act),[21] intermediary liability exemptions have been the central axis of electronic commerce and the information society services legal framework in Europe from the outset.[22]

To date, it has been a commonplace assumption that the intermediary non-liability paradigm has accelerated the expansion and consolidation of the digital environment.

Intermediary liability regime pivots on two key tenets and a legal concept of service providers to define the scope of application. First, the ban of imposing a general obligation to monitor on service providers.[23] Second, a knowledge-and-take-down system.[24] Both tenets constitute the pillars of a negligence-based liability system. Accordingly, those service providers falling under the "safe harbor" provisions are exempted from any general obligation to proactively monitor or filter the information they transmit or the content they store, copy, or search, or to actively seek facts, indicia, or circumstances that might signal illegality.[25] Hence, in the absence of general duties to monitor, service providers must act only upon obtaining knowledge or awareness of the illegal information or activity; then, they have to proceed expeditiously to remove, or disable access to that content or service. Knowledge is essentially obtained from notice mechanisms implemented by service providers to enable users to flag, denounce, or report infringing content, unlawful activities, or any other illegal material. In sum, intermediary service providers do not have any duty to monitor, on a general and proactive basis, any content they transmit,

---

[19] *See generally* World Intermediary Liability Map, http://cyberlaw.stanford.edu/our-work/projects/world-intermediary-liability-map-wilmap (illustrating and providing detailed information about the global regulatory response to intermediary liability) (last visited May 24, 2020).

[20] *Id*.

[21] 17 U.S.C. § 512 (2010).

[22] Council Directive 2000/31 art. 1, 2000 O.J. (L178) 1-16. [hereinafter *Directive on Electronic Commerce*].

[23] *Id*. at art. 15.

[24] *Id*. at art. 14.

[25] *Directive on Electronic Commerce*, *supra* note 22 at Recital 47 in relation to art. 15. Article 15.1 states: "Member States shall not impose a general obligation providers, when providing the services covered by Articles 12, 13 and 14, to monitor the information which they transmit or store, nor a general obligation actively to seek facts or circumstances indicating illegal activity . . . ."

store, copy, or search and they are called to act only when and to the extent they have knowledge or awareness of illegal information or activities.

On the grounds of the above-described two key tenets (no duty to monitor and knowledge-based take-down obligations), intermediary liability regime applies to intermediary service providers, whereas content and service providers other than the latter ones are subject to general liability rules. The underlying assumption is then that intermediary service providers do neither control nor be aware of any information, content, or activity that they transmit, store, search, or anyhow enable. The rationale behind the description of service providers falling under the safe harbor provisions is that they perform a passive, technical and purely instrumental role. Paradigmatically, they provide access, transmission, caching, hosting, or searching services.

However, after almost two decades of evolution, the digital scene has changed considerably. Therefore, the intermediary liability paradigm has been shaken and the continuation of intermediary liability in its current form has come into question. The confluence of several trends has precipitated the debate on the need for a paradigm shift in the intermediary liability system.

First, the transformation of the Digital Economy into the Platform Economy has raised the question about the legal concept of intermediary service providers and therefore the delimitation of the scope of application of the safe harbor provisions.[26]

Second, some ongoing regulatory proposals (namely, under the EU Digital Single Market Strategy)[27] and judicial decisions in different jurisdictions seem to veer toward increasing proactive monitoring and filtering obligations and pave the path for a progressive eroding of the "no monitoring obligations" tenet.

Third, intermediaries play a central role in prevention, civil protection of rights, and voluntary enforcement in the framework of a conspicuous regulatory strategy to promote private ordering increasingly adopted and deployed by governments to face digital challenges[28] – particularly visible

---

[26] *See generally* Teresa Rodríguez de las Heras Ballell, *El régimen jurídico de los Mercados Electrónicos Cerrados (e-marketplaces)* (Marcial Pons, 2006).

[27] European Commission, *A Digital Single Market Strategy for Europe*, COM (2015) 192 final (May 6, 2015).

[28] The governing abilities of platforms mark the approach of the publication *Platform Regulations: How Platforms are Regulated and How They Regulate Us*, Official Outcome of the UN IGF Dynamic Coalition on Platform Responsibility, United Nations Internet Governance Forum Geneva, December 2017 Luca Belli and Nicolo Zingales (eds.), https://juliareda.eu/wp-content/uploads/2019/09/Reda2017_Platform-regulations-how-platforms-are-regulated-and-how-they-regulate-us3.pdf. About the central economic and societal role of platforms, see Alexandre de Streel & Miriam Buiten, Marting Peitz, CERRE Report, *Liability of online hosting platforms.*

in the Digital Single Market for EU.[29] Prevention, control, and enforcement tasks are gradually transferred to and allocated on intermediaries.[30] In that context, platforms and intermediaries have implemented monitoring mechanisms and automatic filtering systems on a voluntary basis to counter fake news, hate speech, copyright infringement, and illegal content addressed to minors.

Should exceptionalism end?, September 2018, https://www.cerre.eu/sites/cerre/files/180912_CERRE_LiabilityPlatforms_Final_0.pdf

[29] Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, *Online Platforms and the Digital Single Market Opportunities and Challenges for Europe*, SWD (2016) 172 final, COM (2016) 288 final Brussels, 25.5.2016, at 3. *See also* the Recitals of the Regulation (EU) 2019/1150 of the European Parliament and of the Council of 20 June 2019 on promoting fairness and transparency for business users of online intermediation services, OJ L 186/57, 11.7.2019, explaining the relevant role of platforms and online intermediation services providers.

[30] The most illustrative examples of this trend are:

First, Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC, OJ L 130, 17.5.2019, p. 92-125, art. 17.4:

> If no authorisation is granted, online content-sharing service providers shall be liable for unauthorised acts of communication to the public, including making available to the public, of copyright-protected works and other subject matter, unless the service providers demonstrate that they have: (a) made best efforts to obtain an authorisation, 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).

Second, Directive (EU) 2018/1808 of the European Parliament and of the Council of 14 November 2018 amending Directive 2010/13/EU on the coordination of certain provisions laid down by law, regulation or administrative action in Member States concerning the provision of audiovisual media services (Audiovisual Media Services Directive) in view of changing market realities, OJ L 303, 28.11.2018, p. 69-92, art. 28b:

> Without prejudice to Articles 12 to 15 of Directive 2000/31/EC, Member States shall ensure that video-sharing platform providers under their jurisdiction take appropriate measures to protect: (a) minors from programmes, user-generated videos and audiovisual commercial communications which may impair their physical, mental or moral development in accordance with Article 6a(1); (b) the general public from programmes, user-generated videos and audiovisual commercial communications containing incitement to violence or hatred directed against a group of persons or a member of a group based on any of the grounds referred to in Article 21 of the Charter; (c) the general public from programmes, user-generated videos and audiovisual commercial communications containing content the dissemination of which constitutes an activity which is a criminal offence under Union law, namely public provocation to commit a terrorist offence as set out in Article 5 of Directive (EU) 2017/541, offences concerning child pornography as set out in Article 5(4) of Directive 2011/93/EU of the European Parliament and of the Council (*) and offences concerning racism and xenophobia as set out in Article 1 of Framework Decision 2008/913/JHA.

These trends reveal that an intermediary liability paradigm is under consideration. Nevertheless, alternative models are not yet well defined. The implications of new models for digital society, the protection of rights, internet neutrality, and the preservation of trust are significant. The current liability model embeds a fair balance between freedom of information, protection of rights, and intermediaries' freedom to conduct their business.[31] A paradigm shift of liability would challenge that balance. Therefore, a debate to reach public consensus on the model of digital society is necessary to have a proper understanding of state-of-the-art technology and its future possibilities. Additionally, serious attempts to produce harmonized rules are imperative.

A. *New Challenges and Orientations: Intermediary Liability Regime Under Consideration*

Today, the paradigms of intermediary liability face significant challenges. While the digital economy evolves and society becomes increasingly digital, the context, the players, and the problems to address under the safe harbor regime have also been transformed. These transformative forces and challenging trends have an impact on the basis of the established paradigm. The stability and soundness of the paradigm and the flexibility of the liability rules to adapt to the new circumstances then come into question. More interestingly, it is discussed whether liability rules in their current form are playing the role attributed thereto.

In this section, the three main challenges and new orientations that were identified earlier will be discussed and further elaborated on as main triggers of the debate for reshaping the liability regime. Fist, the emergence and rapid proliferation of platform operators. Second, the escalation in number, severity and intensity of harming situations. Third, the promotion of private ordering and voluntary enforcement.

1. The Transformation of Digital Economy into a Platform Economy

Electronic platforms are the dominant organizational model[32] for economic activities, social networking, and emerging businesses in today's digital society and have transformed social, political, public, and

[31] Case C-360/10, Belgische Vereniging van Auteurs, Componisten en Uitgevers CVBA (SABAM) v. Netlog NV, 2012; Case C-236/08, Google France SARL and Google Inc. v. Louis Vuitton Malletier SA, 2008; Case C-237/08, Google France SARL v. Viaticum SA and Luteciel SARL, 2009; Case- C-238/08, Google France SARL v. Centre national de recherche en relations humaines (CNRRH) SARL, 2010.

[32] Thomas W. Malone, Joanne Yates & Robert I. Benjamin. *Electronic Markets and Electronic Hierarchies*, 30(6) Communications of the ACM, at 484-97 (1987).

educational contexts. The emergence and increasing popularity of disruptive models, such as sharing-based economy, crowdfunding, or fintech variants, have not only been made possible but greatly stimulated by platform-based solutions. The scaling-up presence of platforms in the digital economy and their growing market power has unveiled a visible disruptive effect on varied angles. Social, economic, and legal disruptions are perceptible, or certainly expected to explode soon. Their social and economic disrupting potential is clearly observed in the transformation of social relationships, market structures, and economic paradigms induced by platform-based emerging models (sharing-driven business models,[33] Fintech variants,[34] crowdfunding[35]). Along with these noticeable social and economic disruptions, the platform model is also proving to be legally disruptive. Their self-regulation power linked to an intense centripetal force that accelerates concentration, the critical role likely to be played by platform operators in prevention and civil enforcement, and the trust-generating capacity of platforms in a digital society have started to strongly attract an increasing interest of regulators and supervisors. With the issuance of public consultations and special reports, and the work undertaken by research groups,[36] first moves have been made at the EU level[37] and in some national jurisdictions[38] showing interest in platform economy.

---

[33] Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, *A European agenda for the collaborative economy,* SWD 184 final (2016).

[34] World Economic Forum, *Beyond Fintech: A Pragmatic Assessment Of Disruptive Potential In Financial Services* (Aug. 22, 2017), https://www.weforum.org/reports/beyond-fintech-a-pragmatic-assessment-of-disruptive-potential-in-financial-services (last visited Sept. 12, 2019); Teresa Rodríguez de las Heras Ballell, *Challenges of Fintech to Financial Regulatory Strategies* (Marcial Pons, 2019).

[35] Teresa Rodríguez de las Heras Ballell, *A Comparative Analysis of Crowdfunding Rules in the EU and U.S.*, Stanford TTLF Working Paper Series*,* Working Paper num. 28, https://law.stanford.edu/publications/no-28-a-comparative-analysis-of-crowdfunding-rules-in-the-eu-and-u-s.

[36] Christoph Busch et al., *Research group on the Law of Digital Services. Discussion Draft of a Directive on Online Intermediary Platforms*, 5 EuCML 164-69 (Apr. 2016). The Project is today a European Law Institute (ELI) Project (Model Rules on Online Intermediary Platforms) approved by the ELI Council on September 7, 2016. The author of this Paper joined the ELI Project Team in 2016 and participated in all Project meetings in Krakow (Jan. 2017), Osnabruck (Mar. 2017) and Berlin (Nov. 2017). Project Rapporteurs are BUSCH, Christoph (Univ, of Osnabrück); DANNEMANN, Gerhard (Humboldt Univ. Berlin); SCHULTE-NÖLKE, Hans (Univ. of Osnabrück and Nijmegen); WIEWIOROWSKA-DOMAGALSKA, Aneta (Univ. of Osnabrück); ZOLL, Fryderyk (Univ. of Krakow and Osnabrück). The opinions expressed in this paper are personal views of the author and do not necessarily represent the Project Team's views.

[37] Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, *Online Platforms and the Digital Single Market Opportunities and Challenges for Europe,* SWD (2016) 172 final, COM (2016) 288 final Brussels (2016).

Optimal liability regime for platform operators is a critical policy concern underlying all these legislative and pre-legislative initiatives. Whether platform operators act as pure intermediaries protected by liability rules, or, in contrast, they should be requested or encouraged to the adoption of proactive measures is a dilemma that finds an effective breeding ground in an expanding platform economy.

As far as the legal framework for the provision of online services is concerned, electronic platform operators can be deemed intermediary service providers (ISPs) in relation to content, activities and behaviours, published, transmitted or performed by their users. Accordingly, a safe harbour regime would be applicable to delimit their liability – articles 12-15 *Directive on Electronic Commerce* with direct antecedents in U.S. legal model divided into the Communications Decency Act of 1996 included as Part V of Telecommunications Act (Pub. L. 104-104, 110 Stat. 56 (codified at 47 U.S.C. § 230) and the Digital Millennium Copyright Act, Pub. L. No. 105-304, 112 Stat 2860 (28 Oct. 1998) (codified at 17 U.S.C. § 512). The European Court of Justice confirmed that assertion when expressly held in *L'Oréal SA and Others* v. *eBay International AG and Others*[39]:

> Article 14(1) of Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market ("Directive on Electronic Commerce") must be interpreted as applying to the operator of an online marketplace where that operator has not played an active role allowing it to have knowledge or control of the data stored.

However, the above-cited decision of the Court differs from the most recent opinion held in the Uber Spain case. Although the opinion is disputable to a certain extent, the European Court of Justice, in *Asociación Profesional Elite Taxi v Uber Systems Spain, SL*,[40] follows the Advocate General's Opinion.[41] As Advocate General Szpunar had proposed, the Court

---

[38] In France, three regulations (decrees) have been adopted to reinforce the transparency and the loyalty of platforms: *Décret N° 2017-1434 du 29 Septembre 2017 Relatif Aux Obligations D'information Des Opérateurs De Plateformes Numériques*, JORF n°0233 du 5 octobre 2017; *Décret N° 2017-1435 du 29 Septembre 2017 Relatif à La Fixation D'un Seuil De Connexions à Partir Duquel Les Opérateurs De Plateformes En Ligne élaborent Et Diffusent Des Bonnes Pratiques Pour Renforcer La Loyauté, La Clarté Et La Transparence Des Informations Transmises Aux Consommateurs,* JORF n°0233 du 5 octobre 2017; *Décret N° 2017-1436 Du 29 Septembre 2017 Relatif Aux Obligations D'information Relatives Aux Avis En Ligne De Consommateurs,* JORF n°0233 du 5 octobre 2017.

[39] Case C-324/09, L'Oréal SA and Others v. eBay International AG, 2011.

[40] Case C-434/15. ECLI identifier: ECLI:EU:C:2017:981, *Asociación Profesional Elite Taxi v Uber Systems Spain, SL,* 2017. Text available at http://curia.europa.eu/juris/celex.jsf?celex=62015CJ0434&lang1=es&type=TXT&ancre= (last visited Sept. 12, 2019).

[41] *Id*.

understands that the service offered by Uber cannot be classified as an "information society service," but it amounts to the organization and management of a comprehensive system for on-demand urban transport. Accordingly, the Court aligns with the Advocate General's arguments and proposes that the service offered by Uber as the platform operator must be classified as a "service in the field of transport." Thus, the separation line between operator and users providing the service dilutes, and the platform operator becomes a direct supplier instead of a provider of intermediary services.

Nonetheless, electronic platforms are contract-based. Such a contractual infrastructure defines the liability regime and indeed allocates duties and liabilities between operators and platform's members. Since "safe harbor" regime is based on lack of knowledge and lack of control, operators manage to preserve their position with a right (but not an obligation) to monitor and supervise so as to enhance confidence without exposing themselves to liability risks.[42] Concurrently, the assumption that to a certain extent the operator of supervisory, sanctioning, or reviewing functions for the purposes of managing the platform may frontally question the assumption that the operator is not playing "an active role" in the meaning of the Court's decision. Therefore, the application of "safe harbor" provisions to platforms may require a further analysis of the functional and operational platform models to specifically assess the nature and the extension of its role.

Furthermore, the analysis becomes more complex due to the fact that there is not a comprehensive, general regulation on platforms. Sector-specific regulations have been adopted at different levels to tackle issues arising from sectorial platforms such as crowdfunding platforms,[43] Alternative Trading Systems[44]/Multilateral Negotiating Systems or

[42] Teresa Rodríguez de las Heras Ballell, *La responsabilidad de las plataformas: Alcance, límites y estrategias*, 369-393 (2006).

[43] *See* European Commission Staff Working Document, Crowdfunding in the EU Capital Markets Union (May 3, 2016), http://ec.europa.eu/finance/general-policy/docs/crowdfunding/160428-crowdfunding-study_en.pdf (last visited Sept. 12, 2019) (enclosing a comparative table of domestic bespoken regimes adopted by Member States). U.S. Rules on Crowdfunding are essentially comprised of the legal provisions of Title III "Capital Raising Online While Deterring Fraud and Unethical Non-Disclosure Act of 2012" of *Jumpstart Our Business Startups Act, JOBS Act* (Apr. 5, 2012), which is implemented by the final rules adopted by the Securities Exchange Commission (SEC) under the Securities Act of 1933 and the Securities Exchange Act of 1934 (17 CFR Parts 200, 227, 232, 239, 240, 249, 269, and 274 [Release Nos. 33-9974; 34-76324; File No. S7-09-13] RIN 3235-AL37). The final rules and forms went into effect May 16, 2016, except that instruction 3 adding part 227 and instruction 14 amending Form ID went into effect January 29, 2016.

[44] Jonathan R. Macey & Maureen O'Hara. *Regulating Exchanges and Alternative Trading Systems: A Law and Economics Perspective*, 28 J. LEGAL STUD. 17 (1999).

Facilities,[45] or the most recent timid, irregular, and to some degree erratic regulatory actions on economy-sharing models.[46] Given their sector-specific scope, these rules do not embrace platforms as a whole, but solely address special features of those platforms falling under their scope of application and for the purposes of protecting certain interests – market stability, transparency, investors' interests, systemic risk, consumer rights, tax collection, and fraud. Under these disparate approaches, platform operators may be required to comply with certain specific duties within the relevant sectoral sphere.

In sum, the transformation of digital economy into a platform economy obscures the binomial classification of service providers – intermediary service providers v. content and general (non-intermediary) service providers – and complicates the application of the intermediary liability regime to the new players (platform operators).

2. The Escalation in Number, Severity and Intensity of Harming Situations and the Role of Intermediaries

The recent controversy about "fake news" and the use of social media for spreading hate speech, violence, or extremist ideologies (e.g., white supremacist, neo-Nazis, alt-right groups) has put intermediaries and platforms in a quandary. Some popular platforms have decided to react, even compromising their neutrality, by removing content, closing accounts, or publicly denying service to certain users, and implementing mechanisms to automatically identify false news. Certainly, none of these situations are new, but they have ultimately exploded with unprecedented virulence arousing social alarm, attracting regulatory attention due to severe policy concerns, and invading international political discourse and diplomacy.[47]

---

[45] For instance, in the European Union, Art.4 (15) MiFID defined "Multilateral trading facility (MTF)" as "a multilateral system, operated by an investment firm or a market operator, which brings together multiple third-party buying and selling interests in financial instruments – in the system and in accordance with non-discretionary rules – in a way that results in a contract in accordance with the provisions of Title II."

[46] *See* Guido Smorto, *Critical Assessment of European Agenda for the Collaborative Economy, on behalf of European Parliament. In-Depth Analysis for the IMCO Committee*. vol. IP/A/IMCO/2016-10, at 9 (2016).

[47] *See* Jean-Baptiste Jeangène Vilmer, Alexandre Escorcia, Marine Guillaume & Janaina Herrera, *Information Manipulation. A Challenge for Our Democracies. A report by the Policy Planning Staff (CAPS, Ministry for Europe and Foreign Affairs) and the Institute for Strategic Research (IRSEM, Ministry for the Armed Forces), (*2018); *See also Report on Initiatives to Counter Fake News in Selected Countries: Argentina, Brazil, Canada, China, Egypt, France, Germany, Israel, Japan, Kenya, Malaysia, Nicaragua, Russia, Sweden, United Kingdom,* Apr. 2019, Law Library of Congress, Global Legal Research Directorate, http://www.law.gov; *see also* Chris Marsden & Trisha Meyer, *Regulating disinformation with artificial intelligence,* Study – Panel for the Future of Science and Technology, European Parliamentary Research Service (Mar.

The array of proactive policies and strategies implemented by platforms and intermediaries in response to such a hostile, menacing context, raises important challenges. First, it means a progressive departure from neutrality. The uncertain consequences of a trip towards a market of biased players are yet unknown. Second, it will require recalibrating liability rules where intermediaries decide to select, assess, remove, and actively monitor. Third, it will very likely head to arbitrage and "platform shopping" in a world without a uniform response yet.

3. The Promotion of Private Ordering and Voluntary Enforcement Mechanisms

Finally, States have realized how weak and ineffective their traditional preventive and enforcement legal machinery is in the digital scene. Accordingly, a progressive and timid but revealing "conveyance" of powers and responsibilities in prevention and civil enforcement from public bodies to platform operators is increasingly visible. The premise inspiring such a conspicuous transfer is that platforms are best situated to detect infringement promptly prevents damages to rights or interests, and effectively enforces rights with contractual-based mechanisms. The collaboration of platform operators and intermediaries enhances the effectiveness of legal enforcement, but also raises legal concerns.

In this context of promotion of private ordering[48] and voluntary enforcement, the intermediary liability paradigm is under consideration. Liability regime is critical to rightly allocate incentives and align interests with policy goals. Should policy goals change to seek greater involvement of intermediaries and platforms in prevention and enforcement, the liability regime might be reshaped.

B. *Signs of Change? – Digital Intermediaries in a Quandary*

A safe harbor-based liability regime for digital intermediaries has remained as a solid foundational pillar of information society services and an electronic commerce legal framework for almost two decades. Its value in reconciling conflicting interests at stake were acknowledged and recognized by case law and legislative policy decisions. The expansion of digital activities and, certainly, the scaling-up emergence of platform-based models in all their variants (collaborative economy, fintech, crowdfunding, social networks, e-marketplaces) have been deeply supported and

2019), https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624279/EPRS_STU(2019)624279_EN.pdf.

[48] Steven Schwarcz, *Private Ordering*, 97 NW. U.L. REV. 319, 319-50 (2002).

encouraged by the intermediary liability paradigm as devised in its original form.

Nonetheless, a dramatic transformation of digital context experienced over the years threatens to destabilize the solidity of current regimen and dilute the rationale behind the liability paradigm. The current system is shaken, and several signs of change are already perceptible. Nevertheless, prospective models resulting from a potential paradigm shift are not clearly delineated yet. The consequences of such a shift will be highly relevant for our digital society and should not be ignored.

Under this section, some perceived signs of change[49] will be exposed and analyzed to envisage afterwards possible alternative models for intermediary liability and discuss their implications and expected outcomes below (*infra* Part IV).

First, recent case law in multiple jurisdictions shows a progressive distancing from intermediary liability tenets and upholds proactive monitoring obligations on intermediaries in relation to a wide array of infringements and illegal activities.[50] Although that departing trend is not consistent and contrasts with other decisions reinforcing the tenets of the current liability paradigm,[51] it depicts a cracked picture.

It is increasingly visible that there is a jurisdictional discourse that stresses the concern about the alarming threat posed by digital means to certain rights, especially intellectual property infringement, privacy violations, defamation and hate speech.[52] Along the lines of that narrative, the digital environment would create unprecedented risks causing massive and persistent damages, unstoppable infringements, and a viral negative impact on rights. Such a reasoning could be paving the path toward a veering from negligence-based liability to strict liability on the grounds of *cuius commode eius et incommoda* principle and would endorse an imposition of monitoring obligations on intermediaries. In that regard, these decisions sustain that insofar as providers obtain economic benefits

---

[49] Giancarlo Frosio, *From horizontal to vertical: an intermediary liability earthquake in Europe*, 12 J. INTELLECTUAL PROP. LAW & PRAC. 565, 565-75 (2016).

[50] Giancarlo Frosio, *The Death of 'No Monitoring Obligations': A Story of Untameable Monsters*, 8(3) J. INTELLECTUAL PROP., INFO. TECH. & E-COMMERCE L. 212 (2017).

[51] *See* Rodriguez M. Belen c/Google y Otro s/ daños y perjuicios, R.522.XLIX, Sup. Ct. of Arg., (Oct. 29, 2014); Reti Televisive Italiane S.p.A. (RTI) v. Yahoo! Italia S.r.l. (Yahoo!) et al, N RG 3821/2011, Milan Ct. App., (Jan. 7, 2015); TF1 v. DailyMotion, Paris Ct. App., (Dec. 2, 2014).

[52] S.T.J., SP No. 1.306.157, Relator: Des. Luis Felipe Salomão, 24.03.2014, 1, Superior Tribunal de Justiça Jurisprudẽncia [S.T.J.J.] (Braz.); Universal City Studios v. Corley, 273 F.3d 429 (2d Cir. 2001); Delfi AS v. Estonia, 2015-II Eur. Ct. H.R. (2015).

(advertisements, mainly) should contribute in blocking or delisting infringing material.[53]

With such arguments, some above-referred judicial decisions held, apparently, under the umbrella of the Recital 40 of the *Directive on Electronic Commerce*, that intermediaries and platforms have the obligation not only to remove infringing material upon notice, but also to prevent repetition of further infringements adopting monitoring measures.[54] Interestingly, another decision links the specific duty to monitor the platforms to those content that prove to be popular, attracting special interests of users with a number of views, visits, or downloads.[55] Accordingly, such indicia of popularity should trigger the duty of the platform to examine the legal status of those contents and, if necessary, to protect it from infringement. That curious delimitation of the duty seems to be inspired by the acknowledgement of a special greater responsibility of platforms and intermediaries to protect rights and interests, employed by other courts as well to declare the reasonableness of proactive monitoring obligations on "new generation" hosting services.[56]

Second, several legislative actions, in particular within the framework of the EU Digital Single Market scheme, apparently point a shift of tendency towards the introduction of filtering and monitoring obligations on

---

[53] Bundesgerichtshof [BGH] [Federal Court of Justice] Aug. 15, 2013, I ZR 79/12, http://juris.bundesgerichtshof.de/cgi-bin/rechtsprechung/document.py?Gericht=bgh&Art=en&sid=64b5038f0e7c75357e0d9a484f2919e9&nr=65240&pos=0&anz=1.

[54] Giancarlo Frosio, *The Death of 'No Monitoring Obligations': A Story of Untameable Monsters*, 8(3) J. INTELLECTUAL PROP., INFO. TECH. & E-COMMERCE L. 212, 204 (2017). First, in *Delta TV v. Google and Youtube*, n. 1928/2017 n. 38112/2013, ORDINARY TRIBUNAL OF TURIN (Jan. 25, 2017), https://www.laleggepertutti.it/wp-content/uploads/2017/04/sentenza_1928_17.pdf, the Court held that "(d)eve dunque affermarsi che per la piattaforma You Tube (essendo ciò pienamente possibile dal punto di vista tecnico, sebbene con un minimo margine di possibilità di insuccesso) sussiste un vero e proprio obbligo giuridico di impedire nuovi caricamenti di video già segnalati come violazione del diritto d'autore . . ." – "(a)ssuming that it is fully possible from a technical point of view, although with a minimum margin for failure, there subsists on YouTube an actual legal obligation to prevent further uploads of videos already flagged as infringing of third-party copyrights" (translation by the author). Second, in the Brazilian decision *Google Brazil v Dafra*, Special Appeal No. 1306157/SP, Superior Court of Justice, Fourth Panel, (Mar. 24, 2014), https://wilmap.law.stanford.edu/news/brazilian-supreme-court-found-google-liable-videos-parodying-dafras-commercials, the Court held that Google had the duty to "certain proactive control" over future uploads, albeit accepting that there are some limitations to that proactive monitoring. Third, the French decision *APC et al v. Google, Microsoft, Yahoo!, Bouygues et Al*, Cour d'Appel Paris, n°040/2016 (Mar. 16 2016), https://juriscom.net/wp-content/uploads/2016/03/16032016caparis.pdf, confirms the measures imposed on the intermediaries aimed to proactively expunge search results from any link to the same websites.

[55] Baidu v. Register.com, 760 F. Supp. 2d 312 (S.D.N.Y. 2010).

[56] Trib. 23 giugno 2014, n. 38113, Foro. It (It.).

intermediaries in some areas and a general support of voluntary prevention and enforcement mechanisms. In order to prevent copyright infringement,[57] the *Directive on Copyright for a Digital Single Market*[58] and the *Audiovisual Media Services Directive*[59] would encourage the adoption of effective content recognition technologies to prevent the availability of infringing content.

Apparently, these cooperative obligations would introduce, at least, duties to prevent future infringements with a more general scope, even if they can be still considered specific in relation to previously identified content or rights.

In regard to harmful content to minors and hate speech on video-sharing platforms, the *Audiovisual Media Services Directive*[60] provides (art. 28b) – certainly, without prejudice of Articles 14 and 15 of the *Directive on Electronic Commerce*[61] – the obligation of video-sharing service providers to adopt adequate measures to protect minors and prevent hate and violent speech (terms of use, report and flag system, age verification, parental control, rating mechanisms, explanation of reporting and flagging).

The system could veer from a notice-and-take-down-based model toward a duty-of-care-centred model. In this context, several regulatory proposals in the EU, as referred to above, seem to reveal a possible replacement of the horizontal liability intermediary regime with a number of vertical sectorial liability regimes introducing filtering obligations, proactive measures, or protective mechanisms for intermediaries and platforms in areas such as copyright infringement, illegal activities, or minors' protection. Which protocols, good practices, and measures that platform and intermediaries could implement to fulfil their "duty of care" will be a key strategical issue likely to affect the sustainability of the model and the stability of the market.

More importantly, and even if such a regulatory change would not materialize, a policy shift from an intermediary liability approach to

---

[57] U.S. Copyright Office, Joint Supplemental Comments of American Federation of Musicians et al. (Feb. 21, 2017), https://www.regulations.gov/document?D=COLC-2015-0013-92433.

[58] *Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives*, 96/9/EC and 2001/29/EC, OJ L 130, 17.5.2019, p. 92-125, art. 17.4

[59] *Directive (EU) 2018/1808 of the European Parliament and of the Council of 14 November 2018 amending Directive 2010/13/EU on the coordination of certain provisions laid down by law, regulation or administrative action in Member States concerning the provision of audiovisual media services (Audiovisual Media Services Directive) in view of changing market realities*, OJ L 303, 28.11.2018, at 69-92.

[60] *Id*.

[61] *Directive on Electronic Commerce*, *supra* note 22, at art. 14, 15.

intermediary responsibility strategy has clearly begun. The Communication *Tackling Illegal Content Online Towards and Enhanced Responsibility of Online Platforms*[62] is an extraordinarily illustrative expression of such a policy trend, subsequently crystallized in the *Commission Recommendation of 1.3.2018 on measures to effectively tackle illegal content online*.[63] Although it might not entail any specific amendments to current legal framework, as stated in the *Tackling Illegal Content Communication*,[64] this responsibility-enhancing strategy settles a new context to interpret current liability regime and reflect on future changes. Even if, at the moment, it is simply projecting a political stance and an effective collaboration between private players and authorities, it will undoubtedly be triggered any time the debate on how responsibility is articulated in specific obligations.

Third, platforms and intermediaries have responded to a challenging environment with the increasing implementation of voluntary monitoring mechanisms, automatic filtering, and self-regulatory actions to prevent illegal activities and enable private enforcement. For example, the ContentID scheme implemented by YouTube allows for singling out digital content in advance for the purposes of blocking, monitoring, or applying monetizing strategies.[65] Tripadvisor has deployed an alerting mechanism to warn users, adopting a proactive and active role beyond the notice-based borders and assuming its own duty accordingly.[66] Facebook has developed strategies to counter fake news and hate speech based on external fact-checkers and internal algorithm-conducted automatic procedures to detect content or sources.[67] Facebook, Microsoft, Twitter, and YouTube, along with other platforms and social media companies have agreed with the

---

[62] *Tackling Illegal Content Online Towards an Enhanced Responsibility of Online Platforms*, COM (2017) 555 final (Sept. 28, 2017) [hereinafter *Tackling Illegal Content Communication*].

[63] *Commission Recommendation of 1.3.2018 on Measures to Effectively Tackle Illegal Content Online*, COM 2018 1177 final (Jan. 3, 2018).

[64] As is stated in the Conclusions of the Tackling Illegal Content Communication (p. 20): "[t]his Communication provides guidance and does not as such change the applicable legal framework or contain legally binding rules." Likewise, it can be also inferred from Recitals 7), 26), 33), 36) and 41), and Chapter I, 3 of the Recommendation, insofar as the measures proposed by the EU Commission shall be applied "without prejudice" of the existing legal framework as defined in Articles 14 and 15 of the Directive 2000/31.

[65] *See* Google Help Center, *What is a Content ID claim?*, https://support.google.com/youtube/answer/6013276 (last visited May 24, 2020).

[66] *See* Lindsay Nelson, T*ripAdvisor's Commitment to Family Safety*, TRIPADVISOR, https://www.tripadvisor.com/blog/tripadvisors-commitment-to-traveler-safety-us (last visited May 24, 2020). Two new features have been implemented in the platform to enhance safety and security and facilitate the access to safety-related information.

[67] Tessa Lyons, *Hard Questions: What's Facebook's Strategy for Stopping False News?*, FACEBOOK (May 2018), https://about.fb.com/news/2018/05/hard-questions-false-news.

European Commission on a code of conduct setting a set of public commitments to counter the spread of illegal hate speech online.[68]

IV. CREDIBILITY-ENHANCING REGULATORY MODELS TO COUNTER FAKE NEWS: POSSIBLE MODELS AND IMPLICATIONS. A CASE FOR HARMONIZATION

Signs of change are perceptible, but whether these signals announce a future regulatory change or simply the raising of policy concerns that will be addressed with cooperation, self-regulation, market-driven solutions, and political initiatives is still uncertain. Yet it is the moment for anticipating possible models and discussing their implications.

A. *An Intermediary-Greater-Responsibility Model: Shift from an Intermediary Liability Approach to an Intermediary Responsibility Strategy*

A shift from an intermediary liability approach to an intermediary responsibility strategy is comprehensible in political and social terms, but it raises complexities to articulate its legal consequences. The model survives without legal reform only to the extent that cooperation, self-regulation, and voluntary measures work effectively.

A case for greater responsibility of intermediaries and platforms to combat illegal activities, hate speech, racism or extremism seems to start crystallizing in several resolutions, communications, and position papers at the European Union. Given the visible loss of protagonism of traditional authoritative sources, a greater-responsibility strategy to counter misinformation might be the expected move of European authorities. Intermediaries and platforms would collaborate with public bodies, and traditional authoritative exponents.[69] However, concurrently, they would emerge as new gatekeepers.[70] That position would give rise to the deterioration of an assumption of neutrality in medium and structures for creating opinion and the multiplication of reference points.

---

[68] EU Code of Conduct, https://ec.europa.eu/info/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/eu-code-conduct-countering-illegal-hate-speech-online_en#theeucodeofconduct (last visited May 24, 2020). The Code of Conduct was agreed on in May 2016 by the Commission with Facebook, Microsoft, Twitter and YouTube. In the course of 2018, Instagram, Google+, Snapchat and Dailymotion joined the Code of Conduct, and in January 2019, Jeuxvideo.com joined.

[69] *Tackling Illegal Content Communication*, *supra* note 62, at 8.

[70] Reinier Kraakman, *The Anatomy of a Third-Party Enforcement Strategy*, 2 J.L. ECON. & ORG. 53 (1986).

Yet, from a regulatory point of view, the decision to devise a model of intermediary responsibility to combat "fake news" has to address two delicate issues: the legal consequences of responsibility and the personal scope of that responsibility.

The first issue to deal with is how to articulate responsibility in legal terms. Unlike liability that is linked to patrimonial or administrative consequences (fines, sanctions, loss of license, prohibition to carry out an activity), responsibility here is configured as a set of commitments whose compliance is highly encouraged, but fundamentally depends upon voluntary measures. Market discipline and reputation play a critical role in this respect. Accordingly, regulatory strategies must principally consist of codes of conduct, EU-wide standards, good practices, and other self-regulation instruments. Such a policy option is convenient and frequently advisable, where a binding regulatory option is unworkable or may endanger the sustainability of the system, as well as where a temporary solution is needed to understand and define the problem in the transition to a more elaborated future regulatory action. The intricacies of the "fake news" problem might recommend such a provisory approach to enable platforms to develop private solutions and to test them in the market and compete before formulating a binding regulatory model.

In sum, a greater-responsibility model shows an appreciable level of adaptability and facilitates the transition to the next regulatory step without distorting the market and encouraging innovation in producing effective solutions. It is, however, a weak approach from the enforcement perspective as it essentially relies on voluntary cooperation of platforms and intermediaries.

The second issue to address is the delimitation of the scope. Bigger platforms arguably contribute in a greater manner to misinformation. The spread of "fake news" is wider, reaches a larger audience, and above all, creates a higher perception of credibility due to popularity indexes. Both the size and the penetration of bigger platforms expose them to greater risks and concurrently, would seem to justify an imposition of greater responsibility. That reasoning appears to be behind the scope of the *Code of Conduct on countering illegal hate speech online*, signed by Facebook, Microsoft, Twitter, and YouTube, as well as the scope of application of the German *NetzDG* that applies to platforms with at least two million registered users (in the Federal Republic of Germany).[71] Both initiatives target illegal or unlawful content, described in relation to typified offenses per the applicable legislation. The rationale behind exempting smaller

[71] Netzdurchsetzunggesetz [NetzDG] [Network Enforcement Act], Oct. 1, 2017, BUNDESGESETZBLATT, Teil I [BGBl I] at 3352 (Ger).

platforms from the obligations of monitoring and removal of content would be to avoid the substantial costs that implement and manage the procedure for handling complaints. Nonetheless, it cannot be ignored that doing so would likely create a competitive disadvantage in the platform market and, more interestingly, may trigger undesired "platform shopping." If the imposed obligations are too onerous and the penalties for non-compliance too stringent, platforms could be dissuaded from growing beyond the regulatory threshold in order to avoid falling under the legal regime.

The other possible interpretation of popularity in a responsibility-based model would be in relation to the content to be protected or monitored, that is, as a trigger for the platform to action. The aforementioned *Baidu* case develops this reasoning. Where digital content becomes popular, the platform should have the responsibility to pay special attention and adopt protective measures. In regard to "fake news," an equivalent analysis would lead to defend that, even if platforms are not subject to a general duty to monitor, they might be expected to carry out specific checking to assess veracity over those content that reach significant levels of popularity. In absence of a standard concept of popularity, such approach does moderately alleviate the burden on platforms to monitor, while the other issues remain unsolved.

B. *Alternatives to Define the Duties of Platforms to Counter Fake News*

As discussed above, one of the key tenets articulating the intermediary liability paradigm is the nonexistence of a general duty to monitor. Intermediaries are only called to act by disabling or removing relevant content upon obtaining knowledge. Thus, the knowledge-and-take-down pillar represents the second key tenet of the current liability regime. As it has been discussed before, there are signs of change pointing toward a possible paradigm shift.

Should a paradigm shift be considered, two scenarios may be envisioned. First, a general duty to monitor is laid out. Second, proactive monitoring duties are encouraged where certain conditions are met (blatantly illegal content, suspicious activities, hate speech, etc.). Under both scenarios, the critical policy decision is to determine whether such duties are *obligations de résultat* or *obligation de moyens.* In other words, should intermediaries be liable in case their implemented state-of-the-art mechanisms to monitor fail to detect illegal/harmful/false content, to do it timely, or to remove content or disable access in an effective manner?

Although it is declared[72] that the adoption of encouraged proactive monitoring does not entail losing the protection of safe harbor protection for collaborative platforms, it is undeniable that it implies obtaining knowledge and therefore triggering the duty to expeditiously react. How the adequacy of the measures will be assessed, and the consequences defective or ineffective measures will have, are relevant issues to discuss. Whether the duty to monitor, even on a voluntary basis, is an obligation of result or an obligation of means is uncertain. Whether proactive monitoring measures serve to provide genuine knowledge of illegality or harmful potential is disputable. Accordingly, intermediaries and platforms will perform their functions in a misty atmosphere.

Effectiveness in proactive monitoring can be significantly enhanced by incorporating automatic filtering, algorithm-based mechanisms, and AI-guided monitoring systems. Nevertheless, automation raises concerning risks of over-removal, and awakes the phantom of censorship. A growing trend toward the increase of transparency in the configuration, operation and self-learning processes of algorithms is echoing such concerns. Full disclosure and clear explanation on platforms' content policies in the terms of the service,[73] on notice-and-action procedures, and on automatic filtering criteria should attenuate those concerns. In market-oriented terms, transparency would increase competition in the market of platforms and enable reasonable choices and educated decisions.

Likewise, other safeguards against over-removal and abuse of the system might be adopted to alleviate the risk of encroaching upon the freedom of speech. Reasonable notice procedures, well-designed and continuously supervised automatic filtering, and balanced removal policy should be complemented with trusted flagging systems, counter-notice procedures, and measures to prevent and penalize bad-faith notices and counter-notices.

1. Alternative Liability Models to Consider

Alternative liability models range between two dimensions. On one hand, it has to be decided among three policy options: no liability, negligence-based liability or strict liability. On the other hand, a model of civil liability, administrative liability, or criminal liability can be devised.

If negligence-based liability were to be replaced by a strict liability system, serious implications on the market, the protection of rights

---

[72] *Tackling Illegal Content Communication*, *supra* note 62, at 10.

[73] Teresa Rodríguez de la Heras Ballell, *Terms of Use, Browse-Wrap Agreements and Technological Architecture: Spotting Possible Sources of Unconscionability in the Digital Era*, 2009 CONTRATTO E IMPRESA EUROPA 841.

(freedom of speech, free access to information, freedom to run a business), and the preservation of neutrality would have to be carefully gauged. Over-removal is a very likely expected result, as platforms and intermediaries will strive hard to minimize their exposure to liability risks.

Criminal liability would likely distort the market in an irreparable manner. As a consequence, freedom of speech, free flow of information, and dialogue values would be severely hampered. A model of administrative liability would penalize any typified contravention of those legal duties set out by regulations with fines, or other administrative sanctions. That has been the path taken in Germany with the enactment of the *NetzDG.* Under this Act, the commission of a regulatory offense as provided for by the law, either intentionally or negligently, will be sanctioned with a fine of up to five million euros.[74] Under this model, infringements are essential to a procedural nature: failure to provide a specified procedure, to supply it correctly, to monitor the handling of complains, to rectify an organizational failure in due time, or to name the required authorized person, among others. In sum, the law delineates a legal model for "a good, responsible platform/intermediary" (strictly speaking, provider of a social network in the terminology used by the law). The law encourages platforms falling under the scope of application to become more responsible in the fighting against illegal content. High fines would act as deterrents for deviation.

Unlike an administrative liability model, a civil liability model depends on the basic triggers for claiming liability. Fundamentally – fault, causation, and compensable damage. Platforms and intermediaries will certainly be encouraged to adopt adequate systems and formulate reasonable content policy to demonstrate diligence. Nevertheless, unlike illegal and harmful content, in cases of false content, damage will be diffuse and very frequently hard to quantify, as content is disseminated by users. An overzealous diligence of platforms to detect and block "fake news" might lead to unreasonable restriction of speech, a biased control of opinions, and a drastic increase of the costs of platforms' activity. The increase in cost and complexity favors big platforms and intensely disfavors small and medium competitors. Excessive costly measures could augment the concentration of the platforms' market.

2. A Case for Harmonization

The intermediary liability paradigm ("safe harbor" provisions), that has been the backbone of the digital living to date, is under consideration. The

[74] Netzdurchsetzunggesetz [NetzDG] [Network Enforcement Act], Oct. 1, 2017, BGBL I at 3355 (Ger).

sound liability regime seems to need a transformation to face new challenges. An alarming spread of "fake news" and a growing international concern about the pervasive penetration of misinformation does precisely defy the continuity of the liability paradigm for intermediaries in its current form. Additionally, the burgeoning Platform Economy represents an important challenge for the liability system. All variants of platforms, aggregators, social networks, sharing-based models, and a wide gamut of other intermediaries not only enable the emergence and blooming of "fake news," but principally trigger its virality as a proxy for credibility in an escalation of uncontrollable misinformation that is very hard and unlikely to counter with objectivity, fact-checking, and deep reflection. There is no time for that, and numbers play against.

The first conclusion of this Article is then that there are signs of change pointing at an eventual liability paradigm shift, but the resulting model is still uncertain and undefined. As discussed above, the implications of different alternative models are significant for the shaping of our digital society, the protection of rights, Internet neutrality, and the preservation of trust. In the case of false content, the need to set a fair balance between rights and interests at stake is trickier and even more imperative, as the collective memory, the global dialogue, and the access to information, knowledge and culture might be endangered.

Upon observation of such a still-uncertain paradigm shift, the second conclusion is that a line must be drawn to distinguish illegal content, harmful content, and false content. An eventual reform on the liability paradigm cannot be undertaken on an all-embracing basis. Both in terms of protected interests and potential harm, falsity-related situations differ from those defined by illegality and harmfulness. Therefore, a distinct and separate approach is needed to interpret the perceptible paradigm shift in the context of "fake news."

Third, if intermediaries and platforms should be forced or encouraged to act against alleged false content, as they are expected to detect, prevent, and remove illegal content and harmful one, the distorting effects may be unprecedented and highly undesired. As the line between untrue content and opinion is very thin, the encouragement of a zealous monitoring, verification, and filtering of potential false content may lead to discrimination and ideological marginalization, biased control, over-removal, or prevalence of dominant informative or ideological lines.

The impact of "fake news" casts over two dimensions: the factual one that represents its degree of veracity, and the social one that determines its credibility, on grounds of its popularity. Whereas inveracity can be fought with fact-checking, trusted flaggers, and authoritative gatekeepers,

any attempt to undermine the credibility perception requires play with popular equivalents. Those alternative models that have been previously discussed aimed to urge or encourage platforms and intermediaries to cooperate with public authorities, trusted third parties, and authoritative voices to detect false content address the first dimension. However, it is doubtful whether platforms should collaborate on infusing artificial popularity, stirring the spread of the fact-checking content or the rectification, or provoking prompt virality.

Considering the above-discussed alternative models are likely to result from a paradigm shift in liability, the following model proposal to enhance credibility and counter fake news is outlined below.

First, a greater-responsibility model to ensure cooperation of platforms and intermediaries with authorities seems to be a reasonable stage to start. It incentivizes innovation, increases competition, and is based on voluntary collaboration. However, it must be a temporary model – testing grounds for the future devising of a regulatory model.

Second, the promotion of transparency on content policy, notice and counter-notice procedure, and automatic filtering design and operation also provide a defensible regulatory solution. It is cautious, prudent, and only slightly invasive. Market discipline works and users make their informed free decisions.

Third, a strict liability regime, as well as the imposition of a general duty to monitor, to counter "fake news" cannot be allowed. The expected consequences would be highly undesirable and detrimental to the development of our digital society. Contrarily, a model aimed to encourage the implementation of clear notice and counter-notice procedure, reliable fact-checking and trusted flaggers, transparent content policy, and effective measures to prevent bad-faith notices and counter-notices appears to be a reasonable policy option to set a fair balance between the right to access to information, the freedom of speech and the liberty to run a business in a competitive market.

The dilemma is still whether to rely on voluntary compliance and self-regulation (code of conducts and collaboration), to articulate a model of administrative liability to sanction any contraventions of the legally established duties (in line with the German *NetzDG*), or to preserve a negligence-based liability model where civil liability will be triggered only upon assessing the concurrence of basic legal requirement (negligence, causation, compensable damage). At the moment, this Article tends to favor the latter liability model in the belief that it sets a fair balance of rights and interests, promotes innovation and competition among platforms to develop

cost-effective monitoring procedures, and safeguards the free exercise of fundamental liberties in an open society.

Fourth, the increasing employment of automatic filtering and algorithm-based monitoring is undeniable. The magnitude of digital society makes their use today reasonable and inevitable to maintain the biggest platforms' functionality. Human-based content-specific monitoring and checking is inconceivable on a proactive and general basis. Automatic detection followed by human assessment would certainly be a more practical model. Nonetheless, automation still raises many concerns and risks, along with its perceptible advantages in processing and monitoring. Automatic discrimination, opacity, or ideological/informational marginalization is credible fears, particularly in our thesis of the two dimensions of "fake news." Automatic filtering heavily impacts on the perception side. Credibility perception would be artificially inflated or deflated by the effect of automatic blockage of certain content. Therefore, careful legislative attention on automatic mechanisms is imperative.

To that end, the recent *EU Regulation on Data Protection*[75] sets an important precedent to make algorithms accountable that, despite the specific scope of the Regulation, it might be extrapolated to the automatic filtering mechanisms to counter false content. As per Article 22 of the *EU Regulation*, the use of automated decision-making, if it has legal consequences for the person whose data is concerned, or if it affects this person in other significant ways, it is prohibited. This general prohibition is limited by three broad exceptions: a specific law authorizes algorithmic decision-making; it is based on an individual's explicit consent; or it is needed for entering into or performing a contract. In this context, safeguards must be implemented to facilitate the exercise of the right to obtain human intervention, the right to express one's point of view and the right to contest the automated decision. Nonetheless, it has still been alleged that such protective measures might not suffice and the need to enshrine a singular right to receive an explanation how the algorithms work and how a specific decision was made has been proposed.[76] A right to explanation that goes beyond a mere duty of transparency may apply to automatic filtering on "fake news" and it may play an effective role in the countering model to be designed.

Finally, this Article concludes that any action to counter fake news should be widely coordinated and harmonized at an international level. In

---

[75] Regulation 2016/679, 2016 O.J. (L 119) 1 (EU).

[76] Sandra Wachter, Brent Mittelstadt & Luciano Floridi, *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation*, 7 Int'l Data Privacy L. 76 (2017).

fact, irrespective of the adopted regulatory model, no change in the liability paradigm should be conducted on a local or regional basis. As it has been noticed, risks of a paradigm shift in intermediary liability are high, but the risks of a non-harmonized action in this issue are immense. Fragmentation, discrepancies among jurisdictions, and legal and regulatory arbitrage would exacerbate the perception of misinformation and lack of credibility in digital scene. More importantly, a disharmonized strategy against "fake news" would likely provoke a new variant of regulatory arbitrage – "platform shopping." Discrepancies in regulations and diversity in platforms' policies and procedure would fragment the digital scene in a plurality of fora. The production and dissemination of "fake news" might circumvent the most stringent regulatory models and the most fake-news-unfriendly platforms with skillful "platform shopping." Only if more rigorous regulatory models and more respectful platforms manage to make their strategies a proxy for credibility, the regulatory competition will produce a positive effect. Then complying platforms would become trusted third parties, returning to a centralized-trust model. Otherwise, if regulatory arbitrage deteriorates confidence and impedes users' ability to identify credibility indicia, misinformation would endure eluding the efforts made to counter it.