PLATFORM VALUES:

CONFLICTING RIGHTS, ARTIFICIAL INTELLIGENCE AND TAX AVOIDANCE

EDITED BY LUCA BELLI AND NICOLO ZINGALES

*CONSOLIDATED AND UNEDITED VERSION OF THE **SPECIAL ISSUE OF THE COMPUTER LAW AND SECURITY REVIEW**, CELEBRATING THE FIFTH ANNIVERSARY OF THE IGF **COALITION ON PLATFORM RESPONSIBILITY***

2019 OUTCOME OF THE IGF COALITION ON PLATFORM RESPONSIBILITY.

PRESENTED AT THE 14TH UNITED NATIONS INTERNET GOVERNANCE FORUM.

27 NOVEMBER 2019, BERLIN.

TABLE OF CONTENTS

- **Platform Value(s):A Multidimensional Framework for Online Responsibility**

Luca Belli and Nicolo Zingales

INTRODUCTORY ESSAYS

- **Governing Digital Societies: Private Platforms, Public Values**

José van Dijck

- **A Constitutional Moment: How We Might Reimagine Platform Governance**

Nicolas Suzor

- **From the Telegraph to Twitter: The Case for the Digital Platform Act**

Harold Feld

CONFLICTING RIGHTS

- **The New City Regulators: Platform and Public Values in Smart and Sharing Cities**

Sofia Ranchordas and Catalina Goanta

- **Sanctions on Digital Platforms: Balancing Proportionality in the Modern Public Square**

Engerrand Marique and Yseult Marique

- **A New Framework for Online Content Moderation**

Ivar Hartmann

ARTIFICIAL INTELLIGENCE

- **Socio-Ethical Values and Legal Rules on Automated Platforms: The Quest for a Symbiotic Relationship**

Rolf H. Weber

- **Democratising Online Content Moderation: A Constitutional Framework**

Giovanni De Gregorio

- **Platform Values and Democratic Elections: How Can the Law Regulate Digital Disinformation?**

Chris Marsden, Trisha Meyer and Ian Brown

TAX AVOIDANCE

- **The Progressive Policy Shift in the Debate on the International Tax Challenges of the Digital Economy: A "Pretext" for Overhaul of the International Tax Regime?**

Alessandro Turina

- **E-commerce and Effective VAT/GST Enforcement: Can Online Platforms Play a Valuable Role?**

Luisa Scarcella

ANNEX

- **Best Practices Platforms' Implementation of the Right to an Effective Remedy**

Collectively elaborated by members of the IGF Coalition on Platform Responsibility

Platform Value(s): A Multidimensional Framework for Online Responsibility

Luca Belli and Nicolo Zingales***

Is there a common understanding on the type of values that should be promoted by platform regulations? What values should be baked into platforms' technical architectures? What values should be promoted and enforced by regulators? Is it acceptable for both platforms and regulators to follow their idiosyncratic views when deciding to prime certain values that will directly affect how individuals communicate, inform themselves, organise their democratic and fiscal systems and conduct businesses globally?

The idea behind this special issue was precisely to stimulate a discussion on the multiform notion of platform value(s). The term "value" is thus construed broadly here to embrace a range of social, ethical and juridical values underpinning digital platforms, as well as the economic value that is generated and extracted within platform ecosystems.

Digital platforms are themselves central to the digital ecosystem, which in turn dramatically affects the structure of offline activities as well. For instance, platforms provide essential means by which connected individuals find information and information is directed consumers, digital marketplaces allowing both individuals and (small and medium) enterprises to exploit the advantages of e-commerce, and various digital services that have transformed education, communication and entrepreneurship. Increasingly, the ways in which platforms operate affect individuals' ability to develop their opinion and personality and engage in a substantial amount of social, political and economic interactions.

Given the scale of their impact, it is crucial to understand that platforms' architectural and regulatory choices are not neutral. The values that drive such choices affect both our personal lives and the functioning of our democracies and markets. To give one paradigmatic example, one may simply consider the platforms' capacity to collect and use personal data. Platforms' design and regulatory choices in that regard affect simultaneously individuals' fundamental rights, the functioning of democratic systems, labour relations and competition in the market. They also have significant tax implications, in particular considering that existing fiscal systems struggle to capture the immense value derived from people's data and the free labour as information producers, and that data processing and monetisation usually takes place in foreign-located servers. On top of this, as societies and economies become increasingly interconnected, (digital) value chains have become global; and the algorithmic systems that organise digital platforms have the potential not only to extract value but also to define values on a global scale, providing extremely effective proxies to exercise influence and control that transcends national borders.[1] In this light, the lack of a clear articulation of platforms' roles and responsibilities in addressing societal challenges has been decried by opposing sides of the political and stakeholder debate, prompting calls for regulation without sufficient consideration of its unintended consequences.

These themes were prominent in the discussions hosted over the past years by the United Nations Internet Governance Forum (IGF) Coalition on Platform Responsibility.[2] In preparation for the 2019 IGF in Berlin, a Call for Papers was collaboratively drafted by Coalition members and circulated more

*Professor of Internet Governance and Regulation, Fundação Getulio Vargas (FGV) Law School, Brasil.
**Associate Professor of Law, University of Leeds, United Kingdom.

[1] In this sense, see for instance: Cédric Villani, For a Meaningful Artificial Intelligence towards a French and European Strategy. (Mission assigned by the Prime Minister Édouard Philippe: A parliamentary mission from 8th September 2017 to 8th March 2018) https://www.aiforhumanity.fr/en/

[2] IGF coalitions are issue-focused multistakeholder working groups that meet at the IGF on a yearly basis, and are encouraged to produce outcome documents in the intervening period.

widely to solicit reflections and gather a diverse range of analytical perspectives on the various dimensions of platform values. The fifth anniversary meeting of the Coalition on 27 November 2019 will thus offer an opportunity to look back over several years of stimulating and productive multistakeholder interactions[3] analysing the roles and responsibilities of platforms and shed light on the Coalition's intellectual journey into new conceptual grounds.

1. Background: The Coalition's Journey to Platform Value(s)

To understand where we are headed, it is important to know where we are coming from. In this spirit, to grasp the reasons behind this work on "Platform Values" it is necessary to appreciate the evolution of the Coalition's endeavours. Thus, the following paragraphs provide a background picture of the origins of the Platform Responsibility debate at the IGF, and its progression to the current state, well beyond the forum's community of stakeholders.

First, and to render unto Caesar the things that are Caesar's, it is fair to say that the biggest achievement of the Coalition is to have coined and promoted the very concept of "Platform Responsibility."[4] Such concepts aim at interrogating, on the one hand, the impact that private ordering regimes designed and implemented by platforms have on individuals' capability to enjoy their fundamental and, on the other hand, the moral, social and human rights responsibility[5] that platforms bear when setting up such regimes.

The initial goal of this Coalition was to stimulate debate and participatory analysis on the meaning of platform providers' responsible behaviour. From the early steps, it was clear to participants that the starting point should be an analysis of the application to digital platforms of the UN Guiding Principles on Business and Human Rights[6], in particular their responsibility to respect Human Rights and to grant effective grievance mechanisms.[7] To lay the foundations of such work, the participants to the inception meeting of the Coalition, in 2014 at the IGF in Istanbul, suggested the development of a set of recommendations on core dimensions of platform responsibility.[8]

The resulting Recommendations on Terms of Service and Human Rights[9] (hereinafter "the Recommendations") presented at the 2015 IGF demonstrated that the cross-disciplinary effort facilitated by the Coalition could lead to concrete outcomes, providing a sound response to all those arguing that the IGF is a mere talking shop, unable to achieve tangible outcomes. The Recommendations provide concrete evidence that the IGF can elaborate solid outputs, including recommendations, as the IGF mandate itself explicitly states, prescribing that the Forum shall "find

[3] For an overview of the Coalition's work, see https://www.intgovforum.org/multilingual/content/dynamic-coalition-on-platform-responsibility-dcpr
[4] See Luca Belli, Primavera De Filippi and Nicolo Zingales, A New Dynamic Coalition on Platform Responsibility within the IGF (Medialaws, 11 June 2014) <http://www.medialaws.eu/a-new-dynamic-coalition-on-platform-responsibility-within-the-igf/> accessed 10 October 2019
[5] See John Ruggie, Guiding Principles on Business and Human Rights: Implementing the United Nations "Protect, Respect and Remedy" Framework, Report of the Special Representative of the Secretary-General on the issue of human rights and transnational corporations and other business enterprises (UN Human Rights Council Document A/HRC/17/31, 21 March 2011) <www.ohchr.org/Documents/Publications/GuidingPrinciplesBusinessHR_EN.pdf> accessed 10 October 2019
[6] Idem.
[7] Luca Belli, Primavera De Filippi and Nicolo Zingales (eds.), Recommendations on terms of service & human rights. Outcome Document n°1 (Internet Governance Forum 2014) <https://www.intgovforum.org/cms/documents/igf-meeting/igf-2016/830-dcpr-2015-output-document-1/file accessed 10 October 2019
[8] See Nicolo Zingales and Luca Belli, Dynamic Coalition on Platform Responsibility: Report of the "inception" meeting at the 2014 IGF, (Internet Governance Forum 2014) <https://www.intgovforum.org/multilingual/index.php?q=filedepot_download/4905/631> accessed 10 October 2019
[9] See Luca Belli, Primavera De Filippi and Nicolo Zingales (eds.), Recommendations on terms of service & human rights. Outcome Document n°1 (Internet Governance Forum 2014) <https://www.intgovforum.org/cms/documents/igf-meeting/igf-2016/830-dcpr-2015-output-document-1/file accessed 10 October 2019

solutions to the issues arising from the use and misuse of the Internet" as well as "identify emerging issues […] and, where appropriate, make recommendations."[10]

Indeed, the Recommendations served as an inspiration for (and were annexed to) both the study on Terms of Service and Human Rights[11], co-sponsored by the Council of Europe and FGV Law School, and the 2017 outcome of the Coalition - a volume entitled 'Platform regulations: how platforms are regulated and how they regulate us' featuring research by an ample range of stakeholders .[12] It also bears noting that the "platform responsibility" approach and a conspicuous number of elements of the Recommendations can be found in the Council of Europe Recommendation CM/Rec(2018)2 of the Committee of Ministers to member States on the roles and responsibilities of internet intermediaries.[13] Fostering this kind of multi-stakeholder and cross-institutional discussion is a core component of the vision behind the creation of the Coalition: to critically analyse challenging questions and collaborative develop potential solutions that, if deemed suitable and efficient, can inspire policymaking exercises.

The Recommendations and of the 2017 volume on Platform Regulations stressed the need to advance further the Coalition's work with two different yet complementary initiatives. First, the elaboration of concrete suggestions on how to implement the right to due process within regard to the remedies provided by online platforms' dispute resolution mechanisms. Such goal was achieved by organising a one year-long participatory process, leading to the Best Practices Platforms' Implementation of the Right to an Effective Remedy[14] that can be found as an annex of this Special Issue. Second, the various debates, cooperative processes and research developed by the Coalition members highlighted the need for a deeper analysis going beyond the notion of platform responsibility and platform regulations, but on the very values underlying the operation of digital platforms.

2. Why Platform Values?

As noted above, digital platforms have acquired a predominant role in digital policy circles and amongst Internet scholars,[15] due to the enormous impact that their choices, activities and self-regulatory initiatives can have on the lives of several billion individuals. Building on that recognition, the rationale behind the present inquiry is to understand and scrutinise the ways in which digital platforms are deploying their influence and power and, particularly, the values that they are willingly conveying, promoting and creating or extracting. As the contributions to this special issue explain, platforms play an active role by, on the one hand, diffusing and instilling specific values and, on the other hand extracting value in a methodically planned fashion. In fact, the ways in which platforms circulate values or siphon value result from explicit business choices. In this perspective it becomes essential to explore three of the most crucial points of contention with regard to the values that underlie the operation of digital platforms.

First, the design of dispute resolution mechanisms and the ways in which such mechanisms deal with conflicting rights and principles. This analysis is crucial because the ways in which such mechanisms are structured and the values based on which disputes are solved have a direct and substantial effect

[10] See Tunis Agenda for the Information Society (18 November 2005). WSIS-05/TUNIS/DOC/6(Rev. 1)-E. para. 72.k and 72.g. <https://www.itu.int/net/wsis/docs2/tunis/off/6rev1.html> accessed 10 October 2019
[11] See Jamila Venturini et al. Terms of service and human rights: an analysis of online platform contracts. (Revan, in collaboration with the Council of Europe and FGV Direito Rio 2016) <https://bibliotecadigital.fgv.br/dspace/handle/10438/19402> accessed 10 October 2019
[12] See Luca Belli and Nicolo Zingales (Eds.), Platform regulations: how platforms are regulated and how they regulate us. (FGV Direito Rio 2017) <https://bibliotecadigital.fgv.br/dspace/handle/10438/19402> accessed 10 October 2019
[13] See http://bit.ly/CoEinternetintermediaries accessed 10 October 2019
[14] The Best Practices can be also found on the IGF website https://www.intgovforum.org/multilingual/index.php?q=filedepot_download/4905/1550 accessed 10 October 2019
[15] See, as an instance, Jacques Crémer. Yves-Alexandre de Montjoye. Heike Schweitzer, Competition Policy for the digital era. European Commission Directorate-General for Competition (2019); Samantha Eyler-Driscoll, Asher Schechter, Camilo Patiño, Digital Platforms and Concentration, (ProMarket and Chicago Booth Stigler Center 2019); BRICS Competition Law and Policy Centre, Digital Era Competition Law: A BRICS Perspective (2019) <https://cyberbrics.info/digital-era-competition-brics-report/> accessed 10 October 2019

on individuals' capability to enjoy an ample spectrum of fundamental rights, including the right to due process of law[16], the right to protection against discrimination, the right to freedom of expression and the right to privacy. In light of these considerations, it is reasonable to argue that the means and costs of resolving disputes through alternative mechanisms designed by digital platforms should be proportionate to the importance and nature of the issues at stake.[17] However, due to their for-profit nature, it is natural to expect that the design and implementation of platforms' dispute resolution mechanisms are largely driven by considerations of cost minimisation and avoidance of potential liability, rather than the maximisation of the protection of individual rights.[18]

Second, the values that can or should be baked into platforms' automated decision-making mechanisms. However, the task of implementing such principles may be extraordinarily complex, particularly considering that the precise meaning of those principles is not universally agreed upon, nor are the processed and methods to bake them into the design of technical systems. The task may become even more complex when conflicts and contradictions amongst values and principles arise, and either the humans designing the systems or the "intelligent" system themselves have to ponder between what should prevail. An automated system of a "gig economy"[19] platform should privilege labour protection or business efficiency? A mechanism aimed at implementing the right to be forgotten in a search engine should privilege personal data protection or freedom of information? A social network timeline should remove or reduce visibility of politically relevant content categorised as misinformation?

Another critical role for platform regulations, in particular regarding artificial intelligence, is to ensure an environment conducive to the production of economic value. Defining what constitutes economic value and the difference between its creation and its extraction is a foundational step in establishing a balanced framework of rights and obligations that serves the advancement of societal well-being. Such framework notably includes the freedom to conduct business for platform users and operators and the rights of appropriation over data generated through platform engagement, which substantially and crucially contribute to the development of AI systems.

Third, the tax avoidance strategies that may be pursued by tech giants to minimise their fiscal responsibility across the multiple jurisdictions in which they provide their services, and the ways in which platforms may become instrumental to the enforcement of national tax regimes. It is becoming increasingly apparent that all nation states would benefit from a clearer and coordinated understanding of roles and responsibilities in this space, including by addressing platforms and other multinational enterprises' capability to exploit existing gaps and mismatches between different countries' tax systems, to erode their tax-base and shift their profits avoiding their fiscal obligations.[20]

3. The Necessary Debate on Platform Value(s)

Digital platforms have become gateways to speech, innovation and value creation and highlighted their ascendance as central elements of our society, economy, and public sphere in redefining the concepts

[16] The Best Practices Platforms' Implementation of the Right to an Effective Remedy are an attempt to provide concrete guidance on how to respect this specific facet of the right to due process of law. See <https://www.intgovforum.org/multilingual/index.php?q=filedepot_download/4905/1550> accessed 10 October 2019

[17] In this sense, see Mark Elliot and Robert Thomas, Tribunal Justice and Proportionate Dispute Resolution, [2012] 71 (2) Cambridge Law Journal

[18] In this sense, see Luca Belli, Pedro Augusto Francisco and Nicolo Zingales, 'Law of the Land or Law of the Platform? Beware of the Privatisation of Regulation and Police' in Luca Belli and Nicolo Zingales (Eds.), Platform regulations: how platforms are regulated and how they regulate us. (FGV Direito Rio 2017) <https://bibliotecadigital.fgv.br/dspace/handle/10438/19402> accessed 10 October 2019

[19] The term "gig economy" is commonly utilized to refer to a free market system where organisations typically contract independent workers for short-term engagements and, consequently, workers typically hold temporary and precarious positions.

[20] See OECD, International collaboration to end tax avoidance. Under the OECD/G20 Inclusive Framework on BEPS, over 130 countries are collaborating to put an end to tax avoidance strategies that exploit gaps and mismatches in tax rules to avoid paying tax, s.d. <https://www.oecd.org/tax/beps/> accessed 10 October 2019

of “private” and “public”, and challenging conventional approaches to regulation and governance. Along those lines, this Special Issue starts from the consideration that, to guarantee the balance and sustainability of governance systems, the exercise of power should be constrained. To do so, a well-informed deliberative process over the aims, mechanisms and boundaries of regulation is needed. To meet such need, this Special Issue offer a selection of ideas, reflections, and proposal whose main purpose is to trigger – when necessary – and nurture the above-mentioned much needed debate.

In fact, when private entities rise to the level of quasi-sovereigns[21] or private regulators, it is natural to expect discussion, shared understanding and scrutiny of the choices and trade-offs embedded in their private ordering. Yet despite the fact digital platforms are becoming instrumental to socialise, communicate and do business, in many countries, particularly in the least developed ones, there is little discussion of the ways in which these all-important players are generating, shaping and championing values. Accordingly, more work is needed to question what counts as value and how value judgment ought to be made in these hybrid spaces, exploring the elements that should underpin legal and policy-making initiatives and the risks that may occur when decision-making remains in the sole province of contracts and self-regulation.

At this juncture, it is important to think more clearly about the boundaries and accountability mechanisms to frame platform responsibilities. For instance, is it appropriate for deliberations over platform values and user rights to be exclusively driven by the economic imperatives of shareholders? What are the best strategies to take into considerations the broader set of concerns and expectations of the stakeholders affected by platform regulations? And what is the role of the market in restoring a level playing field?

To introduce this Special Issue, we decided to feature a selection of short essays penned by a group of authors that have distinguished themselves by recently publishing innovative contributions to the debate on platform values.

In her introductory essay on “**Governing Digital Societies: Private Platforms, Public Values**,” José van Dijck emphasises that dominant platforms are facing a serious ‘techlash’ due to an incredibly ample spectrum of scandals and abusive practices, including the spread of disinformation and hate speech, elections manipulation, massive privacy breaches and security leaks, tax evasion and the undermining of fair labor laws. Van Dijck argues that the neoliberal architecture of the GAFAM (Google-Alphabet, Amazon, Facebook, Apple, and Microsoft) goes beyond market value, as the platforms increasingly influence the very texture of society and the process of democracy. However, the values conveyed but the GAFAM can be challenged and different platform governance strategies can emerge. She emphasises that Europe should articulate its own approach based on its appraisal of a strong public sector, independent institutions, fair taxation, and the common good. A loss of public trust is ultimately a loss of business value, and societal value must be pursued in addition to market value: both types of values are integral part of a nation’s economic strength.

In a complementary perspective, Nicolas Suzor’s essay on “**A Constitutional Moment: How We Might Reimagine Platform Governance**”, argues that the states’ yearning for regulation – of both user behavior and platform responsibilities – and the increasing awareness of platforms’ role as focal points of control should be seized as an opportunity for us all to rethink how the Internet should be governed. Suzor posits that we are currently facing a constitutional moment for the future of the Internet, and that the hallmark for legitimacy in this process should be the principle that the rules of a society should be created and enforced in a way that is predictable and fair. In this spirit, it is of utmost importance that we collectively debate and deliberate on how power is held to account, and whose values prevail on digital platforms.

[21] In this perspective, see e.g. Lawrence Lessig, Code and Other Laws of Cyberspace (Basic Books 1999); Rebecca McKinnon, Consent of the Networked: The Worldwide Struggle for Internet Freedom, (Basic Books 2012); Luca Belli, De la gouvernance à la régulation de l’Internet, (Berger-Levrault 2016)

Harold Feld's essay on "**From the Telegraph to Twitter: The Case for the Digital Platform Act**" goes a step further, calling for the elaboration of a new type of institution, able to frame platforms' roles and responsibilities. Feld stresses that digital platforms have unique features and make up a unique sector of the economy that does not behave like a traditional competitive market for goods and services. The combination of multisided market structure, the importance of personal data and information to these markets, and the strength of the network effects enjoyed by successful platforms, allow digital platforms to exercise both market dominance and outsized influence in society as a whole. The influence and power that some of these actors have achieved is such that some authors, including Feld, are now calling for the establishment of dedicated regulatory agencies to address both concerns about competition and broader societal concerns. In the words of Feld, such regulator should be a "Digital Authority" that he describes in his call for a Digital platform Act.

In the perspective of stimulating a constructive exchange of ideas on platform values, and the strategies available to understand and, when needed, regulate such actors, this volumes features a selection of articles providing analyses and putting forward concrete solutions and policy proposals in relation to three main dimensions of this all-important debate. The next sections present the contributions exploring the three dimensions dedicated to conflicting rights, artificial intelligence systems and tax strategies.

3.1. Conflicting Rights

The first set of governance questions analysed in this volume pertain to the intersection of conflicting rights and values. A number of questions have orientated the analyses provided in this work, such as what should be the value to be actively promoted by platforms? Should platforms prioritise certain rights or principles over others? Are platforms the best-placed entities to identify which rights should be privileged when regulating social interactions? How should such balancing be conducted between conflicting rights of the same nature, for example between conflicting economic freedoms or conflicting fundamental human rights? What is the relevance of the sources of those rights, for instance in conflicts between rights enshrined in terms of service and diverging conceptions of those rights under the "law of the land"? Should principles, community guidelines and rules of practice (including internal precedents) be weighed any differently as part of balancing? Should balancing be ruled out for certain conflicts?

In "**The New City Regulators**," Sofia Ranchordas and Catalina Goanta note that when discussing these value questions, it is essential to understand the stakeholders in relation to which certain values are held: for instance, efficiency and other market-based platform values may be a boon to users and shareholders, but often come at the expense of the interest of workers or the broader society. Behind such conflicts lie often significant divergences between the values carried forward by digital platforms and those upon which our societies have been grounded. For this reason, they propose a framework based on a set of shared values between platforms and public authorities to promote the use of technology in the public interest, and the creation of a municipal clearinghouse for negotiation between cities and technology companies wishing to launch products and services with a direct impact on public infrastructure.

Subsequently, in "**Sanctions on Digital Platforms: Balancing Proportionality in the Modern Public Square**," Engerrand and Yseult Marique revisit the concept of proportionality, joining a nascent stream of literature on its application in the context of digital platform. In their contribution, they highlight the crucial connection of proportionality to the type of sanctions imposed, and therefore usefully distinguish four different scenarios of conflicting rights. To allow for the adaptation of the proportionality framework to the necessities of the evolving business models of digital platforms while maintainting oversight and accountability, they propose the creation of a public forum where different operators could share

aggregate data about types of violations, modalities of sanctions, user complaints, and results of their dispute resolution process.

Lastly, Ivar Hartmann's article on "**A New Framework for Online Content Moderation** contextualises and explores the reflection on platforms values in the field of content moderation." Hartmann reviews the profound transformations brought by the Internet to the moderation of speech. This includes fundamental structural changes, such as the decentralisation in the dissemination of information and the privatisation of substantial decision-making, which make judicial review both unfeasible and undesirable. In this context, he sketches a proposal for a new framework, strictly compliant with due process, which shifts more decision-making power onto user communities. In the framework proposed by the author, legislators establish the general procedural guidelines for content self-regulation systems and courts review the extent to which platforms have respected such guidelines, while the administrative State plays a role in between- to review procedural elements enshrined in the system as architecture choices.

3.2. Artificial Intelligence

The second set of questions can be seen as twofold. On the one hand, it relates to value appropriation, in particular in the "scramble for data"[22] and insights that can be extracted from it to power a new breed of artificial intelligence applications. Since data is a key input for the improvement of algorithms, profiling, and the elaboration of new cognitive services, should data subjects and other players in the platform ecosystems share in the value generated by their marginal input? Should platforms be the only beneficiaries of this learning process, or should the law constrain their ability to exclude others (including consumers, workers, competitors and complementors) from sharing in the benefits generated by the platform ecosystem? Is the surrender of democracy and (informational) self-determination the true value paid to enjoy the supposedly "free" services provided by platforms?

On the other hand, the development and implementation of artificial intelligence systems to automatise decision-making functions calls into question the values that should be "baked" into such systems in order to minimise negative consequences and strive towards the design and development of ethical automated systems. In this respect, what are the fundamental values that should orientate the design, development and deployment of artificial intelligence within platforms? How can those values be appropriately incorporated into artificial intelligence solutions implemented within platforms? Are the principles of transparency, non-discrimination and due process sufficient to prevent unfair value extraction, or do we need stronger intervention?

In the "**Socio-Ethical Values and Legal Rules on Automated Platforms: The Quest for a Symbiotic Relationship**," Rolf H. Weber convincingly argues that solutions implemented by platforms – and the values that orientate the development and functioning of such systems –impact on both society and the competitive environment. For this reason, the deployment of artificial intelligence on automated platforms needs to go hand in hand with the development of legal frameworks safeguarding socio-ethical values as well as fundamental rights, particularly the self-determination and the non-discrimination principle. Notably, the author maintains that a trust-based approach focused on human values can mitigate a potential clash between a solely market- and technology-oriented use of artificial intelligence and a more inclusive multistakeholder approach. In this perspective, the regulatory tools are to be designed in a manner that leads to a symbiotic relationship between ethics and law. Weber stresses that both socio-ethical and legal elements play a role within a framework of automated platforms and of AI governance in general. However, even if both are necessary, neither is sufficient

[22] See Luca Belli, The scramble for data and the need for network self-determination, (openDemocracy, 15 December 2017) <https://www.opendemocracy.net/luca-belli/scramble-for-data-and-need-for-network-self-determination> accessed 10 October 2019

and neither can substitute the other. The two disciplines act in a complementary way and can inspire each other.

Importantly, the choice to let one value prime over another or to determine whether and how socio-ethical considerations should influence regulation is typically a political one made by governments and legislators rather than platforms providers. Is the adoption of "AI ethics guidelines" an adequate response? Or is such strategy simply passive acceptance of the status quo? Any automated system is built, trained (using data-sets that may convey specific values or bias[23]), tested and overseen by humans that have the possibility to shape such systems according to their own values (and bias). Is the current lack of dedicated enforcement or governance mechanisms compatible with an alignment of AI with the collective good, or is it more properly seen as a delegation to private platforms (or more correctly to platforms developers and executives) of the power to define winners and losers? Is the automatization of identification and removal of highly nuanced and vague items such as content configuring (or not) "hate speech," "fake news" or "obscene material" a suitable option?

In his article on "**Democratising Online Content Moderation: A Constitutional Framework**," Giovanni De Gregorio provides useful elements to answer the abovementioned questions. The author stresses that freedom of expression is one of the cornerstones on which democracy is based but the troubling evolution of the current algorithmic society challenges both freedom of expression and democracy, subjecting them to opaque artificial intelligence technologies aimed at governing the flow of information online. De Gregorio notes that digital platforms establishing such technical tools are usually neither accountable nor responsible for contents uploaded or generated by the users. Nevertheless, online content moderation policies and tools designed by platforms affect users' fundamental rights, especially when assessing requests to remove flagged contents whose illegal nature is not always so evident.

Crucially, despite their crucial role in governing the flow of information online, social media platforms are not required to ensure transparency and explanation of their decision-making processes. Aware of such context, De Gregorio's work illustrates that ensuring that public actors do not interfere with the right to freedom of expression online is not enough to promote a democratic online environment in the algorithmic society. Although public actors' non-interference is still paramount, it is necessary to enhance the positive dimension of this fundamental right by establishing new users' rights based on transparency and accountability vis-à-vis online platforms.

Another important reflection on the impact and relevance of the disinformation debate and how to address it properly is provided by Chris Marsden, Trisha Meyer and Ian Brown, in their article on "**Platform Values and Democratic Elections: How Can the Law Regulate Digital Disinformation?**." The authors examine how governments can effectively regulate the values of social media companies that themselves regulate disinformation spread on their own platforms. Stressing that that disinformation initiatives directly impact on freedom of expression, media pluralism and the exercise of democracy, the authors focus on the responses to such phenomenon elaborated by the member states and institutions of the European Union. Importantly, Marsden, Meyer and Brown stress that regulating fake news should not fall solely on national governments or supranational bodies like the European Union, neither should the companies be responsible for regulating themselves. Instead, the authors argue favour co-regulation, highlighting that the companies develop – individually or collectively – mechanisms to regulate their own users, which in turn must be approved by democratically legitimate state regulators or legislatures, who also monitor their effectiveness.

[23] See for instance: Safiya Umoja Noble, *Algorithms of Oppression. How Search Engines Reinforce Racism*, (NYU Press 2018); Cathy O'Neil, *Weapons of Math Destruction: How Big data Increases Inequality and Threatens Democracy* (Crown Random House 2016).

3.3. Tax Avoidance

Finally, it is necessary to appreciate whether platforms provide long-term value with their functionalities (for example, bringing together different sides) or rather primarily engage in value extraction (for instance, limiting choice and deriving advantages in favouring certain kinds of behaviours or business models) and regulatory arbitrage. Defining how and where the value is created is crucial in determining the tax regime that is applicable to their activities, and in identifying unfair or fraudulent transfers of wealth. How should value be constructed for tax purposes, and how should regulators around the world deal with global tech giants? Are recent legislative initiatives on digital VAT marking the beginning of an inevitable race to the bottom to attract investment by global platforms, or do they set the foundations for interstate cooperation? Are existing reflections, such as the OECD's works on transfer pricing and Base Erosion and Profit-Shifting sufficiently mature to be implemented by states? And, most importantly, are states willing and able to implement existing proposals? Is a national or local tax on intermediaries for data collection and aggregation a viable way to account for the transfer of value that takes place between users and platforms?

Alessandro Turina reviews some of these challenges in his contribution entitled "**The Progressive Policy Shift in the Debate on the International Tax Challenges of the Digital Economy: A "Pretext" for Overhaul of the International Tax Regime?**". He presents some of the ongoing proposals for reform of taxation in light of the digital economy, focusing in particular on the difficulties associated with fitting the concept of "value creation" within the pre-existing framework based on "source" and "residence". For instance, the importance of user participation in the creation of value is not a concept that is peculiar to the digital economy, and therefore does not seem to justify a "ring-fencing" of certain activities to address the challenges of profit shifting. Similarly, insisting on physicality in the definition of permanent establishment to broaden the notion of "significant economic presence" seems to miss the mark of the intangible nature of digital assets. Furthermore, these solutions would imply a wholesale rethinking of some of the core tenets of the international tax system, including the rules on transfer pricing. For this reason, Turina argues in favor of a residence-based approach of redistribution of value for digital supplies, requiring the adoption of a new international tax treaty. He suggests that the current tendency to adopt sector-specific responses to the rise of the digital economy, particularly at the national level, may signal a momentum for the overhaul of the international tax regime.

A second contribution on tax governance is Luisa Scarcella's article on "**E-commerce and Effective VAT/GST Enforcement: Can Online Platforms Play a Valuable Role?",** which rather than focusing on profit shifting by platforms constructively explores how to involve platforms into tax-law enforcement. The author notes that just taxation is a value that should be promoted by platforms not only with regard to their own activity, but also with regard to that of their users. Scarcella underscores that, while the global volume of e-commerce sales is on the raise, online sales have put the enforcement of traditional VAT/GST rules to the test, resulting in a higher risk of tax evasion and fraud. The author affirms that these types of risk are mainly associated with the qualification of taxable persons, the nature of transactions (C2C or B2C) and imports of low-value goods. In such context, platforms can play in the effective enforcement of VAT/GST rules through data sharing and enhanced co-operation between tax authorities and online marketplaces. Scarcella conducts a comparative analysis of the legislations adopted in the UK, Germany, Australia and of the EU VAT e-commerce package, assessing the main benefits and limits of such new provisions. As the author emphasises, even if there is room for improvement, provisions strengthening the role of platforms for VAT/GST enforcement are in any case a valuable measure for states to adopt in order to create a level playing field for businesses and protect national tax revenues.

4. Platforms as *Locus* of Convergence: The Interplay of Value and Values

The above sections illustrates the plethora of issues within the scope of platform value(s), and thus the fascinating and wide-ranging discussions that await the Coalition over the next few years. To guide those discussions and conclude this editorial, we would like to articulate a vision of interconnectedness between different spheres of action for platforms and regulators alike, which can help foster a harmonious development of platform law and policy-making. Notably, our suggestion is to more closely link the economic and social dimension of platform regulation.

Traditionally, economic and social regulation are treated separately due to the different rationales for intervention: one stems from the need to correct market failures, while the other is driven by concerns of protection of human rights and social solidarity[24]. There is, however, a significant potential for spill-over and cross-pollination between the two, which led to the choice of bringing them together in this volume.

First, social regulation has the essential role in creating boundaries to the market enterprise, so as to prevent the commoditisation of social values in the pursuit of efficiency. Second, the approach of regulators in the enforcement of those boundaries defines the "level playing field", and for this reason, where not effective can leave a significant margin for abuse by actors placed in a pivotal position for the provision of social value. Third, the consistent production of economic value allows operators to grow and achieve scale, which fosters their ability to deal with challenges threatening existing social values- a clear example of that being the regulation of fake news that is specifically addressed in this volume. Fourth, there is an un underexplored relationship between the attainment of economic value and the expectation of diligence for economic operators, which typically translates into an insufficient consideration of market power outside "traditional" asymmetric regulation (such as antitrust or sector-specific regimes). Fifth, and finally, there are clear spill-over effects on innovation: if the boundaries of social regulation are too fuzzy, they may deter efforts that constitute normal methods of competition in the market; conversely, if a clear path is highlighted for provision of social value, this will unlock the potential of market forces to compete and innovate on those terms.

This complex but crucial relationship between different manifestations of value should drive a healthy dose of scepticism towards economic surplus generated in ways that are inconsistent with the "social contract"[25] -for instance, profits obtained by lowering or circumventing the standards of protection of privacy (or other fundamental rights that businesses have a corporate responsibility to respect), environmental, or labour protection. This scepticism is particularly compelling in the case of the owners of large productive assets, who in regulating access to their property get not just *dominion* over things but also, and more importantly, *imperium* (or sovereignty) over people[26]. Implied in this reasoning is an acknowledgment of the importance of social values in economic regulation and of the need for a broad understanding of "economic rent," a term that refers to a situation in which a market participant enjoys durable profits considerably exceeding what is socially necessary.[27]

Platform operators may be in a position to extract such rents due to structural market conditions, for instance informational asymmetries affecting the ability of other producers to compete, and

[24] See R. Baldwin, M. Cave and M. Lodge, *Understanding Regulation: Theory, Strategy and Practice* (Oxford University Press, 2nd ed. 2011), 22; T. Prosser, 'Regulation and Social Solidarity' (2006) 33 Journal of Law and Society 364-87.
[25] See e.g. Celeste Friend, 'Social Contract,' INTERNET ENCYCLOPEDIA OF PHILOSOPHY http://www.iep.utm.edu/soc-cont/
[26] See M. Feintuck, 'Regulatory Rationales Beyond The Economic', in R. Baldwin, M. Cave and M. Lodge, *The Oxford Handbook of Regulation*, citing M. R. Cohen, 'Property and Sovereignty', 13 Cornell L. Rev. 8 (1927), 13.
[27] It is fair to acknowledge that "economic rent" is a disputed concept, due to competing theories of value in economics. The definition chosen here is taken from the labour theory of value advanced by classical economists, such as Adam Smith and David Ricardo, and subsequently by Karl Marx. The theory holds that value corresponds to the amount of labour necessary to produce a marketable commodity, where "necessary" tended to be interpreted as equal to the value of the goods and services that a working-class family needed to be kept in present occupation. See H. G. Brown, 'Economic Rent: In What Sense a Surplus?', 31 The American Economic Review 4 (1941), 833-835. This theory contrasts most notably with the subjective theory of value, introduced by the so-called "marginalist revolution", that equates value to the usefulness of a good in satisfying a want and its scarcity. See K. S. Taylor, *Human Society and the Global Economy* (2001) ch. 6, available at http://www.d.umn.edu/cla/faculty/jhamlin/4111/2111-home/value.htm

infrastructural or technological advantages that cannot easily be replicated.[28] Importantly, the choices made by platforms executives and the values that platform designers decide to bake into platforms' architectures and artificial intelligence systems have become essential to the definition of the above-mentioned such market conditions and asymmetries. In such scenario, regulatory intervention must be designed strategically to cater to the multiform nature of value(s) and not just focus on the existence of supra-competitive profits.

While maintaining an environment stimulating the production of economic value is a crucial priority for society, the law must draw boundaries to conduct that is acceptable within that notion, and economics can provide the tools to allow wider measures of welfare to be taken into account. Accordingly enlightened platform regulation should "follow the value" more holistically by: (1) reflecting a clear understanding of where and how value is created, as opposed to extracted, in the perspective of long-term societal well-being; and (2) crafting measures designed to stir the production of value in a direction that aligns the incentives of its targets with those of broader society.

In this spirit, we call for a deeper reflection on the pursuit of value(s) in platform regulation. This Special Issue was conceived to elicit concrete elements of reflection for researchers, regulators, platform providers and well-informed platform users to spark a much-needed debate. Although in some specific countries the significance of values in platform regulation has entered the political debates, this is still a very sporadic practice that, so far, has led to very few concrete results. While awareness of the power and importance of dominant platform is raising, only very narrow topics are publicly scrutinised (and, frequently, in rather confused and politicised ways, as use of the "fake news" formula in televised or social media debates tellingly explains). The aim with this project was to trace pathways upon which constructive debates can be structured. What values should be promoted by platforms and how? What values should underpin the creation and enforcement of regulatory frameworks? What forms and shapes is value acquiring in the context of platforms and how can such value identified and properly taxed? This work does not pretend to provide definitive answers, but offers helpful material for a clearer and better-informed debate.

5. References

Luca Belli, De la gouvernance à la régulation de l'Internet, (Berger-Levrault 2016)

Luca Belli, The scramble for data and the need for network self-determination, (openDemocracy, 15 December 2017) <https://www.opendemocracy.net/luca-belli/scramble-for-data-and-need-for-network-self-determination> accessed 10 October 2019

Luca Belli and Nicolo Zingales (Eds.), Platform regulations: how platforms are regulated and how they regulate us. (FGV Direito Rio 2017) <https://bibliotecadigital.fgv.br/dspace/handle/10438/19402> accessed 10 October 2019

Luca Belli, Primavera De Filippi and Nicolo Zingales, A New Dynamic Coalition on Platform Responsibility within the IGF, (Medialaws, 11 June 2014). <http://www.medialaws.eu/a-new-dynamic-coalition-on-platform-responsibility-within-the-igf/> accessed 10 October 2019

Luca Belli, Primavera De Filippi and Nicolo Zingales (eds.), Recommendations on terms of service & human rights. Outcome Document n°1 (Internet Governance Forum 2014) <https://www.intgovforum.org/cms/documents/igf-meeting/igf-2016/830-dcpr-2015-output-document-1/file accessed 10 October 2019

[28] See M. A. Peteraf, 'The cornerstones of competitive advantage: A resource- based view', 14(3) Strategic Management Journal (1993) 179–191; J. B. Barney, *Gaining and sustaining competitive advantage* (Upper Saddle River, NJ: Pearson Prentice Hall 1997)

BRICS Competition Law and Policy Centre, Digital Era Competition Law: A BRICS Perspective (2019) <https://cyberbrics.info/digital-era-competition-brics-report/> accessed 10 October 2019

Jacques Crémer. Yves-Alexandre de Montjoye. Heike Schweitzer, Competition Policy for the digital era. European Commission Directorate-General for Competition (2019

Mark Elliot and Robert Thomas, Tribunal Justice and Proportionate Dispute Resolution, [2012] 71 (2) Cambridge Law Journal

Samantha Eyler-Driscoll, Asher Schechter, Camilo Patiño, Digital Platforms and Concentration, (ProMarket and Chicago Booth Stigler Center 2019)

Lawrence Lessig, Code and Other Laws of Cyberspace (Basic Books 1999)

Rebecca McKinnon, Consent of the Networked: The Worldwide Struggle for Internet Freedom, (Basic Books 2012).

Safiya Umoja Noble, Algorithms of Oppression. How Search Engines Reinforce Racism, (NYU Press 2018)

OECD, International collaboration to end tax avoidance. Under the OECD/G20 Inclusive Framework on BEPS, over 130 countries are collaborating to put an end to tax avoidance strategies that exploit gaps and mismatches in tax rules to avoid paying tax, s.d. <https://www.oecd.org/tax/beps/> accessed 10 October 2019

Cathy O'Neil, Weapons of Math Destruction: How Big data Increases Inequality and Threatens Democracy (Crown Random House 2016).

John Ruggie, Guiding Principles on Business and Human Rights: Implementing the United Nations "Protect, Respect and Remedy" Framework, Report of the Special Representative of the Secretary-General on the issue of human rights and transnational corporations and other business enterprises (UN Human Rights Council Document A/HRC/17/31, 21 March 2011) <www.ohchr.org/Documents/Publications/GuidingPrinciplesBusinessHR_EN.pdf> accessed 10 October 2019

Tunis Agenda for the Information Society (18 November 2005). WSIS-05/TUNIS/DOC/6(Rev. 1)-E. <https://www.itu.int/net/wsis/docs2/tunis/off/6rev1.html> accessed 10 October 2019

Jamila Venturini et al. Terms of service and human rights: an analysis of online platform contracts. (Revan, in collaboration with the Council of Europe and FGV Direito Rio 2016) <https://bibliotecadigital.fgv.br/dspace/handle/10438/19402> accessed 10 October 2019

Nicolo Zingales and Luca Belli, Dynamic Coalition on Platform Responsibility: Report of the "inception" meeting at the 2014 IGF, (Internet Governance Forum 2014) <https://www.intgovforum.org/multilingual/index.php?q=filedepot_download/4905/631> accessed 10 October 2019

Governing Digital Societies: Private Platforms, Public Values.

José van Dijck

Distinguished University professor in media and digital societies, Utrecht University (The Netherlands).

1. Introduction

Online digital platforms have deeply penetrated every sector in society, disrupting markets, labor relations and institutions, while transforming social and civic practices. Moreover, platform dynamics have affected the very core of democratic processes and political communication. After a decade of platform euphoria, in which tech companies were celebrated for empowering ordinary users, problems have been mounting over the past three years. Disinformation, fake news, and hate speech spread via YouTube, Twitter, and Facebook poisoned public discourse and influenced elections. The Facebook—Cambridge Analytica scandal epitomized the many privacy breaches and security leaks dogging social media networks. Further compounded by charges of tax evasion and the undermining of fair labor laws, big tech companies are facing a serious 'techlash'. As some argued, the promotion of longstanding public values such as tolerance, democracy, and transparency are increasingly compromised by the global 'exports' of American tech companies which dominate the online infrastructure for the distribution of online cultural goods: news, video, social talk, and private communication (Geltzer & Gosh, 2018).

The digitization and 'platformization' of societies involve several intense struggles between competing ideological systems and their contesting actors, prompting important questions: Who should be responsible for anchoring public values in platform societies that are driven by algorithms and fueled by data? What kind of public values should be negotiated? And how can European citizens and governments guard certain social and cultural values while being dependent on a platform ecosystem which architecture is based on commercial values and is rooted in a neo-libertarian world view?

2. The Platformization of European Digital Space

Europe has become increasingly dependent on the American platform ecosystem dominated by the Big Five tech companies (Google-Alphabet, Amazon, Facebook, Apple, Microsoft), which techno-commercial architecture is rooted in neoliberal market values. But beyond market value, the platform ecosystem revolves around societal power and influence. The Big Five increasingly act as gatekeepers to all online social traffic and economic activities; their services influence the very texture of society and the process of democracy. In other words, they have gained rule-setting power. There have been many clashes between American tech companies and European regulators as well as national legislators over public values, including privacy (resulting in the GDPR), fair competition (resulting in the EU levying substantial fines on Google-Alphabet), tax evasion (resulting in Facebook changing its tax base policy), and the condemnation of fake news and hate speech (resulting in the German parliament imposing a 24-hour deadline on social networks to take down such expressions).

We often hear from Silicon Valley CEOs that Europe is 'cracking down' on American Big Tech out of 'jealousy' (Solon, 2018). I take a different stance on this issue: the American platform ecosystem hardly allows for public space on the internet and tends to favor commercial benefits

and private interests over public ones. Therefore, Europe should articulate its own governance strategy based on its appraisal of a strong public sector, independent institutions, fair taxation, and the common good. Protecting the Rhineland model of a social market economy should not be considered an economic liability but rather an asset: a loss of public trust is ultimately a loss of business value[29]. As Mariana Mazzucato (2018) argues, it is important to assess what constitutes *societal* value in addition to market value, because both types of values are integrally part of a nation's economic strength.

Platformization has disrupted not just markets and sectors, but has started to uproot the infrastructural, organizational design of societies (Helmond, 2015; Plantin et al. 2016). It is crucial to study *how* platform ecosystems operate, because we know very little about big platforms' technical operations, their governance and business models—partly as a result of those being trade secrets (Van Dijck, 2013). As we explain in our recent book, the Big Five operate about seventy strategic *infrastructural platforms*: social networks, web hosting, pay systems, login and identification-services, cloud services, advertising agencies, search engines, audiovisual platforms, map and navigating services, app stores, analytics services, and so on (see also Van Dijck, Poell & De Waal, 2018, chapter 1). Together, these infrastructural platforms form the backbone of an ecosystem that is boundary-and-border-agnostic. Besides owning and operating a core of infrastructural platforms, the Big Five are also branching out in a variety of *sectors* that are progressively interwoven with this online infrastructure.

Indeed, platformization affects *all* sectors in society, both private (e.g. transport, finance, retail) and public (e.g. education, health), hence also affecting the common good. Power is exercised *between* infrastructural and sectoral platforms, as well as *across sectors*. Tech companies leverage control over data flows and algorithmic governance not just through operating a few major infrastructural platforms (e.g. Alphabet-Google in Search and Cloud services) but by extending these powers across many sectors (e.g. Google Apps for Education, Google Health, Google Shopping, etc). Unprecedented network effects across the global online ecosystem are thus gained through the potential of horizontal, vertical, and 'diagonal' integration of data flows, creating user lock-ins and path-dependency.

The platform mechanisms underpinning the ecosystem are largely opaque and out of sight for users and governments. Platformization is overwhelmingly driven by commercial interests which often take precedence over societal values. Some of the main problems are an almost total lack of transparency into how data flows are steered within and between sectors, how algorithms influence user behavior, how selection mechanisms impact the visibility of content, and how business models favor economic transactions over the public interest. In addition, public sectors that historically serve and protect the common good, such as education and health, are rapidly encapsulated in the American platform ecosystem, where they risk to be turned into privatized commodities. Platform companies inadvertently take over vital functions from state and public bodies once they become major gatekeepers in the circulation of health and educational data flows as well as in news and information cycles. Platforms thus increasingly become the new infrastructural providers. As Mark Zuckerberg observed in 2017, Facebook wants to be a 'social infrastructure'—a term that resonates with the notion of public utilities. Global social infrastructures, as we know, come with awesome responsibilities not just for the welfare of the company and its shareholders, but for the wellbeing of the people as societal stakeholders.

3. Who is responsible for public values and the common good?

[29] According to Peters and Weggeman (2010), the Rhineland model presumes an active government that is involved in major social issues, such as minimizing poverty and environmental protection, advocating a strong public sector and government regulation and enforcement.

If European societies want to guard public values and the common good in an online world, they first need to articulate *what kind of public values* they want to foreground when designing an ideal digital society. Norms and values are often left implicit. Looking at regulator's disputes with tech companies over the past few years, it seems clear that values such as privacy, security, accuracy, and transparency are at stake. Europeans insist on protecting their private information, securing their internet access, relying on accurate information, and pursuing transparency in terms of service. But beyond these principles relating directly to the internet as a digital environment, there is also a need to articulate values that pertain to much broader societal issues, such as democratic control of the public sphere, a level playing field for all actors, anti-discrimination practices, fairness in taxation and labor, and clarity with regards to (shared) responsibility and accountability. Public values are not a simple set of rules that you can buy 'off the shelf' and implement in society; on the contrary, they are disputed and negotiated at every level of governance – from schools and hospitals to local city councils, and from national governments to supra-national legislators.

The negotiation of public values is historically anchored in institutions or sectors, where—after extensive deliberation—they are moored in laws, agreements, or professional codes. For instance, in news journalism, public values such as accuracy and fairness in reporting are (self-) regulated via professional codes; in education, the norms for privacy, fairness and accessibility are controlled partly by the government and partly by a school's agreements with parents; urban transport is regulated by city councils and local governments. Over the past decade, platform companies have preferred to bypass institutional processes through which societies are organized – sectoral regulation, public accountability, and responsibility – by claiming their exceptional status[30]. Facebook, Google, Uber and other big platforms have argued they are mere 'facilitators', connecting users to creators or producers, and connecting content to users; insisting on their status as 'connectors' and avoiding regular legal categories, platforms and their operators have avoided taking responsibility. Until 2017, Facebook firmly denied its functioning as a 'media company' although more than half the news consumed by Americans comes to them through Newsfeed. And Uber's refusal to accept its status as a 'transportation company' was fought all the way up to the European court, where it was finally confirmed in December 2017.

So who *is* responsible for guarding public values in a digital society? The European Rhineland model ideally balances off the powers of state, market, and civil society actors in multi-stakeholder organizations. Obviously, these multiple stakeholders do not have the same interests, so government bodies need to take the roles entrusted to them as legislator, regulator, moderator, and enforcer to negotiate the public interest. However, because the architecture of the American ecosystem is uniquely engineered by market actors—and its infrastructure is dominated mostly by the Big Five—it is difficult for state and civil society actors in Europe to put their stamp on these negotiations. Governing the platform society has turned out to be a big struggle over public values and the common good.

Most visible to the public eye are the outcomes of a wide range of negotiation battles; the concerns underlying these negotiations involve a variety of public values, but it is not always immediately evident what the common denominators are. We read about EU-regulators levying big fines upon American tech firms, and understand this is about the principle of 'fair access' and a 'level playing field' of markets. We witness national governments like Germany impose strict rules on social networks to ban hate speech and fake news; of course, such judgement involves a fine balancing act between the right to free speech vis-à-vis the public values of accuracy, fairness, and

[30] This exceptional status has a legal basis in Section 230 of the American Communication Decency Act of 1996, which provides immunity from liability for providers and users of an "interactive computer service" who publish information provided by third-party users.

nondiscrimination. Cities like Amsterdam and Barcelona have set limits to short-term online rentals, curbing the free reign of Airbnb while protecting a fair housing market and livable cities. Municipalities, schools, and hospitals negotiate contracts with big tech giants such as Google to exchange data for platform services while bartering their citizens', students', and patients' right to privacy and accessibility. Each negotiation between private platform companies, government agencies, independent institutions, and citizens discloses how interests sometimes clash, sometimes converge when negotiating public values. Many of these tradeoffs boil down to a set of fundamental questions such as: who owns and exploits data flows, who controls algorithmic governance, and who is *responsible* and *accountable* for their impact?

4. Conclusion

The ideal platform society does not exist, and it will be hard to recalibrate the Western-European Rhineland model to make it fit with the American ecosystem's infrastructural architecture that privileges commercial values over public ones. Indeed, its architecture is currently firmly cemented in an American-based neoliberal set of principles that defines its operational dynamics. If European countries and the EU as a supra-national force want to secure their ideological bearings, they need to understand the ecosystem's underpinning mechanisms before they can start fortifying their legal and rinstitutional structures built on it. The implications of platformization on societies are profound, as platform ecosystems are shaping not only norms and values, but the very fabric of society.

Governing digital societies in Europe takes a serious effort at all levels, from local municipalities to national governments, from schools to collaborating universities, and from city governments to the European Parliament. European countries need to realize the limitations and possibilities of these competing networked infrastructures and articulate their position in the wake of emerging online superpowers (such as China, India, and of course the US) which ideologies and value systems are substantially different. Public values and the common good are the very stakes in the struggle over platformization around the globe. Viewed through a European looking glass, governments at all levels, independent public institutions, and nonprofits can and should be proactive in negotiating those values on behalf of citizens and consumers. Implementing public values in the technological and socio-economic design of digital societies is an urgent European challenge which cannot be left to companies alone. If we want the internet to remain a democratic and open space, it requires a multi-stakeholder effort from (supra-) national and local governments, companies, civil society organizations, and citizens; legislation is and should be the result of value-negotiations between all actors who are jointly responsible for governing our digital societies.

5. References

Geltzer, J. & Ghosh, D. Tech companies are ruining America's Image. Foreign Policy 14 May 2018. https://foreignpolicy.com/2018/05/14/tech-companies-are-ruining-americas-image/

Helmond, A. The Platformization of the Web: Making Web Data Platform Ready. *Social Media & Society* 1, no. 2 (2015). http:// journals.sagepub.com/ doi/ pdf/ 10.1177/ 2056305115603080

Mazzucato, M. *The Value of Everything. Making and Taking in the Global Economy*. New York: Allen Lane, 2018.

Peters, J., and M.Weggeman. *The Rhineland Model. Reintroducing a European Style of Organization*. Amsterdam: Business Contacts, 2010

Plantin, J. C., C. Lagoze, P. N. Edwards, and C. Sandvig. Infrastructure Studies Meet Platform Studies in the Age of Google and Facebook. *New Media & Society* (2016): 1– 18. Available at: http:// journals.sagepub.com/ doi/ 10.1177/ 1461444816661553.

Solon, O. Peter Thiel: 'Europe is cracking down on Silicon Valley out of "jealousy".' *The Guardian*, 10 March 2018. https://www.theguardian.com/technology/2018/mar/15/peter-thiel-silicon-valley-europe-regulation

Van Dijck. *The Culture of Conectivity. A Critical History of Social Media*. New York: Oxford University Press, 2013.

Van Dijck, J., Poell, T. & De Waal, M. *The Platform Society. Public Values in a Connective World.* New York: Oxford University Press, 2018.

A Constitutional Moment: How we might Reimagine Platform Governance

Nicolas Suzor[31]

We are at a constitutional moment for the future of the internet. Nation states around the world are launching major new initiatives to regulate the internet, both directly against users and by regulating the companies that provide access to telecommunications infrastructure and content services. The giant technology companies that control the bulk of the commercial internet are themselves under unprecedented scrutiny for the policies they set, the decisions they make, and the choices that go into designing their architecture. In my new book, *Lawless*, I argue that in this moment of change there is a major opportunity for us all to rethink how the internet should be governed, how power is held to account, and whose values prevail.[32]

1. Digital Intermediaries are the Focal Points of Control over the Internet

Digital intermediaries govern the internet. The telecommunications companies that provide the infrastructure, the standards organizations that design the protocols, the software companies that create the tools, the content hosts that store the data, the search engines that index that data, and the social media platforms that connect us all make decisions that impact how we communicate.[33] They govern us, not in the way that nation states do, but through design choices that shape what is possible, through algorithms that sort what is visible, and through policies that control what is permitted.[34] The choices these intermediaries make reflect our preferences but also those of advertisers, governments, lobby groups, and their own visions of right and wrong.

Technology companies now find themselves at the center of many different battles to control what people do and say online. They are the focal points of control of the internet, and governments and private organizations around the world are rapidly learning how to influence their rules and their code. These companies play a major role in governing our actions, but the power they have over us is wielded in a way that does not at all live up to the standards of legitimacy we have come to expect of governments.

Internet intermediaries enjoy a broad discretion to create and enforce their rules in almost any way they see fit. They make decisions based on their own vision for how they want users to behave, their business plans, and commercial interests, as well as in response to their exposure to legal risk and potential bad publicity. They provide little in the way of due process, leaving their users to wonder how and why decisions affecting them were made and creating deep suspicions about hidden bias and overt discrimination.[35] At the same time, nation states are rapidly learning how to influence intermediaries in order to effectively regulates users and content, sometimes in

[31] Professor, School of Law, Queensland University of Technology: n.suzor@qut.edu.au.
[32] Nicolas P Suzor, *Lawless: The Secret Rules That Govern Our Digital Lives* (Cambridge University Press 2019).
[33] Kate Klonick, 'The New Governors: The People, Rules, and Processes Governing Online Speech' (2017) 131 Harvard Law Review 1598.
[34] Tarleton Gillespie, *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media* (1st edition, Yale University Press 2018).
[35] West, Sarah Myers. "Censored, Suspended, Shadowbanned: User Interpretations of Content Moderation on Social Media Platforms." *New Media & Society*, May 8, 2018. https://doi.org/10.1177/1461444818773059.

ways that avoids or bypasses their own constitutional limitations and systems of judicial oversight.[36]

The broad discretionary powers they exercise are the antithesis to *legal* means of making decisions. The role of law in democratic societies is to create a set of rules that reflect the public interest and the morals of the populace. Laws are made legitimate through democratic institutions that are supposed to work in the public interest and constitutional limitations that protect the rights of citizens. The hallmark of legitimacy in law is the rule of law: an underpinning principle that the rules of a society should be created and enforced in a way that is predictable and fair.[37] Legislative systems are designed to ensure that the rules themselves reflect the public interest and the will of the people, and judicial systems exist as a way to check that laws are validly made and fairly enforced. Legal systems are by no means perfect, but they create the infrastructure that allows for public oversight of the rules that we live by.

Technology companies govern, but they are losing popular support. Slowly, technology companies have been losing our collective consent to govern our shared social spaces in the way that they have been.[38] The pressure to be more accountable has been building for years because technology companies have been making decisions that affect us all behind closed doors, without any real accountability.[39] It increases with every shock and controversy that casts doubt on whether the industry has our best interests at heart or is doing as much as we would like to fight all manner of bad actors online.[40] This pressure is fed by established media industries whose power has been disrupted by decentralization, and by governments who want to protect their citizens from the dangers of the internet or to better control the flow of information. All this pressure is building to a moment of profound potential change that opens up the possibilities to imagine new forms of regulating the internet.

2. A New Constitutionalism

Because online intermediaries play such a crucial role in regulating how users behave, we should find a way to ensure that their decisions are legitimately made. For this, we need a 'digital constitutionalism'.[41] Traditional constitutionalism focuses on power exercised by the state and is not well adapted to ensuring that the decisions of private actors are legitimately made. A more modern view of regulation can help us to understand that the type of power that intermediaries exercise over users is a type of governance power and that this power is subject to influence by a wide range of different actors.[42] In short, digital constitutionalism requires us to develop new ways of limiting abuses of power in a complex system that includes many different governments,

[36] Niva Elkin-Koren and Eldar Haber, 'Governance by Proxy: Cyber Challenges to Civil Liberties' (2016) 82 Brooklyn Law Review 2016; Michael D Birnhack and Niva Elkin-Koren, 'The Invisible Handshake: The Reemergence of the State in the Digital Environment' (2003) 8 Va. JL & Tech. 6.
[37] Brian Z Tamanaha, *On the Rule of Law: History, Politics, Theory* (Cambridge University Press 2004).
[38] MacKinnon, *Consent of the Networked: The Worldwide Struggle For Internet Freedom* (2012).
[39] Nicolas P Suzor and others, 'What Do We Mean When We Talk About Transparency? Toward Meaningful Transparency in Commercial Content Moderation' (2019) 13 International Journal of Communication 1526.
[40] Mike Ananny and Tarleton Gillespie, 'Public Platforms: Beyond the Cycle of Shocks and Exceptions' (2016) <http://blogs.oii.ox.ac.uk/ipp-conference/2016/programme-2016/track-b-governance/platform-studies/tarleton-gillespie-mike-ananny.html> accessed 10 July 2017.
[41] Edoardo Celeste, 'Digital Constitutionalism: A New Systematic Theorisation' (2019) 33 International Review of Law, Computers & Technology 76; Dennis Redeker, Lex Gill and Urs Gasser, 'Towards Digital Constitutionalism? Mapping Attempts to Craft an Internet Bill of Rights' (2018) 80 International Communication Gazette 302; Brian F Fitzgerald, 'Software as Discourse: The Power of Intellectual Property in Digital Architecture' (2000) 18 Cardozo Arts & Entertainment Law Journal 337; Paul S Berman, 'Cyberspace and the State Action Debate: The Cultural Value of Applying Constitutional Norms to "Private" Regulation' [2000] University of Colorado Law Review 1263.
[42] Scott Burris, Michael Kempa and Clifford Shearing, 'Changes in Governance: A Cross-Disciplinary Review of Current Scholarship' (2008) 41 Akron Law Review 1; Julia Black, 'Decentring Regulation: Understanding the Role of Regulation and Self-Regulation in a "Post-Regulatory" World' (2001) 54 Current Legal Problems 103.

businesses, and civil society organizations. The difficult task of digital constitutionalism is to build consensus about how power over the internet should be shared and limited, how those limits may be imposed, and by whom.

Constitutionalism is the difference between lawlessness and a system of rules that are fairly, equally, and predictably applied.[43] There is no simple, single definition of what it means to govern legitimately. It is impossible to define, because it is a concept that depends fundamentally on context and constantly changes. People who exercise power have legitimacy because we collectively give it to them.[44] So whether social media platforms, search engines, content hosts, telecommunications companies, governments, and other entities are acting legitimately when they shape our actions and our environment depends on how much we expect from them. This is still very much up for grabs; we are still in the early days of the commercial internet, and we do not yet have an easy answer or even common agreement on the exact shape of the limits people want to see imposed on the power of tech companies.[45]

Working out what limits we, as a society, want to impose on the exercise of power in the digital age is the first challenge of digital constitutionalism. Human rights is probably the most powerful tool we have to encourage intermediaries and governments to make their governance processes more legitimate.[46] The language of human rights provides a universally agreed-upon set of values that governments and businesses should work to promote. These values — and the responsibilities that accompany them — provide a useful way of making explicit concerns over the constitution of our shared online social spaces. The voluntary component of human rights compliance is already helping to set standards for what intermediaries should do, and it provides a guide for civil society to work cooperatively to amplify the pressure for more legitimate processes. The frame of human rights can also guide governments to implement better laws, with binding legal obligations. Human rights do not enforce themselves, and they are not sufficient to hold either governments or technology companies accountable, but they do provide a common language that we can use to build consensus about what we expect from those who govern us.

Part of the answer here is for digital platforms to work more closely with democratic governments around the world to set the standards that should apply in their countries. But there is a limit to how much legitimacy can come from the actual laws of different countries. The problem is that Governments can only really set minimum standards. The policies of platforms will always play an important role in addition to the minimum standards set by laws of the various countries where they operate. At other times, platforms need the support of the international human rights community to resist pressure from nation states to regulate their networks in ways that are directly harmful. Because social media platforms and other intermediaries cannot rely on governments to set all the rules for them, they need another system to create rules that are accepted as legitimate by their users and their critics.

There is a lot of work for digital platforms and telecommunications intermediaries to do to develop new, more legitimate systems of governance. The key next steps towards improving accountability are both straightforward and very difficult. First, platforms and telecommunication

[43] Nicolas P Suzor, 'Digital Constitutionalism: Using the Rule of Law to Evaluate the Legitimacy of Governance by Platforms' (2018) 4 Social Media + Society 1.
[44] Julia Black, 'Constructing and Contesting Legitimacy and Accountability in Polycentric Regulatory Regimes' (2008) 2 Regulation & Governance 137.
[45] Robert Gorwa, 'What Is Platform Governance?' (2019) 22 Information, Communication & Society 854.
[46] Rikke Frank Jørgensen, *Framing the Net: The Internet and Human Rights* (Edward Elgar Pub 2013).

companies should develop clearer rules that are better justified, and they must start to experiment with new systems of independent review and appeals processes that adequately deal with inevitable mistakes. We don't really what how these checks and balances might look like, but now is the time for more bold new ideas and experimentation to help make social media platforms and other intermediaries worthy of our trust.[47] Intermediaries of all types should immediately improve their transparency practices, focusing on how they can help people understand decisions that affect them and their systems as a whole. They should hire human rights lawyers and empower them to review and advise about improving technical features and business practices. They will need to reach out more to others in working through some of the tough decisions they will have to make — they should cultivate stronger relationships with experts, civil society groups, government regulators, and find some new ways to encourage genuine participation from their user communities.

None of this will be easy, and it will not happen without a great deal of effort from a diverse range of stakeholders. The second challenge of digital constitutionalism is building enough consensus and enough social pressure to force technology companies to create and enforce their own constitutional limits. Rulers usually do not give up power voluntarily; the creation of constitutional limits takes an 'immense external pressure, as the result of fierce constitutional battles'.[48] We are at a constitutional moment now, where change might be possible but is by no means guaranteed. For all of us who care about how the internet is governed, now is the time to work together to hold power accountable. We need to make visible the influence that technology companies have on our lives and the influence that others have on them, in turn. We need to trace how governments and private interests regulate how we behave and communicate; what we can see and share; and how we live, love, and work through the technologies that we use. And then we will need to mobilize. We will need to seize this moment to marshal and coordinate pressure on technology companies to fundamentally change their cultures — to recognize that, as powerful governors of our social lives, they owe us real accountability. At the same time, we need to resist the efforts of governments around the world to introduce new restrictions that unjustifiably limit our freedoms or threaten the conditions for autonomy and innovation that can make the internet so great.[49]

All of this means that we need new collaborations. We do not yet have the institutions that are able to regularly and consistently hold power to account at scale. A digital constitutionalism requires not just change from platforms, but new structures that can monitor compliance and address wrongdoing. There is a role for courts and legislatures here, but there is also a need for new institutions that can more effectively marshal social pressure in day to day governance where the legal system is too cumbersome.[50] These new institutions require some imagination — we will have to invent them. If a new constitutionalism is to be effective, then academics, activists, journalists, and others will have to work together to engage tech companies, governments, and concerned users.[51]

---

[47] See, for example, Evelyn Douek, 'Facebook's "Oversight Board:" Move Fast with Stable Infrastructure and Humility' [Forthcoming] North Carolina Journal of Law and Technology <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3365358&download=yes> accessed 3 May 2019; Thomas Kadri and Kate Klonick, 'Facebook v. Sullivan: Building Constitutional Law for Online Speech' [Forthcoming] Southern California Law Review <https://papers.ssrn.com/abstract=3332530> accessed 9 April 2019.

[48] Gunther Teubner, *Constitutional Fragments: Societal Constitutionalism and Globalization* (Oxford University Press 2012) 82.

[49] David Kaye, *Speech Police: The Global Struggle to Govern the Internet* (Columbia Global Reports 2019).

[50] Nathalie Maréchal, 'Ranking Digital Rights: Human Rights, the Internet and the Fifth Estate' (2015) 9 International Journal of Communication 3440.

[51] Nicolas P Suzor, Tess Van Geelen and Sarah Myers West, 'Evaluating the Legitimacy of Platform Governance: A Review of Research and a Shared Research Agenda' (2018) 80 International Communication Gazette 385.

As for concerned users, it is easy to feel disempowered, but there is great power in collective action. For all of us, it is time to continue to participate in the emerging debates about how we want our shared social spaces to be governed, to make our concerns heard to governments and technology companies, and to lend our support to the activists and the civil society organizations fighting for our rights. Achieving real change is not going to be easy, but what is at stake is the possibility of constructing an internet that is vibrant, diverse, and accountable. There is a lot of work ahead of us, but never has there been a better opportunity to make serious change than now.

From the Telegraph to Twitter: The Case for the Digital Platform Act

Harold Feld

In The Case for the Digital Platform Act[52] *I argue that digital platforms form a distinct part of the economy that requires its own rules tailored to the unique features of the digital platform space, and its own regulator to police the sector. This regulator would have jurisdiction to promote competition, protect consumers, and address questions of content moderation. Rather than creating regulatory models from scratch, governments should look to the regulation of disruptive communications technologies of the past to evaluate what worked to promote values essential to our democracy and economy, and what did not work. With this experience firmly in mind, we can adapt previous regulatory models to advance our fundamental values for society.*

Within the last year, major studies in the United States, the United Kingdom, and the European Union have underscored the growing need to regulate "digital platforms" to promote competition and protect consumers. In the United States, the Stigler Center at the University of Chicago issued a series of white papers[53] on the economic and societal impact of digital platforms. The United Kingdom released a 150-page report[54] on digital platforms from a panel of experts headed by economist Jason Furman. The European Commission issued its own expert report[55], often referred as the "Vestager Report." While the reports differ to some degree in detail and analysis, they reach consensus on several important facts.

First, the three reports recognize that digital platforms constitute a unique sector of the economy that does not behave like a traditional competitive market for goods and services. The combination of multisided market structure, the importance of personal data and information to these markets, and the strength of the network effects enjoyed by successful platforms, digital platforms can exercise both market dominance and outsized influence in society as a whole. Second, traditional rules of antitrust and existing regulatory agencies cannot keep pace with the impact of digital platforms. Many of these platforms challenge traditional definitions of what constitutes a "market," and challenge traditional definitions of consumer harm. Third, all three reports suggest the need for some sort of "digital authority" or regulatory agency to address both concerns about competition and broader societal concerns.

In my book *The Case for the Digital Platform Act*, I propose the creation of such a comprehensive "digital authority." The good news is, we have been here before. Although the first electronic communications technologies of over a century ago -- the telegraph and radio communication – seem primitive and remote from the social media and online shopping sites of today, these technologies has similar disruptive effects on the world and raised similar difficult questions as to the nature of sovereignty in a world where instant communication enabled economic transactions and the flow of information to take place globally at breakneck speed. At the same time, the rise in these technologies enabled the rise of monopolization of information and the commerce that depended on the new, global and broadcast communications.

---

[52] Harold Feld, The Case for the Digital Platform Act*: Breakups Starfish Problems and Tech Regulation* (Roosevelt Institute: New York City 2019) https://www.digitalplatformact.com/ accessed on 15 October 2019

[53] See https://research.chicagobooth.edu/stigler/events/single-events/antitrust-competition-conference/digital-platforms-committee accessed on 15 October 2019

[54] Digital Competition Expert Panel, Unlocking digital competition. Report of the Digital Competition Expert Panel (March 2019) https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/785547/unlocking_digital_competition_furman_review_web.pdf accessed on 15 October 2019

[55] , Jacques Crémer. Yves-Alexandre de Montjoye. Heike Schweitzer, Competition Policy for the digital era. European Commission Directorate-General for Competition (2019).

The rise of foreign correspondents able to route around government censorship was accompanied by the rise of fake news, and ultimately the monopolization of international news by the Associated Press. The telegraph made possible peace negotiations and trade, but it also enabled those who controlled the telegraph to rig elections and manipulate markets. Radio made culture and news available in every home. But it also made the wild spread of race hatred and philosophies such as Nazism and Facism. We cannot ignore that Facebook enabled the genocidal assault against the Rohinga in Mynmar. But we should not forget that the radio and telephone made Kristallnacht possible in 1938, and the genocide of Tutsis by Hutus possible in 1994.

The point of reexamining this history is neither to invoke a feeling of futility that we can never solve these problems, or to suggest that simply lifting the more successful policies from the last 150 years of communications regulation will neatly fit with the latest iteration of disruptive communications technology. But the history does confirm for us that the same *type* of policy problems tend to reemerge time and again because certain important facts remain the same. The importance of communications in all areas of human activity from culture to commerce makes the rules governing the communications sector uniquely important. The power of "network effects" and other economic factors that make the sector uniquely susceptible to concentration. Finally, while open networks enhance to possibility for democratic discourse and promote the ability of traditionally marginalized communities and individuals to engage in the public sphere, open networks also allow bad actors to spread misinformation and undermine both democracy and the rule of law.

Armed with the knowledge of how we successfully (or in some cases, unsuccessfully) met these challenges in the age of the telegraph, the telephone, radio and television, we can now consider how to meet these challenges again in the age of broadband and social media. As before, our dialog must be guided by a firm understanding of fundamental values that serve the public interest, rather than abandoning our responsibility to govern ourselves to "free markets" run by monopolists, oligarchs or cartels.

Critically, we must not separate regulation designed to promote competition from that designed to protect consumers, public safety, and democratic principles. As the history of communications demonstrates, these matters are inextricably intertwined, and sometimes require balancing and trade offs that should be made by governments, not outsourced to the private sector. Enhancing competition through data portability may make it more difficult to protect privacy, whereas laws designed to hold platforms accountable for content may result in preserving in place the existing industry structure in ways that stifle competition or that create barriers to civic engagement for the disadvantaged and marginalized.

For example, requiring broadcasters to have government licenses made them more accountable, but it excluded all but the wealthiest from broadcasting and excluded content unacceptable to the mainstream. It is inevitable that policymakers will need to make tradeoffs between fundamental social values. These should be made mindfully and carefully by an expert agency empowered to address any unintended outcomes or changes in technology, rather than divided among several agencies with competing agendas and only a partial understanding of the problem.

1. Where Do We Go From Here?

The Case for the Digital Platform Act provides enough concrete detail to begin this debate in a meaningful way. The book proposes a universal definition of "digital platform" necessary to construct a sector-specific regulator, capable of providing adequate sector-specific oversight to promote competition, address the issues of content moderation, adequately protect consumers, and provide for public safety. Because digital platforms often challenge traditional concepts of antitrust markets and traditional metrics of market power, *The Case for the Digital Platform Act*

proposes a new metric, the “cost of exclusion” (CoE). Recognizing that the power of digital platforms derives from the powerful network effects possible with flexible, multi-sided markets, CoE measures the cost to an individual or business based on exclusion from the platform. This measure is designed to be flexible enough to apply both to the economic cost of exclusion and to the cost to individuals from being excluded from platforms increasingly central to civic discourse and the public sphere.

The book reviews the wide area of pro-competitive tools used globally to transition from monopoly telecommunications and media regulation to a competitive framework that also serves the public interest. Rather than recommend a specific solution, the book recommends that governments empower a specific national regulator capable of addressing the myriad of issues created by digital platforms in a comprehensive manner. Although written by an American advocate for the United States Congress and American regulators, *The Case for the Digital Platform Act* offers a useful starting framework for global policymakers as well. To distill the recommendations to four simple bullet points, *The Case for the Digital Platform Act* advises governments to:

- Embrace comprehensive sector-specific regulation by empowering a single regulator to oversee digital platforms, rather than using multiple agencies to apply a mix of competition policy, content moderation policy, consumer protection, and public safety. Only by recognizing the unique nature of digital platforms can governments ensure comprehensive and appropriate policy across the board.

- Provide guidance to the sector regulator based on enduring values of promoting competitive markets, protecting consumers, encouraging free expression and news production while protecting vulnerable members of society from harassment, and utilizing the capacities of the sector to protect public safety.

- Governments should recognize that while we must not allow the complex nature of the technology and the difficult social and economic tradeoffs to freeze us into immobility, we should not rush to pass broad laws of general applicability that will generate unforeseeable consequences. Rather, governments should provide the new sector regulator with a broad array of tools to address the myriad issues raised by such an important and diverse sector of the economy.

- We must recognize that effective, comprehensive regulation of such an important sector takes time to establish. No law will work perfectly on Day 1, or even Day 1000. We will continue to revisit these important questions for the near future.

To conclude, I believe that this book provides the right questions, and the right framework for answering these questions. It is my hope that we can move forward in a way that centers our enduring values, and ensures that digital platforms will serve the public interest.

The New City Regulators:

Platform and Public Values in Smart and Sharing Cities

Sofia Ranchordás[*] and Catalina Goanta[**]

Abstract

Cities are increasingly influenced by novel and cosmopolitan values advanced by transnational technology providers and digital platforms, which differ from the traditional public values protected by national and local laws and policies. This article contrasts the public values created by digital platforms in cities with the democratic and social national values that the platform society is leaving behind. It innovates by showing how co-regulation can balance public values with platform values. In this article, we argue that despite the value-creation benefits produced by the digital platforms under analysis, public authorities should be aware of the risks of technocratic discourses and potential conflicts between platform and local values. In this context, we suggest a normative framework which enhances the need for a new kind of knowledge-service creation in the form of local public-interest technology. Moreover, our framework proposes a negotiated contractual system that seeks to balance platform values with public values in an attempt to address the digital enforcement problem driven by the functional sovereignty role of platforms.

Keywords: digital platforms; smart cities; Internet-of-things; public values; privacy; urban law

1. Introduction

The digital revolution is not only a technological revolution, but it is primarily a revolution of powers and values.[56] In the last decade, it has become clear that the services facilitated by digital platforms (*e.g.*, Facebook, Airbnb, Google, Uber) were not as value-neutral, unbiased, and impartial as they originally claimed.[57] Rather, digital platforms are now well-known for being self-serving, opaque, and imbued with values that determine the types of services offered, shape the interactions between users and service providers, and define who has a voice and who does not.[58] Yet, their central role in promoting innovation and growth, creating new communication opportunities, and removing market entry barriers to small and medium enterprises is indisputable.[59] Thus far, it has remained nonetheless challenging to establish the precise value created by these platforms, how the values conveyed by these platforms differ from national public

[*] Professor of European and Comparative Public Law and Rosalind Franklin Fellow, Faculty of Law, University of Groningen, The Netherlands.We would like to thank the anonymous reviewers for their insightful comments as well as Luca Belli, Nicolo Zingales, and Luã Fergus Oliveira da Cruz.
[**] Assistant Professor of Private Law and postdoctoral Researcher at Studio Europa (Maastricht Working Group on Europe), Maastricht University, Maastricht Private Law.

[56] Benoît Thieulin, 'Gouverner à l'heure de la revolution des pouvoirs' (2018) 164 *Pouvoirs* 19.
[57] Lucas Introna and Helen Nissenbaum, 'Shaping the Web: Why the Politics of Search Engines Matters' (2000) 16 *The Information Society* 169; Frank Pasquale, 'Platform Neutrality: Enhancing Freedom of Expression in Spheres of Private Power' (2016) 17 *Theoretical Inquiries in Law* 487.

[58] See for instance, Christopher S. Yoo, 'Free Speech and the Myth of the Internet as an Unintermediated Experience' (2010) 78 *George Washington Law Review* 697. See also the developments regarding content moderation and intermediary liability at European Union level, Opinion of Advocate General Szpunar, Case C‑18/18, Glawischnig-Piesczek v Facebook Ireland Limited, ECLI:EU:C:2019:45; Daphne Keller, 'Dolphins in the Net: Internet Content Filters and the Advocate General's Glawischnig-Piesczek v. Facebook Ireland Opinion' (*Stanford Center for Internet and Society,* 4 September 2019) <https://cyberlaw.stanford.edu/files/Dolphins-in-the-Net-AG-Analysis.pdf>.

[59] Communication from the European Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: Tackling Illegal Content Online. Towards an Enhanced Responsibility for Online Platforms. COM (2017) 555 final.

values, and whether they are contributing to the emergence of a novel source of values parallel to those of domestic law and policy.[60]

A growing number of scholars from different fields has delved into the phenomenon of "platformization" which seeks to convey the impact of digital platforms on cultural industries, politics, and the economy.[61] Legal scholars have contributed to this body of scholarship by explaining in general terms how the growing emergence of platform power and values is threatening fundamental rights, competition rules, and democracy.[62] This strand of literature has particularly delved into the shortcomings of technology (*e.g.*, opacity, complexity, biased algorithmic decision-making or discrimination).[63] However, scholars may sometimes overlook that the impact of digital platforms is also experienced in the physical world at the most basic and local levels. Citizens see their neighborhoods depleted of affordable houses due to the rise of Airbnb or alike tourist accommodation, are surrounded by e-scooters and are often affected by the accidents, urban nuisance and vandalism that has accompanied their proliferation.[64] At the same time, citizens also experience the prevalence of digital platforms at other levels as the number of digital municipal services provided by sophisticated platforms grows, or they realize that the tech companies contracted by their cities collect data on every single step they take.[65] The recent destruction of multiple smart lightposts in Hong Kong in August 2019 as part of the demonstrations against the local government, show the citizens' growing rejection of this loss of privacy.[66] In spite of these developments, the impact of platform power and values at local level has nonetheless remained overlooked.[67] The reliance on digital technology provided by Big Tech companies (*e.g.*, Google/Alphabet) is not only putting the protection of human rights at stake, but it is also changing

[60] Luca Belli, Pedro Francisco and Nicolo Zingales, 'Law of the Land or Law of the Platform? Beware of the Privatisation of Regulation and Police' in Luca Belli and Nicolo Zingales (eds), *Platform Regulations. How Platforms Are Regulated and How They Regulate Us* (FGV, 2017) 59; Eyal Benvenisti, 'Upholding Democracy Amid the Challenges of New Technology: What Role for the Law of Global Governance?'(2018) 29 European Journal of International Law 9; On platform values, see José van Dijck, Thomas Poell and Martijn de Waal, *The Platform Society: Public Values in a Connective World* (Oxford University Press 2018).

[61] See *e.g.*, Vera Demary, The Platformization of Digital Markets, IW Policy Paper 39/2015, available at https://www.iwkoeln.de/fileadmin/publikationen/2015/257401/Digital_Markets_policy_paper_IW_Koeln.pdf; David B. Nieborg and Thomas Poell, 'The Platformization of Cultural Production: Theorizing the Contingent Cultural Commodity'(2018) 20 (11) *New Media & Society* 4275; David B. Nieborg and Anne Helmond, 'The Political Message of Facebook's Platformization in the Mobile Ecosystem: Facebook Messengers as a Platform Instance'(2019) 41 (2) *Media, Culture & Society* 196; José van Dijck, Thomas Poell and Martijn de Waal, *De Platformsamenleving* (Amsterdam University Press 2016) 17.

[62] See for instance Orla Lynskey, 'Regulation by Platforms: The Impact on Fundamental Rights' in Luca Belli and Nicolo Zingales (eds), *Platform Regulations. How Platforms Are Regulated and How They Regulate Us* (FGV 2017).

[63] See for instance Frank Pasquale, *The Black Box Society: How Secret Algorithms Control Money, Information* (Cambridge, MA: Harvard University Press, 2016); Nicolas P. Suzor, *Lawless: The Secret Rules that Govern our Digital Lives* (Cambridge: Cambridge University Press, 2019). See also Rathenau Institute, Digitaliseren vanuit publieke waarden (*Rathenau Institute*, 6 March 2019) <https://www.rathenau.nl/nl/digitale-samenleving/digitaliseren-vanuit-publieke-waarden>; Tarleton Gillespie, The Relevance of Algorithms, in Tarleton Gillespie, Pablo J Boczkowski, Kriste A Foot (eds) *Media Technologies: Essays on communication, materiality and society* (MIT Press 2014) 167-194.

[64] James A. Allen, 'Disrupting Affordable Housing: Regulating Airbnb and Other Short-Term Rental Hosting in New York City'(2017) 26 Journal of Affordable Housing and Community Development Law 151; Dayne Lee, 'How Airbnb Short-Term Rentals Exacerbate Los Angeles's Affordable Housing Crisis: Analysis and Policy Recommendation'(2016) 10 *Harvard Law & Policy Review* 229.

[65] Privacy International, 'Smart cities: Utopian vision, Dystopian reality' (*Amnesty International*, 2017) <https://privacyinternational.org/report/638/smart-cities-utopian-vision-dystopian-reality>; Liesbeth van Zoonen, 'Privacy Concerns in Smart Cities' (2016) 33(3) Government Information Quarterly 472; Lilian Edwards, 'Privacy, Security and Data Protection in Smart Cities: A Critical EU Perspective' (2016) 3 European Data Protection Law Review 28.

[66] See for instance Ellen Ioanes, 'Hong Kong protesters destroyed "smart" lampposts because they fear China is spying on them' (*Business Insider*, 26 August 2019) < https://www.businessinsider.nl/hong-kong-protesters-smart-lampposts-are-spying-on-them-2019-8?international=true&r=US >. See also Raj Gaire, Ratan K Ghosh, Jongkil Kim, Alexander Krumpholz, Rajiv Ranjan, R K Shyamasundar et al., 'Crowdsensing and Privacy in Smart City Applications' in Danda B Rawat and Kayhan Zrar Ghafoor (eds) *Smart Cities Cybersecurity and Privacy* (Elsevier 2019) 57; Kati Brock, Elke den Ouden, Kees van der Klauw, Ksenia Podoynitsyna and Fred Langerak, 'Light the way for smart cities: Lessons from Philips Lighting' (2019) 142 Technological Forecasting and Social Change 194; Daniel van den Buuse and Ans Kolk, 'An exploration of smart city approaches by international ICT firms' (2019) 142 Technological Forecasting and Social Change Volume 220.

[67] For reflections on platform power exercised globally, see for instance Orly Lobel, 'The Law of the Platform' (2016) 101 Minnesota Law Review 87.

the fulfilment of the mandate of public functions, particularly because of the lack of scrutiny.[68] Therefore, it is important to understand the underlying governance choices made by public authorities and the values they decide to imbue them with.[69] This article addresses this gap by inquiring into the role and practices of digital platforms in urban centers in the contexts of smart cities and the sharing economy.[70] Both phenomena are inserted in similar recent debates on the digitalization of urban centers, the promotion of innovation, efficient allocation of urban resources, and sustainability.[71] Nonetheless, both sharing-economy and smart-city enabling platforms have been accused of not being as citizen-centric, sustainable, and protective of public values as they claim.[72] In addition, in both fields we find platforms with significant market power developed by Big Tech that have the capacity to impose their own values on public authorities. Small local smart-city and sharing-economy platforms are thus outside the present analysis, as our focus lies within Big Tech.

In the sharing economy and smart cities, platforms mediate the relationship between citizens and government, reshaping it with their private data-driven and profit-oriented values. Platforms do so because they track, collect, process, and predict information regarding cities and citizens and they support decision-making by relying on big data analysis techniques such as machine learning.[73] While focused on the influence of platforms in the regulation of local values, this article seeks to touch upon a crucial question with public policy implications: What values do platforms convey in a city, and how do they differ from public values?

In answering this question, we explore the potential conflicts between private, profit-oriented platforms whose priorities are defined by shareholders and their online communities, and the heterogeneous interests of local communities, citizens (including underrepresented minorities), and public actors.[74] In this context, we question the limited transparency of platforms and how this undermines the task of determining the underlying platform values. From a methodological perspective, this article draws its analysis on an interdisciplinary literature review (*e.g.,* law, communication sciences, business, public administration, new media studies) on smart cities, platform values and value creation, as well as on the qualitative content analysis of the terms of service of Airbnb and Lime, and the promotional materials used on the websites of Sidewalk Labs and IBM Smarter Cities as examples of sharing economy and smart city platforms.[75]

We argue that despite the value-creation benefits produced by digital platforms, public authorities should be aware of the risks of technocratic discourses and potential conflicts between

---

[68] Rikke Frank Jørgensen, 'What Platforms Mean When They Talk about Human Rights' (2017) 9 Policy and Internet 280; Lorna McGregor, 'Accountability for Governance Choices in Artificial Intelligence: Afterword to Eyal Benvenisti's Foreword'(2019) 29 European Journal of International Law 1079, 1084.

[69] Lorna McGregor, 'Accountability for Governance Choices in Artificial Intelligence: Afterword to Eyal Benvenisti's Foreword' (2019) 29 European Journal of International Law 1079.

[70] Paula Gori, Pier Luigi Parcu and Maria Luisa Stasi, 'Smart Cities and Sharing Economy' (*EUI*, 2015) <https://cadmus.eui.eu/bitstream/handle/1814/38264/RSCAS_2015_96.pdf?sequence=1&isAllowed=y>.

[71] See Duncan McLaren and Julian Agyeman, *Sharing Cities: A Case for Truly Smart and Sustainable Cities* (MIT Press 2015); Anthony Townsend, *Smart Cities: Big Data, Civic Hackers, and the Quest for a New Utopia* (W W Norton 2013).

[72] For a criticism of smart cities, see for instance, Jiska Engelbert, Liesbet van Zoonen and Fadi Hirzalla, 'Excluding citizens from the European smart city: The discourse practices of pursuing and granting smartness' (2019) 142 Technological Forecasting and Social Change 347; Paolo Cardullo and Rob Kitchin, 'Smart Urbanism and Smart Citizenship: The Neoliberal Logic of 'Citizen-Focused' Smart Cities In Europe' (2018) 37 Environment and Planning C: Politics and Space 813. On the sharing economy, see for instance, Andrea Geissinger, Christofer Laurell, Christina Öberg and Christian Sandström, 'How sustainable is the sharing economy? On the sustainability connotations of sharing economy platforms' (2019) *206* Journal of Cleaner Production 419; Koen Frenken and Juliet Schor, 'Putting the sharing economy into perspective' (2017) 23 Environmental Innovation and Societal Transitions 3.

[73] For a general overview of machine learning, see for instance Tom Mitchell et al., 'Machine learning' (1990) 4(1) *Annual Review of Computer Science* 417.

[74] Tarleton Gillespie, *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media* (Yale University Press 2018).

[75] Klaus Krippendorff, *Qualitative Content Analysis: An Introduction to Its Methodology* (Sage 2004) 15; Marnix Snel and Janaína de Moraes, *Doing a systematic literature review in legal scholarship* (Boom Juridische Uitgevers 2017).

platform and local values.[76] It is in this context that we aim to offer a normative framework for this problem through a co-regulatory or negotiated system that seeks to balance platform and public values.[77]

This article's contribution to existing literature is twofold: first, it offers an innovative legal analysis of the broader impact of digital platforms on public values in the urban context (where platforms tend to have a stronger influence); second, it suggests a normative framework for the protection of public values, based on the notion of local public-interest technology as well as on the introduction of an obligation to take into account the broader impact of private services on public infrastructure.

The article is organized as follows. Section 2 briefly describes how digital platforms have evolved from their traditional role as business matchmakers to influential urban intermediaries. It first defines the notion of value, makes an overview of the values of the platforms under analysis on the grounds of their terms of service, and compares them with public values identified on the basis of literature and public policy documents. Section 3 explores the tension between platform and public values within the context of smart cities and the sharing economy services which have an impact on local communities. Section 4 reflects upon the role of digital platforms in smart cities and explains how these actors are conveying their values as providers of public services, to contextualize the normative framework we propose for aligning the values promoted by digital platforms and cities. Section 5 concludes.

2. Digital Platforms and Their Values

2.1. The Emergence of the Urban Platform Economy

In the early 1970s, students from Stanford University's Artificial Intelligence Laboratory concluded the world's first digital peer-to-peer transaction, using ARPANET, the Internet's precursor network, to purchase drugs from fellow students from the Massachusetts Institute of Technology.[78] Two decades later, the New York Times was writing about the first sale made on the Internet as we know it today, which entailed a transaction consisting in a Sting CD.[79] With the advent of microcomputing and the rise of Internet penetration in individual households, e-commerce became the first industry that shaped the notion of digital platforms as we currently identify them, by turning the intermediation of consumer transactions into a lucrative business model. As Internet users found more familiarity in the virtual sphere, digital platforms (*e.g.*, Google, Facebook, Twitter, eBay), already regarded as intermediaries, started providing an increasingly wider range of information society services.[80] The emergence of such intermediaries contributed to the development of online transactions, as these 'matchmakers' hosted information, facilitated the intermediation of transactions between strangers, and matched supply and demand.[81] About a decade ago, Airbnb, Uber, Lyft, and other 'sharing-economy' platforms (broadly defined) started relying on this technology to offer services that would allow individuals to share their apartments, vehicles or other goods with strangers.[82] The sharing-economy

[76] See also Sara Hofman et al., 'The Public Sector's Role in the Sharing Economy and the Implications for Public Values' (2019) *Government Information Quarterly* (forthcoming).
[77] For a very brief analysis of co-regulation in the context of an earlier form of the sharing economy, see Michele Finck and Sofia Ranchordas, 'Sharing and the City' (2016) 49 Vanderbilt Journal of Transnational Law 1299.
[78] John Markoff, *What the Dormouse Said: How the Sixties Counterculture Shaped the Personal Computer Industry* (Penguin 2006) 20. See also Jamie Bartlett, *The Dark Net* (Melvile House 2014).
[79] Peter H Lewis, 'Attention Shoppers: Internet Is Open' (*New York Times*, 21 August 1994) <https://www.nytimes.com/1994/08/12/business/attention-shoppers-internet-is-open.html>.
[80] See for instance, David Evans, 'The Antitrust Economics of Two-Sided Markets' (2003) 20(2) Yale Journal on Regulation 327; also more recently, David Evans and Richard Schmalensee, *Matchmakers: The New Economics of Multisided Platforms* (Harvard Business Review Press 2016).
[81] Evans and Schmalensee fn 25.
[82] For early scholarship on the implications of the sharing economy see for instance Charles Gottlieb, 'Residential Short-Term Rentals: Should Local Governments Regulate the Industry' (2013) 65 Planning & Envtl L 4; Anne-Marie Slaughter, 'Filling Power Vacuums in the New Global Legal Order' (2013) 36 B C Int'l & Comp L Rev 919.

disrupted at first existing regulated sectors (*e.g*., hotels, taxis) and was at the outset of significant litigation throughout the world. In the last eight years, the sharing economy, the 'gig economy' or the platform economy have occupied hundreds of legal scholars throughout the world.[83] As national and local governments start bending or revising their legal frameworks to address the challenges of unregulated sharing or 'gig' services, legal literature has shifted its interest to other topics.[84] However, this shift overlooks one of the key impacts of the platform economy that has become visible with the consolidation of this new economic system: sharing-economy platforms are changing the landscape of cities and have a profound influence on local values.[85]

Although it has been clear for almost two decades that digital platforms would change our economy, platforms have thus far been regulated as value-neutral hosts of information.[86] This traditional view no longer encompasses the current impact of digital platforms on our society, economy, and politics.[87] Digital platforms have become regulators, vehicles of communication, innovation, online dispute resolution, and value creation.[88]

Big Tech platforms in particular have become the new essential infrastructures for information, economic and political influence.[89] Although this phenomenon should not come as a surprise, the growing power of private platforms at the local level is nonetheless problematic for three reasons. First, urban centers with the ambition to become smart cities are currently partnering up with Big Tech to contract not only for software, but also to implement interconnected digital sensors and systems that influence the way cities are planned, how citizens move in a city, and the type of services offered.[90] While, for example, Huawei offers useful digital platforms for cities, it is also well-known that this company has been under investigation in different countries on suspicion of espionage and alleged trade-secret theft.[91] This extreme example does not necessarily reflect the practices of other Big Tech companies, but it helps us illustrate the risks of a potential misalignment between public and private interests, the existence of hidden interests, and the lack of transparency of digital platforms. It is this lack of transparency that extends to the relationship between public authorities and private platforms not only in the context of public procurement and outsourcing of IT-services but also when it comes to regulated platforms. Another illustration is the relationship between Airbnb and the municipality of Amsterdam. In the early days of home-sharing in Amsterdam, the municipality celebrated a confidential

[83] See for instance Vanessa Katz, 'Regulating the Sharing Economy' (2015) 30 Berkeley Tech LJ 1067; John O McGinnis, 'The Sharing Economy as an Equalizing Economy' (2018) 94 Notre Dame L Rev 329; Erez Aloni, 'Pluralizing the Sharing Economy' (2016) 91 Wash L Rev 1397; Matthew T Bodie, 'Lessons from the Dramatists Guild for the Platform Economy' (2017) 2017 U Chi Legal F 17; Leah Wing, 'Social Impact and Technology: Issues of Access, Inequality and Disputing in the Collaborative Economy' (2014) 1 IJODR 150; Michael Todisco, 'Share and Share Alike: Considering Racial Discrimination in the Nascent Room-Sharing Economy' (2014-2015) 67 Stan L Rev Online 121; Irina Domurath, 'Platforms as Contract Partners: Uber and beyond' (2018) 25(5) Maastricht Journal of European and Comparative Law 565; Christoph Busch, 'The Sharing Economy at the CJEU: Does Airbnb Pass the Uber Test' (2018) 7(4) Journal of European Consumer and Market Law 172; Sofia Ranchordas, 'Peers or Professionals: The P2P-Economy and Competition Law' (2017) 1 Eur Competition & Reg L Rev 320.

[84] See for instance Andrew G Malik, 'Worker Classification and the Gig-Economy' (2017) 69 Rutgers UL Rev 1729, 1745; Michael L Nadler, 'Independent Employees: A New Category of Workers for the Gig Economy' (2018) 19 NC JL & Tech 443, 445.

[85] See for instance Sarah Kessler, 'The "Sharing Economy" Is Dead, And We Killed It' (*Fast Company*, 14 September 2015) <https://www.fastcompany.com/3050775/the-sharing-economy-is-dead-and-we-killed-it>.

[86] See for instance, Richard Posner, 'Antitrust in the New Economy' (2000) 68 *Antitrust L. J.* 925.

[87] Van Dijck, Van Poell and de Waal fn 8.

[88] For a thorough analysis of the different roles of platforms, see Rory van Loo, 'The Corporation as Courthouse' (2016) 33 Yale J. Reg. 547, 553; Tarleton Gillespie, 'The Politics of 'Platforms' (2010) 12(3) New Media & Society 347; Tarleton Gillespie, *Moderation, and the Hidden Decisions That Shape Social Media* (Yale University Press 2018).

[89] K Sabeel Rahman, 'Regulating Informational Infrastructure: Internet Platforms as the New Public Utilities' (2018) 22 Georgetown Law Technology Review 234; Tarleton Gillespie, 'Platforms Are Not Intermediaries' (2018) 22 Georgetown Law Technology Review 198.

[90] Francisco Klauser, Till Paasche and Ola Söderström, 'Smart Cities as Corporate Storytelling' (2014) 18(3) *City* 307; Anthony Townsend, *Smart Cities: Big Data, Civic Hackers, and the Quest for a New Utopi*a (WW Norton & Company 2014) 64.

[91] Adi Robertson, 'Huawei executive accused of helping steal trade secrets' (*The Verge*, 22 May 2019) <https://www.theverge.com/2019/5/22/18636237/huawei-cnex-trade-secrets-lawsuit-eric-xu-accusation-trial-hearing>. See also Michael Plachta and Bruce Zagaris, 'Sanctions and Int'l Human Rights' (2019) 35 IELR 62, 65.

memorandum of understanding with Airbnb allowing this platform to operate temporarily "beyond local law."[92] The platform was able to establish itself very quickly to the growing discontent of Amsterdam residents and attract thousands of hosts and guests. These type of arrangements or informal partnerships are also found in the context of smart cities in the mobility sector. Lime, the global leading platform for electric scooters, has recently partnered with the cities of Omaha, Detroit, and Charlotte to reduce traffic congestion. This partnership has also been designed as a pilot to test the efficacy of micro-mobility (*e.g.*, electric scooters) to improve urban mobility. However, as this article later explains, such arrangements are not as unproblematic as they seem.

Second, as a result of the expansion of digital platforms in cities, their values and related global trends (*e.g.*, cosmopolitan tourism) appear to have started to prevail in the context of these contractual or informal 'partnerships' and over national public values, resulting in the decharacterization of neighborhoods, exclusion of residents from the city center, and gentrification of traditional urban centers. Third, as cities become imbued with platform values, we observe a new shift in the power dynamics from public authorities to private actors that do not pursue the public interest with a certain measure of democratic legitimacy. The cooperation between public and private actors results thus not only in the privatization of public services but also in the transformation of public values.[93] This phenomenon is connected to the more general problem of misalignment of public and private interests. While both public and private entities tend to attend to the interests and needs of their customers, it is well-known they do it in very different ways.[94] The boundary between the public sphere and the respective rights and duties of citizens has become thus blurred due to the growing power of digital platforms that not only offer commercial services to consumers but also disrupt once regulated services and digitize local services, reshaping the relationship between public authorities and citizens.[95]

2.2. Platforms as Generators of Values

Thus far, the notion of 'digital platform' has been used to depict BigTech intermediaries, whether within the realm of the 'sharing' or 'gig' economy, or outside it, for example on social media.[96] The core technology transacted by these companies has been software (*e.g.*, apps, online platforms) developed on the basis and/or for the enabling of big data collection. With the increase of functionalities performed by connected machines on the Internet of Things ('IoT'), we advance the idea that the next generation of digital platforms will be defined by companies that add electrical engineering expertise (*e.g.* hardware). New digital platforms will not only be present in smartphones and other personal computing devices, but also in machines from the physical public or private space that did not traditionally include a computer and urban furniture, such as

[92] This document is available at https://www.binnenlandsbestuur.nl/Uploads/2016/2/2014-12-airbnb-ireland-amsterdam-mou.pdf (last accessed on October 7, 2019). As it was initially unclear whether Airbnb fitted within existing legal qualifications and the platform did not violate directly any national or local rules, it is important to underline that the municipality merely agreed not to enforce existing rules on tourist accommodation as regards Airbnb hosts. This position has changed considerably since 2014. The platform has signed more recent agreements with the municipality to help the latter enforce new rules that restrict the maximal rental period.

[93] For a critical analysis of the privatization of city services, see Gerald Frug, 'City Services' (1998) 73 N. Y. U. Law Rev. 23, 29-30.

[94] Jean Damascene Twizeyimana and Annika Andersson, 'The public Value of E-Government—A Literature Review' (2019) 36 Government Information Quarterly 167, 168.

[95] Ida Lindgren, Christian O. Madsen, Sara Hofmann and Ulf Melin, 'Close Encounters of the Digital Kind: A Research Agenda for the Digitalization of Public Services' 1(2019) 36 Government Information Quarterly 427, 432.

[96] Social media is often not included in scholarship on the gig economy, although Youtube, launched in 2005, predates Airbnb and Uber, launched in 2008 and respectively in 2009, and it has a similar business model: connecting a broadcaster to a peer audience. *See* Catalina Goanta and Sofia Ranchordás, 'The Regulation of Social Media Influencers: An Introduction' in Catalina Goanta and Sofia Ranchordás, *The Regulation of Social Media Influencers* (Edward Elgar Publishing, 2020, forthcoming).

lightposts.[97] This article thus uses a broader definition of 'digital platforms' which also includes platforms that are developed to support different types of sensors.

Before delving into the matter of what values are conveyed by digital platforms, we must acknowledge what we mean by 'value'. 'Value' is a concept that can be interpreted in a plethora of ways, as it has importance for philosophy, economics, sociology, public administration and law, to name a few examples. In this article, we employ it to reflect on moral qualities, as values 'are the *principia* of practical thought.'[98] Given that the morality dimension implies the consideration of what is right and what is wrong,[99] the notion of 'value' as described in the following sections encompasses what digital platforms associate with these two directions.

When discussing values, it is equally essential to understand not only who holds the values, but also to identify the stakeholders in relation to which such values are held. On the basis of this distinction, the values of digital platforms are manifold, and reflect a diverse ecosystem of stakeholders. For instance, the value of providing affordability to a customer may come at the expense of the interests of workers hired by the platform. In another example, public institutions are supposed to be the embodiment of values endorsed by society at large (*e.g.* welfare), yet have an equally important interest in maintaining healthy markets.

This brings us to another necessary clarification, namely the difference between values and interests.[100] To illustrate this, in their role as privately-held companies, platforms are guided by one central interest: profit maximization.[101] In itself, this may signal values such as responsibility towards shareholders. Yet, there may be other values companies adhere to, or claim they adhere to, which could potentially be – at least in some ways – contrary to their interests.[102]

While many taxonomies already map and classify public values,[103] there seems to be no consensus regarding what may be considered as a value, whether in the public or the private sector. It could be argued that the public sector may be defined by having to represent the social values held by society at large (*e.g.,* what society perceives as right or wrong), while in the private sector, the immediate interpretation of the concept of 'value' reflects the economic values of

[97] Some of these platforms are not new. An example is IBM, one of the oldest companies in the history of computing, see James W. Cortada, *IBM: The Rise and Fall and Reinvention of a Global Icon* (MIT Press 2019). See for instance Bruce Schneier, *Click Here to Kill Everybody* (WW Norton & Company 2018).

[98] Peter Railton, *Facts, Values and Norms: Essays Towards a Morality of Consequence* (Cambridge University Press 2003) 43. See also Nicolai Hartmann, Moral Values (Routledge 2017) 2.

[99] Salomon Rettig and Benjamin Pasamanick, 'Changes in Moral Values over Three Decades, 1929-1958' (1959) 6 Soc Probs 320, 321. See also Helmut Coing, 'Analysis of Moral Values by Case-Law' (1987) 65 Washington University Law Quarterly 711; Thomas Lee Hazen, 'The Corporate Persona, Contract (and Market) Failure, and Moral Values' (1990) 69 North Carolina Law Review 273; James Hitchcock, 'Church, State, and Moral Values: The Limits of American Pluralism' (1981) 44 Law & Contemporary Problems 3.

[100] *See*, *e.g.*, Margaret Jane Radin, 'Government Interests and Takings: Cultural Commitments of Property and the Role of Political Theory' in Stephen E Gottlieb (eds.), *Public Values in Constitutional Law* (University of Michigan Press 1993) 69.

[101] See for instance Ian B Lee, 'Corporate Law, Profit Maximization, and the Responsible Shareholder' (2005) 10 Stan JL Bus & Fin 31.

[102] See for instance Shlomit Azgad-Tromer, 'The Virtuous Corporation: On Corporate Social Motivation and Law' (2017) 19 U Pa J Bus L 341; Angus Corbett, 'Corporate Social Responsibility - Do We Have Good Cause to be Sceptical about It' (2008) 17 Griffith L Rev 413; Tan Seng Teck and Chang Jau Ho and Liau Chee How and Nanthakumar Karuppiah and William Chua, 'A Theorisation on the Impact of Responsive Corporate Social Responsibility on the Moral Disposition, Change and Reputation of Business Organisations' (2018) 8 J Mgmt & Sustainability 105.

[103] See for instance Sara Hofmann, Øystein Sæbø, Alessio Maria Braccini and Stefano Za, 'The public sector's roles in the sharing economy and the implications for public values', Government Information Quarterly 1 <https://www.sciencedirect.com/science/article/pii/S0740624X18304106#bb0305>; Mayuree Yotawut, 'Examining progress in research on public value' (2018) 39(1) Kasetsart Journal of Social Sciences 168; Mark H Moore, *Recognising Public Value* (Harvard University Press 2013); Briggite Unger, Daan van Der Linde, Michael Getzner, Public or Private Goods? Redefining Res Publica (Edward Elgar Publishing 2017) 232; John Benington and Mark H Moore, *Public Value: Theory and Practice* (Palgrave Macmillan 2011); Hal G Rainey and Barry Bozeman, 'Comparing public and private organizations: Empirical research and the power of the a priori' (2000) 10(2) Journal of Public Administration Research and Theory 447.

markets.[104] Public and private values are nonetheless not strictly divided, as companies should embrace social values just as much as government embraces economic values.[105]

Extracting values is a research exercise which may entail a wide array of methods, whether qualitative or quantitative.[106] This article combines a small qualitative content analysis with a literature review to identify, analyze and compare platform values and public values.[107]

2.3. Platform Values

2.3.1. The Platform Economy and Its Vision

The platform economy benefited from particular regular leniency in its early days. In the light of the regulatory subsidies they have received in the past decades,[108] platforms have traditionally governed themselves through self-regulation.[109] However, given their growing power even as private actors, platforms may even be considered as 'norm-creating actors besides or within the state' in a legal pluralist understanding.[110] According to this understanding, platforms create their own legal orders, which complement or compete with the sovereignty of the state in making rules. From this perspective, the self-defined standards enacted by digital platforms give expression to private values which may have an economic or social nature, and are in turn aligned to the platform's interests. These interests may be different from those of the public served by the platform, yet their imposition is possible due to disparities in bargaining power.

This section first addresses the nature of the private/self-regulatory instruments drafted by platforms, and subsequently discusses selected examples of platform values which may be extracted from the terms of service ('ToS') and community guidelines of platforms that have an impact on cities. For this purpose, we have selected four representative platforms: Airbnb and Lime (as sharing economy platforms), and Sidewalk Labs and IBM Smarter Cities (as smart city platforms).

2.3.2. Voluntary and Mandatory Values

Norm creation by platforms takes the form of ToS, policies and community guidelines. It is also in these documents that we may find the vision that a platform would like to convey as well as the values it holds dear and imposes on its users. As their goal is to define transactional behavior, platforms employ mainly contracts as their primary self-regulatory instruments. Typically, sharing economy platforms create a contractual relationship between the platform and the user on the basis of the platform's general terms. Once all relevant conditions are met (*e.g.,* offer and acceptance), the standard terms delineate the rights and obligations of the parties, and in

[104] See for instance Robert T Slee, *Private Capital Markets: Valuation, Capitalization, and Transfer of Private Business Interests* (Wiley and Sons 2011); Mariana Mazzucato, *The Value of Everything: Making and Taking in the Global Economy* (Penguin Books 2018).
[105] Moore, fn 47. See also Phillip Crowson, 'Adding public value: The limits of corporate responsibility' (2009) 34(3) Resources Policy 105.
[106] Elena Harman, Tarek Azzam, 'Incorporating public values into evaluative criteria: Using crowdsourcing to identify criteria and standards' (2018) 71 Evaluation and Program Planning 68.
[107] See fn 17.
[108] For instance, in the light of the Cambridge Analytica incident, showcasing the lack of accountability mechanisms for sharing user data with third parties, Mark Zuckerberg has even called for more government regulation, The Economist, 'Mark Zuckerberg says he wants more regulation for Facebook' (*The Economist*, 6 April 2019) <https://www.economist.com/business/2019/04/06/mark-zuckerberg-says-he-wants-more-regulation-for-facebook>. For the regulation of sharing-economy platforms, see for instance, Stephen R. Miller, 'First Principles for Regulating the Sharing Economy' (2016) 53 Harv. J. on Legis. 147; Daniel E. Rauch and David Schleicher, 'Like Uber, But for Local Governmental Policy: The Future of Local Regulation of the "Sharing Economy"' (2015) Geo. Mason L.& Econ. Research Paper; Lobel, fn 11.
[109] Christoph Busch, 'Self-Regulation and Regulatory Intermediation in the Platform Economy' forthcoming in Marta Cantero Gamito and Hans-Wolfgang Micklitz (eds) *The Role of the EU in Transnational Legal Ordering: Standards, Contracts and Codes* (Edward Elgar 2019).
[110] Vanessa Mak, 'Pluralism in European Private Law' (2018) 20 Cambridge Yearbook of European Legal Studies 202, 219.

principle—though depending on the jurisdiction—the terms are binding on the parties to this contract. Due to the scale at which it is used, this contract cannot be negotiated, which results in the platform acting as a self-regulator who defines and imposes its own values onto its users. In contrast, smart city platforms negotiate contracts with local authorities, and these contracts shape the private regulatory framework governing the relationship between the transacting parties. In practice, the ability of local public authorities to truly negotiate these terms may also depend on the dimension and economic power of the city in question.

In this section, we distinguish between two types of values conveyed by platforms to the communities they serve through their services. *Voluntary values* reflect standards which are not required by state-made law, such as economic values arising out of the provision of customer service. Efficiency and effectiveness are in fact vital for businesses to establish a standard of care for the consumer's needs and build their reputation as trustworthy contracting partners.[111] In the public sector, public bodies also embrace efficiency and effectiveness as public values in order to ensure that public bodies are pursuing the public interest in the best possible way. However, voluntary values may also produce negative externalities. Consumer-oriented values may come at the expense of other stakeholder interests: to meet delivery deadlines, Amazon employees are assigned a performance rate considered by many of them inhumane.[112] Moreover, in the absence of clear public legal standards, platforms may not only embrace what is explicitly allowed, but also what is not explicitly prohibited. An example in this respect are arbitration clauses, which are always unfavorable for consumers because they restrict access to justice and impose unnecessarily high costs on resolving disputes arising out of business-to-consumer ('B2C') transactions. In jurisdictions where they are not prohibited, companies impose them on consumers by virtue of the 'take-it-or-leave-it' nature of the terms of service.

As digital platforms adapt to new regulations, compliance becomes an actively pursued interest which increases the level of protection offered to individuals, and translates into an adoption of public values into the private legal framework.[113] With the emergence of stringent regulatory frameworks focused on individual protections (*e.g.*, the Unfair Contract Terms Directive, the European General Data Protection Regulation), the bargaining power gap between digital platforms and users is somewhat reduced.[114] The values promoted by platforms in their compliance efforts may be considered as *mandatory values*.

A subsequent question in the case of sharing economy platforms arises with respect to the nature of the contract concluded between the digital platform and the user. While the specific qualification may depend on the nature of the industry in which the platform is active, the intermediation provided by the platform is an information service. Under the framework of the European consumer protection applicable to B2C-transactions, a contract regarding an information service is considered a contract for digital content, either 'allowing the creation, processing or storage of data in digital form' or 'allowing sharing of and any other interaction with data in digital form provided by other users of the service'.[115]

---

[111] See for instance J. Rose, J.S. Persson, L.T. Heeager and Z. Irani, 'Managing e-Government: value positions and relationships' (2015) 25(5) Information Systems Journal 531.

[112] Josh Dzieza, '"Beat The Machine": Amazon Warehouse Workers Strike To Protest Inhumane Conditions' (The Verge, 16 July 2019) < https://www.theverge.com/2019/7/16/20696154/amazon-prime-day-2019-strike-warehouse-workers-inhumane-conditions-the-rate-productivity>.

[113] Lya G. Soeteman-Hernandez, Margarita D. Apostolova, Cindy Bekker, Susan Dekkers, Roland C. Grafström, Monique Groenewold, et al. 'Safe innovation approach: Towards an agile system for dealing with innovations' (2019) Materials Today Communications <https://www.sciencedirect.com/science/article/pii/S2352492818304239>.

[114] Council Directive 93/13/EEC of 5 April 1993 on unfair terms in consumer contracts [1993] OJ 1993, L 95; Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC [2016] OJ 2016, L 119.

[115] Article 2 (b) and (c), Proposal for a Directive on certain aspects concerning contracts for the supply of digital content, COM(2015) 634 final.

What is more difficult to ascertain is the legal nature of policies and community guidelines which tend to be the source of platform values or at least the documents in which they are conveyed to the public. There are three ways in which these instruments can generally be interpreted. First, policies and community guidelines can be considered to be part of the standard terms, and thus binding to the extent allowed by national contract law. Second, they may be qualified as codes of conduct adopted by the platforms, and thus with a more limited binding force.[116] Third, depending on the nature of the provisions referred to in the various complementary instruments, it could be argued that some clauses may be legally binding (and thus part of the standard terms), while others may not (and consequently be seen as provisions from a code of conduct). This distinction is highly relevant when dealing with the enforcement of standards that reflect values. For instance, a value such as accountability may be expressed by a platform in its community guidelines. However, if guidelines are considered not to have any binding force, the value expression does not lead to any rights or remedies that could enforce it in practice. To illustrate these distinctions, the following section delves into the private/self-regulatory instruments employed by Airbnb, Sidewalk Labs and IBM Smarter Cities and further discusses selected provisions.

2.3.3. Platform Values in Terms of Service and Marketing Materials

When users make accounts on Airbnb, they agree to the platform's ToS, a document with almost 25,000 words which will represent the basis of their agreement. The ToS includes clauses referring to the platform's content and reflects the intermediary nature of Airbnb, as it accounts for its relationship with hosts, but also with tenants/guests. The ToS showcase the bargaining power exercised in the transactional triangle (Airbnb – hosts – guests). For instance, Airbnb applies the same standard to both hosts and renters when it comes to content. On the one hand, Airbnb grants users a 'limited, non-exclusive, non-sublicensable, revocable, non-transferable license' to download and access content on the platform, including that of other users.[117] On the other hand, the content created by members themselves is licensed to Airbnb using a 'non-exclusive, worldwide, royalty-free, irrevocable, perpetual (or for the term of the protection), sub-licensable and transferable license'.[118] However, when looking at the exclusion of liability clauses, Airbnb imposes exclusions not only with respect to its own obligations,[119] but also to those of hosts.[120] The ToS do acknowledge such exclusions may be unlawful in some jurisdictions, and thus specify that they only apply to the maximum extent permitted by law. Just like every other platform under scrutiny, Airbnb also has a privacy policy, and in addition, a copyright policy, and a cookie policy. Many of the clauses included in such policies deal with disclosures, mostly mandated by law (e.g., data retention and erasure[121]).

A more recent type of sharing economy service that currently enjoys enormous popularity in smart cities, is that of micromobility. Popular platforms like Lime change the contractual constellation through their business model, by intermediating and providing access to a fleet of company-owned electric scooters deployed on the streets of a given city.[122] While Uber built its business model on not purchasing cars, but rather relying on the cars of its riders, e-scooter businesses entail the ride-sharing company is the one making the entire infrastructure available to the public. This infrastructure generally consists in the digital platform (website and app), and the e-scooter fleet (including parking locations). As the consumer only interacts with the platform,

[116] Anna Beckers, 'Towards a Regulatory Private Law Approach for CSR Self-Regulation? The Effect of Private Law on Corporate CSR Strategies' (2019) 27(2) European Review of Private Law 221-244.
[117] Clause 5.4 Airbnb ToS.
[118] Clause 5.5 Airbnb ToS.
[119] Clause 17 Airbnb ToS.
[120] Airbnb Experiences Guest Release and Waiver, <https://www.airbnb.com/terms/experiences_guest_waiver>.
[121] Clause 6.4 Airbnb Privacy Policy <https://www.airbnb.com/terms/privacy_policy#sec6>.
[122] Cornelius Hardt Klaus Bogenberger, 'Usage of e-Scooters in Urban Environments' (2019) 37 Transportation Research Procedia 155.

both for data tracking (*e.g.*, using the platform digital infrastructure for finding the e-scooters on a map), as well as the rental of goods, a B2C-contract arises. The rental of Lime products (the company also has a fleet of bikes, e-bikes and ridesharing vehicles) is at the core of what it calls the 'User Agreement', an approximately 18,000 word-long set of standard terms that overall emphasize values such as safety, the protection of minors, but also general user obligations when using the e-scooter. In addition, Lime imposes on the user the company's release of any collective claims relating to its products, showing that whatever fairness value may be imbued in the general terms, the company tries to restrict it by limiting the customer's access to justice. Interestingly, the User Agreement extends this release to third parties such as employees, agents or affiliates, but more importantly, it also stipulates this release to the benefit of 'municipalities and public entities (including all of their respective elected and appointed officers, officials, employees and agents) which authorize Lime to operate any of its Services.'[123] In other words, the standard terms consumers need to agree to before using Lime products and services exclude the possibility of bringing collective claims against public administration entities on the basis of this use.[124]

Another point that deserves particular attention is reflected by data collection and transfers. GPS tracking as that found at the core of the Lime application is infamously sensitive, as it can trace an individual's location history, thereby raising serious data protection (but also moral) questions.[125] Hackers who disassemble Lime GPS modules have revealed that these are devices (micro-computers) that run on Android and have a 4G-SIM card, which entails that e-scooters are connected to the Internet.[126] In combination with the acknowledgement that 'Lime may disclose aggregate and other data about [the consumer] in accordance with applicable law, including, without limitation, general latitude and longitude data for [consumer] addresses (provided this would not allow any individual's address to be separately identified)',[127] this may raise additional concerns. Sharing allegedly anonymized data with third parties without any further specifications may mislead consumers when using the Lime app. Especially since Lime seems to collaborate with Uber,[128] and this collaboration entails data sharing between the two companies, transparency ought to be one of the guiding values promoted by Lime, but it is not.

It could be argued that a smart city platform like Sidewalk Labs uses technology to create urban development solutions to problems such as rising rents, traffic congestion, or air pollution. These types of projects require collaboration between the platform local government and local communities. This role places such a service provider in a different type of intermediation, whereby local government, itself unable to generate public-interest technology, outsources this process to tech companies. [129] As its clients are not peers, but local governments, contractual frameworks will most likely take place under strict rules of national and local administrative law and European public procurement. These contractual terms are thus not made fully available on the company's website which means that our value analysis is complemented here by the literature and the media discussion of specific projects developed by Big Tech platforms. This lack of transparency is also applicable to additional websites made for specific projects belonging to Sidewalk Labs, such as Replica, an urban planning tool.[130] As an example of the way in which Sidewalk Lab operates, we can shift our attention to the Sidewalk Toronto project, meant to 'shape

[123] Clause 5.1(iii) Lime User Agreement.
[124] This clause, just as the rest of the User Agreement, is valid only to the extent it does not contravene to mandatory rules at the national level. For instance, according to European consumer law, this clause may be judicially deemed unfair according to the Unfair Contract Terms Directive, fn 58.
[125] Associated Press, '"Tracking every place you go": Weather Channel app accused of selling user data' (*The Guardian*, 5 January 2019) <https://www.theguardian.com/technology/2019/jan/04/weather-channel-app-lawsuit-location-data-selling>.
[126] Scootertalk, see <https://scootertalk.org/forum/viewtopic.php?t=1370>.
[127] Clause 10.1 Lime User Agreement.
[128] Lime mentions this on its website, see <https://www.li.me/lime-uber-electric-scooter>.
[129] David Eaves, Kevin Frazier and Lavanya Singh, 'Public Interest Technology' (*Medium*, 22 February 2019) <https://medium.com/digitalhks/entry-i-public-interest-technology-a-brief-look-to-the-past-to-identify-pit-in-the-present-bb72cadd593c>.
[130] See Replica, <https://replicahq.com>.

the city's future and provide a global model for inclusive urban growth'.[131] Sidewalk published the project Master Innovation and Development Plans (MDIPs) on its website, and while it is not clear whether these are the final plans considered for implementation, this publication can be seen as an attempt to embrace transparency towards project stakeholders. In addition, by taking into account digital accessibility needs and ensuring that such plans can be read by citizens facing various physical barriers,[132] another value that can be underlined is that of accessibility. As for the content of the MDIPs, Sidewalk Labs lists its own eight commitments of the Proposed Innovation and Funding Partnership, including the deployment of 'cutting-edge technologies to improve urban life', 'spur[ring] economic development', or 'sharing profits associated with certain technologies with the public sector', which generally reflect economic and social values tailored to the needs of a public administration client.

The same approach is taken by IBM Smarter Cities, as it already works with cities such as Busan (Korea), Palermo (Italy), San Isidro (Argentina), San Jose (US), and Yamagata City (Japan) in the context of a *pro bono* system where IBM would offer consultancy on matters such as public safety, economic development, affordable housing and even social services.[133] IBM Smarter Cities is the vision and set of technology solutions touting potential contributions to what it calls 'cognitive government' and covering policy areas such as public safety, smart buildings and urban planning. Similarly to Sidewalk Labs, IBM Smarter Cities does not target consumers, and therefore its business model does not need to account for any contractual B2C framework. Out of the areas of interests listed on its website, IBM Smarter Cities safety, social services and affordable housing can be translated into the values of citizen safety, care (welfare), and affordability, and additional values may be inferred – albeit with less clarity - from other areas of interest (*e.g.*, economic development may or may not promote the value of equality).[134]

The examples of digital intermediaries we reflected upon so far in this section account for a wide range of contractual practices employed for the private governance of (mostly B2C) intermediated transactions. Where intermediaries publicize ToS, as was the case for Airbnb and Lime, we can observe a dramatic contrast between the marketing language used to entice consumers, and the overwhelmingly lengthy and carefully worded contractual clauses that primarily aim to limit the platform's liability, and create frameworks that might appear compliant with legal standards. Yet, what are the true values and interests of the platform when drafting such terms? For instance, Lime claims that agreeing to its ToS entails giving Lime 'the right to photograph, videotape, and otherwise record [the consumer's] appearance and voice related to [the consumer's] use of the Services, at any time and from time to time.'[135] It is unclear what this right aims to achieve. Does it entail that the SIM cards in the GPS module can be used for recordings? The mere consideration that an important contractual clause like this may leave too much space for interpretation should raise concerns regarding the commercial intentions leading to the ToS. Lime may tout its services as 'cleaner and less expensive than a rideshare', or claim that it is 'working with city, university and community partners to enable smart micro-mobility around the world',[136] thereby implying to embrace values such as affordability, collaboration and sustainability. Still, limitation of liability clauses, especially when imposed in legal systems that do not specifically prohibit them, may also show an overarching economic interest that can be said to overpower a value such as fairness. Similarly, Airbnb's marketing speaks about 'unforgettable trips', 'adventures nearby or in faraway places and access unique homes, experiences, and

[131] See Sidewalk Toronto, <https://www.sidewalktoronto.ca/accessible-midp>.
[132] Ibid. Public documents are said to be 'WCAG 2.0 AA compliant and have been validated on PAC3, Adobe Acrobat Pro DC Accessibility Checker'.
[133] Smart Cities World, 'Five cities land IBM Smarter Cities Challenge grants' (*Smart Cities World*, 21 July 2017) <https://www.smartcitiesworld.net/news/news/five-cities-land-ibm-smarter-cities-challenge-grants-1918>.
[134] Ibid.
[135] Clause 12.7 Lime User Agreement.
[136] See Lime website, <https://www.limebike.com>.

places around the world,'[137] which may reflect values related to improving the human experience, such as increasing the livelihood of the global citizen. However, as mentioned above, this sometimes comes at the cost of other stakeholders not targeted by these values, such as locals whose cities become overcrowded by tourism, and who need to bear the negative effects of the imposed platform values. Given the nature of their services and the lack of user agreements, Sidewalk Labs and IBM Smarter Cities list values in their mission statements.[138] These values are the same as those portrayed by the sharing economy platforms in the discussion above, yet it remains to be seen what kind of values are taken over in the contractual framework with their clients. As contracts capture the intention of the parties with respect to a specific transaction, or even within a broader context than the transaction itself, they are useful in interpreting what this intention actually is. In principle, platforms may have well-articulated visions about their role in society. However, a closer look at the values these platforms claim to support shows not only that transnational (or 'cosmopolitan') values are applied with disregard for national and local values but also that the implementation of platform values in their business practices may vary considerably.

In the next section, these values will be discussed in comparison with public values driven by public interest, in order to better gauge the potential conflicts emerging out of the public/private divide as applied to the context of smart cities.

3. Platform Values versus Public Values in the Smart City

3.1. Defining Public Values

The protection of public values is inherently linked to the pursuit of the public interest. Yet, these two concepts are distinct. The 'public interest' represents an ideal that changes with time and place rather than an identifiable content, it refers to the pursuit of the outcomes that best 'serve the long-term survival and well-being of a social collective constituted public.'[139] Public values are those normative judgments that reflect 'a consensus about rights, benefits, and prerogatives to which citizens should and should not be entitled to; the obligations of citizens to society, the state, and one another; and the principles on which government and policies could be based.'[140] Public values are thus broader than rights. For example, accountability, inclusiveness, and efficiency refer to rights that citizens may have (for example, the right to have access to certain documents or the right not to be discriminated) but they also include a moral dimension that goes beyond legal rights.[141] Drawing on this characterization, it is clear that any list of national public values is by definition incomplete. In this section, we focus on the public values that are particularly important for cities and we identify a set of public values that are mentioned on a regular basis in national legislation, local policy documents, and scholarship.[142] To illustrate this point, while each English city has its own policies and local values, the Local Government Act of 2000 states that the objective of any local authority should be 'the promotion or improvement of the economic, social and environmental well-being of their area.'[143] This disposition can be interpreted as a reference to a number of public values such as affordability of public services, sustainability, inclusiveness, and promotion of the local economy.

---

[137] See Airbnb website, <https://www.airbnb.com>.
[138] For instance, Sidewalk Labs: standards of sustainability, affordability, mobility, and economic opportunity, see <https://www.sidewalklabs.com/mission/>.
[139] Barry Bozeman, *Public Values and Public Interest: Counterbalancing Economic Individualism* (George Washington University Press 2007) 12.
[140] Ibid. 13-14.
[141] Frank Bannister and Regina Connolly, 'ICT, Public Values and Transformative Government: A Framework and Programme for Research' (2014) 31 *Government Information Quarterly* 119, 120.
[142] See for instance Saskia Sassen, *The Global City* (Princeton University Press, 2001); Peter J Taylor amd Ben Deruder, *World City Network: A Global Urban Analysis* (Routledge 2015); Mark Amen et al. (eds), *Cities and Global Governance: New Sites for International Relations* (Routledge, 2016); Nestor Davidson and Geeta Tewari (eds*), Global Perspectives in Urban Law: The Legal Power of Cities* (Routledge, 2019).
[143] Local Government Act of 2000, Section 4.

Before delving into an overview of these public values, it is important to distinguish between the creation of public value which aims at the production of value for society and the protection of public values as such. The creation of public value is a broader approach which ensures that a public organization meets the needs and expectations of citizens. This approach is based on the so-called 'public value management paradigm' which seeks to gain a legitimate mandate from citizens to pursue the public interest by advancing the efficient performance of public authorities, accountability, responsiveness to public needs, and trust.[144] In order to achieve legitimacy, public authorities need to show that they are transparent, accountable, and open to the input of citizens.[145] Achieving public value in the context of the digitization of public services has been regarded as a way to improve efficiency in government, improve public services to citizens, and social values such as inclusion, democracy, transparency, and participation.[146] Since the liberalization movement, the conflict between individual and public values has made it more difficult to find a balance between the creation of public value in an economic sense and the protection of public values.

A first set of public values that is often mentioned in scholarship and policy documents pertaining to local public authorities refers to the quality and affordability of public services.[147] Cities also have a particular interest in safeguarding the public values of availability, stability, and sustainability of certain services of general economic interest such as energy.[148]

Accountability and transparency are often presented as key public values that are being affected in different ways by public authorities' reliance on digital platforms. These two values are for example underlined as key public values of Bristol's social policy and all 'governance arrangements are to be agreed, in order to achieve transparency, and ensure accountability to all of our stakeholders, including our customers, contractors, suppliers, our partners and auditors'.[149]

Public values that refer to public services also go beyond their quality and affordability. They also include the neutrality of their provision to citizens, that is, public authorities should be politically neutral and objective in their communication with citizens and provide services to all citizens without imposing certain political views.[150] Reliance on digital platforms for the provision of public services rarely fulfills this mission. Platforms values which are primarily driven by individualism, tend not to service the primary interests of society, but rather see public administration as a contracting client or as a hurdle that needs to be overcome, in order to have legal access to the market. For example, while Airbnb may contend that the platform aims to support local communities, it not primarily driven by this public value but by their own financial interests. A second aspect where this neutrality may easily disappear refers to the use of digital platforms in the context of smart cities to influence the behavior of citizens in smart cities, for

[144] See Gerry Stoker, 'Public Value Management: A New Narrative for Networked Governance?' (2006) 36 American Public Administration Review 41; Collin Talbot, 'Measuring Public Value—A Competing Values Approach' (The Work Foundation, 2008).
[145] This public-value movement towards openness and transparency with reliance on technology was particularly visible during the Obama Administration in the United States and is present in numerous European openness strategies, see Beth Noveck, *WikiGovernment: How Technology Can Make Governments Better, Democracy Stronger, and Citizens More Powerful* (Brookings 2010).
[146] Twizeyimana and Andersson, fn 38.
[147] Hans de Bruijn and Willemijn Dicke, 'Strategies for Safeguarding Public Values in Liberalized Utility Sectors' (2006) 84(3) Public Administration 717.
[148] Marga G Edens and Saskia Lavrijssen, 'Balancing Public Values during the Energy Transition—How Can German and Dutch DSOs safeguard sustainability' (2019) 128 Energy Policy 57; Catherine Butler, Christina Demski, Karen Parkhill, Nick Pidgeon and Alexa Spence, 'Public values for energy futures: Framing, indeterminacy and policy making' (2015) 87 Energy Policy 665.
[149] Bristol City Council, 'Social Value Policy: Creating Social Value in Bristol' (Bristol.gov.uk, January 2019) <https://www.bristol.gov.uk/documents/20182/239382/Social+Value+Policy+-+approved+March+2016-1.pdf/391b817b-55fc-40c3-8ea2-d3dfb07cc2a0>.
[150] Alberto Alemanno and Alessandro Spina, 'Nudging Legally: On the Checks and Balances of Behavioral Regulation' (2014) 12 I•CON 429.

example, through nudging techniques.[151] When information is filtered, omitted or transmitted in a non-neutral way in order to influence the choices of citizens, the autonomy of citizens may be significantly affected.[152] Public authorities have the duty to protect information neutrality and diversity.

A second set of public values that we often identify in legislation and policy documents have a social nature. This set includes for example inclusiveness, equality of treatment and access, affordability of (public and private) housing, safety, and the livability of cities. While these values may resonate with most of us nowadays, it is worth underlining that equality of access and treatment when it comes to public services are relatively recent public values.[153] Digital platforms can on the one hand ensure that more citizens and visitors have access to digital services but on the other they may also exclude less tech-savvy citizens if the services are only available online.[154] In many cities throughout the world (including Western countries) the digital divide and the limited digital literacy of many thousands of citizens is deepening inequality and excluding many residents from city services.[155]

The Toronto Public Service By-law mentions explicitly the need to promote diversity as an integral part of Toronto's civic identity.[156] Bristol also comprises inclusiveness as one of the key values of the city' social value policy.[157] As a consequence of the growth of Airbnb and other home-sharing platforms, investment in private houses for tourism has become such an important source of income that residents are leaving cities. While platforms values convey flexibility in housing, this has meant that poor residents living in touristic areas have been terrorized to leave their houses which will later be transformed into Airbnb-houses.

Third, economic growth and the promotion of local economy appear to be also public values highly underlined in local policy documents. For example, the city council of Bristol enhances the importance of promoting 'the local economy, so that micro, small and medium sized enterprises and the voluntary and community sector in Bristol can thrive,' 'creating or promoting local employment, training and inclusive economic sustainability'.[158]

Fourth, Amsterdam as well as other Dutch smart cities have also enhanced the need to advance a new set of public values in recent policy documents, such as privacy, autonomy, and broad democratic participation.[159] In order to protect these public values in data-driven urban contexts, local public bodies have invested in the development of ethical and data protection impact assessments and hiring their own data scientists and analysts to assess the quality of the data collected in smart cities.

---

[151] Sofia Ranchordás, 'Nudging citizens through technology in smart cities' (2019) International Review of Law, Computers and Technology (https://www.tandfonline.com/doi/full/10.1080/13600869.2019.1590928).
[152] Brent D Mittelstadt et al., 'The Ethics of Algorithms: Mapping the Debate' (2016) Big Data & Society 1, 9.
[153] Bannister and Conolly, fn 85, 124.
[154] Beth Simone Noveck, *Smart Citizens, Smarter State: The Technologies of Expertise and the Future of Governing* (Harvard University Press 2015); Beth C Weitzman, Diana Silver and Caitlyn Brazill, 'Efforts to Improve Public Policy and Programs through Data Practice: Experiences in 15 Distressed American Cities' (2006) Public Administration Review 386, 387. See also Simone Noveck, 'Peer to Patent": Collective Intelligence, Open Review, and Patent Reform' (2006) 20 Harvard Journal of Law & Technology 123.
[155] See for instance Harlan Yu and David G Robinson, 'The New Ambiguity of 'Open Government' (2012) 59 UCLA Law Review Discourse 178.
[156] Chapter 192 of Toronto's Municipal Code incorporated in 2015 the Toronto Public Service By-law which includes a list of public values that aim to guide public services and the management of resources in this city.

[157] Fn 92.
[158] Bristol City Council, 'Social Value Policy: Creating Social Value in Policy', available at https://www.bristol.gov.uk/documents/20182/239382/Social+Value+Policy+-+approved+March+2016-1.pdf/391b817b-55fc-40c3-8ea2-d3dfb07cc2a0 (last accessed on October 7, 2019).
[159] Rathenau Instituut, 'Hoe beschermen gemeenten publieke waarden in de slimme stad?' [How Do Municipalities Protect Public Values in the Smart City?], available at https://www.rathenau.nl/nl/digitale-samenleving/hoe-beschermen-gemeenten-publieke-waarden-de-slimme-stad (last accessed on October 7, 2019).

To conclude, traditional cities tend to emerge as a result of a complex interaction between different elements: geography, economy, existence of raw materials.[160] In the digital age, technology is transforming the planning, organization, and governance of cities by their ability to forecast new events and needs (*e.g.*, criminality, sustainability) and thus better allocate city resources.[161] However, technology should nonetheless be used to pursue these values and not the other way around.

3.2. Balancing Platform Values with City Values

At first sight, digital platforms privilege specific values in the platform economy: convenience and short-time access over long-term engagements, flexibility over stable employment, sharing of information, objects, and experiences over ownership and discretion.[162] Platforms in the sharing economy also claim that they promote sustainable transactions. Many citizens have come to accept these values and, in most cases, national and local governments have found a way to regulate them without interfering with the essence of these services. To illustrate, in most cities Airbnb hosts are allowed to rent rooms to tourists without obtaining a license as long as they do so only for a short period of time. Airbnb also claims on a regular basis that they provide 'healthy, diverse, inclusive and sustainable' travel and aim to benefit 'all of its stakeholders, including (….) communities'.[163]

Despite this appearance of harmony with local communities and possibly their values, the platform economy is one of the different urban contexts where we observe the expansion of platform values. Legal literature has nonetheless not yet discussed the broader phenomenon underlying the advancement of platform values at local level. This is particularly important as it has become clear that platform values are not always aligned with national or local values established in existing legal frameworks. Local residents may not wish to benefit from the flexibility and cosmopolitan interaction that Airbnb or other platforms seek to promote. Rather, the safety, affordability, and family-friendliness of their neighborhoods may be the values that they prefer to hold on to and have entrusted their local representatives to protect.

What is more, digital platforms seek to advance more than economic values. As important vehicles of news, advertisement, and political influence, digital platforms also appear to have intrinsic values regarding for example hate speech, voting behavior, sustainability, and the protection of human rights.[164] These values are implicitly or explicitly listed in large platforms' community guidelines. Platforms advance these values for example through the promotion of messages to their users on online community forums or the publication of codes of conduct (*e.g.*, Airbnb's Non-discrimination policy). Platforms encourage users who detect content contrary to their 'values' to report it and enforce it themselves by sanctioning users with the removal of content or shutting down accounts. The promotion of platform values is nonetheless problematic for several reasons: first, it is unclear what the nature and relevance of these values are. As platforms become essential infrastructures for communication, business, social and political influence, platform values are starting to affect the public sphere and the public interest. However, here a second problematic aspect arises: platform values may differ from national values. In a certain jurisdiction, the legal and social acceptance of renting out (even if only sporadically) apartments to strangers or even the definition of 'hate speech' may be perceived in very different terms from those adopted by a platform's online community guideline. Despite the alleged good intentions of platforms, the merit of many of their initiatives to reduce discrimination, and their attempt to take into account local customs, the law and values of platforms are not always aligned

[160] See generally Steven B Johnson, *Emergence: The Connected Lives of Ants, Brains, Cities, and Software* (Scribner, 2001).
[161] Terry Farrell, 'City Making: Many Hands, Over Time'(2018) 13 Journal of Planning & Environmental Law 6.
[162] Lobel, fn 11.
[163] Airbnb, 'About us' <https://press.airbnb.com/about-us/>.
[164] Keller, fn 3.

with the law and values of the land. This tension has become particularly challenging in the last years as platforms started playing a growing role in the provision of public services (*e.g.*, crowd-management), for example, in the context of smart cities.[165]

In smart cities, IBM, Sidewalk Labs (a subsidiary of Alphabet to which Google also pertains) or Huawei collect and process personal and urban data through Internet-of-Things, big data, and algorithms.[166] In Toronto, Sidewalk Labs is designing a new district to 'tackle the challenges of urban growth' that would collect data from a wide range of sources to facilitate mobility, logistics, and sustainability solutions.[167] In April 2019, the Canadian Civil Liberties Association sued Waterfront Toronto, the publicly funded entity responsible for the project, and the Canadian government at three levels (federal, provincial, and municipal powers) over this plan. This innovative plan has been shrouded in secrecy and opacity and has been accompanied by raising concerns (for example, the limited protection of the privacy of Toronto residents).[168] The media has reported that resigning members of the Waterfront Toronto and the civil society are particularly concerned with the protection of Canadian values and the fact that SideWalk Labs is the one defining the values fed into the digital technology employed in the city rather than democratically elected officials.[169] Toronto is one of many examples analyzed in this article, of a controversial partnership where digital platforms seek to interfere with local values by promoting a technocratic discourse that is susceptible of violating important public values (*e.g.*, privacy and autonomy of citizens) and the limit the participation of less tech-savvy citizens.

In the last decade, a growing number of cities and local authorities have embraced digital technology either to improve the efficiency and sustainability of their services or with the ambition to transform their urban centers into so-called 'smart cities.'[170] Since there is no consensual definition of 'smart city'—and this article does not only focus on smart-city platforms—we will refer to urban centers that rely more generally on digital platforms to improve the quality of living of their citizens and visitors as 'digital cities'.[171]

In a smart city, citizens and visitors can use different digital platforms to obtain both public and private services (*e.g.*, finding tourist accommodation, identify the fastest route to go from one point to the other). Thanks to platforms, citizens have become more mobile, several services are more convenient, and cities have the potential to become more sustainable.[172] However, public authorities may only collect this data, contract with private tech companies providing information services, and regulate local services to promote tourism in the strict pursuit of the public interest and safeguard of public values.[173]

---

[165] See, for example, Rob Kitchin, 'The Real-time City? Big Data and Smart Urbanism' (2014) 79 GeoJournal 1-14; Rob Kitchin, 'The Promise and Peril of Smart Cities' (2015) 26(2) Computers and Law.

[166] For a legal analysis of the privacy implications of the widespread collection of data in cities, see van Zoonen, fn 9.

[167] Laura Bliss, 'How Smart Should a City Be? Toronto Is Finding Out' (*Citylab*, 7 September 2018) <https://www.citylab.com/design/2018/09/how-smart-should-a-city-be-toronto-is-finding-out/569116/>.

[168] Ibid.

[169] Ibid.

[170] For a literature review on the definition of smart city, see for instance, Margarita Angelidou, 'The Role of Smart City Characteristics in the Plans of Fifteen Cities' (2017) 24 Journal of Urban Technology 3; Andrea Caragliu, Chiara del Bo and Peter Nijkamp, 'Smart Cities in Europe' (2011) 18 Journal of Urban Technology 6.

[171] The concepts of 'smart city' and 'digital city' are distinct. See Renata Dameri, 'Comparing Smart and Digital City: Initiatives and Strategies in Amsterdam and Genoa. Are They Digital and/or Smart?' in Renata Dameri and Camille Rosenthal-Sabroux (eds) *Smart City: How to Create Public and Economic Value with High Technology in Urban Space* (Springer 2014).

[172] Kitchin, fn 106.

[173] See Christopher Bovis, *Public-Private Partnerships in the European Union* (Routledge, 2014).

4. Normative Approaches to Public-Private Values Supporting Local Public-Interest Technology

The rationale behind the existence of public administration is to give an institutional setting to the enactment of public values in society.[174] As seen in section 3, these values shape public policy, public morality, and define various groups of individuals and their preferences. Within an increasingly digitized society, public values are at risk from two perspectives. First, Big Tech may replace public values with private values, which may be opaque and undesirable. Second, by enforcing privately-held socio-legal standards, Big Tech may be seen to compete for the sovereignty of law-making. Each of these points will be discussed below, in order to propose new theoretical and practical solutions for the tensions that we have seen to arise between the public and the private spheres.

Throughout this article we have tried to give illustrations of both public values and platform values. At first sight, these two notions seem to clash when platforms present themselves as guardians of public values: fairness and equality as legal standards and public values will not be interpreted in the same way by the private sector. A telling example in this respect are the lengthy exclusion or limitation of liability clauses that companies like Airbnb and Lime unilaterally impose on their customers. If a property on Airbnb or a Lime e-scooter have a hidden defect that causes a loss to their respective landlords or renters, the law deems it fair for the victim to have a means of both a remedy and an action for them to be placed in a position where the loss would not have occurred.[175] Yet in their ToS, both companies take any precaution possible not to be held liable for losses that mandatory law may impose on them. Therefore, they try to exclude their potential accountability.

However, in other ways, private and public values may be very similar, if not complementary.[176] The sharing economy is said to have led to the creation of a market niche that promotes sustainability because of its increasing profitability.[177] If additional mandatory values are imbued in the private sector through top-down regulation (*e.g.,* fuel-related limitations and restrictions), sustainability may very well become a shared value. Additional private values we identified earlier, such as collaboration or affordability, may be associated with the public values of participation and citizen care (welfare), as citizens are expected to actively contribute to democratic decision-making or standard setting.[178] Moreover, the dynamics between values and the interests of institutions or companies upholding them have similarities as well. On the one hand, private values try to reconcile customer centricity with the inherent economic interests of a given business. On the other hand, public values are caught in between the promotion of the greater good of society and the political influence exercised on this process.

[174] Hofmann, Sæbø, Braccini and Za, fn 47.
[175] Unfair Contract Terms Directive, Annex, point a shows there is a presumption that a term limiting the legal liability of a seller in the case of a personal injury is unfair: '(a) excluding or limiting the legal liability of a seller or supplier in the event of the death of a consumer or personal injury to the latter resulting from an act or omission of that seller or supplier. Point (b) of the Annex specifies that a standard term excluding or limiting access to justice is equally presumed to be unfair: (b) 'inappropriately excluding or limiting the legal rights of the consumer vis-d-vis the seller or supplier or another party in the event of total or partial non-performance or inadequate performance by the seller or supplier of any of the contractual obligations, including the option of offsetting a debt owed to the seller or supplier against any claim which the consumer may have against him'.
[176] Unger, van Der Linde and Getzner, fn 47, 230. See also Konstantin Petrichev, Susan Thorp, 'The private value of public pensions', (2008) 42 Insurance: Mathematics and Economics 1138; Bram Klievink, Nitesh Bharosa and Yao-Hua Tan, 'The collaborative realization of public values and business goals: Governance and infrastructure of public–private information platforms' (2016) 33(1) Government Information Quarterly 67.
[177] Harald Heinrichs, 'Sharing economy: A potential new pathway to sustainability' (2013) 22(4) GAIA-Ecological Perspectives for Science and Society 228; Boyd Cohen and Jan Kietzmann, 'Ride on! Mobility business models for the sharing economy' (2014) 27(3) Organisation & Environment 279.
[178] See for instance Juho Hamari, Mimmi Sjöklint and Antti Ukkonen, 'The sharing economy: Why people participate in collaborative consumption' (2016) 67(9) Journal of the association for information science and technology 2047.

We thus posit that in order to better understand how the private values of Big Tech platforms and the public values of state institutions interact and affect one another, it is necessary to move from a narrative of opposition to a model of complementarity at a level which goes deeper than existing approaches to co-regulation. Differences in interests do not always generate differences in values, and if digital platforms see municipalities as more than clients, but as co-creators of business opportunities which benefit local communities, this can shape a new model of local public-interest technology, dependent on the values shared by both platforms and local authorities. Any transportation company, not just Lime, will make safety one of its core values, because its profits depend on public trust, which can be broken easily.[179] Traffic rules and standards adopted by national and municipal authorities equally reflect safety as a value, as do sanctions meant to enforce them. By following a narrative that divides values into public versus private, there is no room left for reflecting on how the private sector can strengthen the public sector and vice-versa. Indeed, there are many risks that need to be carefully considered when blurring this line, and we critically explored some of them in section 3. As we have explained above, one of the arguments against digitalizing public infrastructure through private services has been that platforms may reshape public values in smart cities and advance a technocratic discourse that may exclude a number of underrepresented groups and less tech-savvy citizens. Yet, exclusion is not only a problem associated with the Big Tech platforms of the 21st century, given that exclusion arising out of technology adoption often reveals more traditional causes, such as 'inequality and social exclusion in the e-society are partly rooted in the capability to access and use information rather than just in the access to technological resources'.[180]

This brings us to the second point of this normative section. As functional sovereigns, digital platforms not only generate the private economic and legal standards that define their interaction with the users and thus society at large, but they are also the administrators of these standards. In the Airbnb example mentioned earlier in this paper, the need behind an agreement between Airnbn and the municipality of Amsterdam arose out of the convergence of two general interests. On the one hand, Big Tech companies want to retain as much independence as possible in setting their own limits to products and services. On the other hand, municipalities simply do not have the capacity to enforce all their rules in the platform age as this would involve in some cases daily door-to-door inspections to verify who is renting their house legally and who is not.[181] If municipal regulations limit the number of days for which a home may be rented out, or imposes licensing requirements, the consistent and fair enforcement of such standards is impossible, because the resources necessary for digital monitoring and e-enforcement are at the moment too high. Moreover, local government may not have access to platform data, which renders monitoring attempts somewhat powerless. In consequence, local government needs to collaborate with digital platforms just as much as digital platforms need the support of local government. As Cohen and Sundararajan put it, 'digital platforms [should be utilized] as partners in the regulation of exchange, rather than […] as adversaries or entities that require governmental regulation.[182]

The approach we propose consists in the creation of a legal framework to facilitate this collaboration. Numerous technology companies aim to disrupt the market with aggressive business models which take advantage of legal uncertainty.[183] This is in some cases possible

[179] James Tapper, 'Television presenter Emily Hartridge dies in electric scooter crash' (*The Guardian*, 14 July 2019) <https://www.theguardian.com/uk-news/2019/jul/13/tv-presenter-emily-hartridge-dies-in-scooter-crash>.
[180] Mike Cushman and Rachel McLean, 'Exclusion, inclusion and changing the face of information systems research' (2008) 21(3) Information Technology & People 213.
[181] See for instance the legal struggles of the municipality of Amsterdam in its e-enforcement strategies, in a case where user information was scraped by the municipality from Airbnb. District Court of Amsterdam, judgment of 27 June 2018, ECLI:NL:RBAMS:2018:4442.
[182] Molly Cohen and Arun Sundararajan, 'Self-Regulation and Innovation in the Peer-to-Peer Sharing Economy' (2015-2016) 82 U Chi L Rev Dialogue 132. See also Antonio Cordella and Leslie Willcocks, 'Government policy, public value and IT outsourcing: The strategic case of ASPIRE' (2012) 21(4) The Journal of Strategic Information Systems 295.
[183] Lobel fn 11, 92.

because laws that were made to fit other decades need time to be adapted by the judiciary or by law-makers, and disruptive innovation thrives – at least temporarily – in this uncertainty.[184] Evidently, legal uncertainty cannot be fully removed, but it can be improved. One such improvement we propose is the creation of a legal framework at municipal level for technology companies launching new products and services that have a direct impact on public infrastructure (and thus on public values). This legal framework can, instead of regulating specific technologies, focus on a legal duty to negotiate the conditions of the economic activity with the municipality in good faith. In the case of SideWalk Labs and IBM Smart Cities, this is already happening. As the public sector takes on the role of customer, success stories have already developed, such as the case of government-driven sharing economy services in Seoul.[185] This does not mean, however, that success comes without criticism. According to Hofmann et al., in the sharing economy setting, the public sector is dependent on the functionality of the provider, and this dependency 'can endanger the robustness of the public sector'.[186] However, these arguments can be made for any functionality that is outsourced by local government through tendering procedures (which may very well apply to digital services), including services as trivial as catering for the employees of a given municipality. On a positive note, outsourcing infrastructural needs (*e.g.*. micromobility, urban planning) for data-driven solution has a wealth of benefits. Firstly, cities would be able to have access to state-of-the-art services they do not have the resources to design internally, and could thus better serve communities. As an example, bike-sharing schemes can be a saving service for large cities that struggle with air pollution and traffic congestion: in 2018, Romanian company Pegas launched its bike-sharing system consisting in the deployment of over 2,000 bikes in predefined parking spaces, with the goal of improving the livelihood of locals and tourists alike, while promoting sustainable urban mobility,[187] which the municipality of Bucharest tried to achieve on earlier occasions with limited success. Secondly, digital platforms can generate new business models in the form of public-interest technology provided to public authorities instead of to consumers or other businesses. Public-interest technology is an umbrella term for a plethora of options bringing together technologists and public administration, and its burgeoning significance for the convergence of public and private interests raises a multiplicity of additional questions for legal and interdisciplinary research.[188] Lastly, public-private partnerships where the different parties have the real ability to discuss terms and their underlying values can help promote mandatory values more consistently in order to protect the public values that may be at stake with the rise of data-driven innovation.[189]

5. Conclusion

Nowadays, most digital services are built around platforms, which we understand to be a digital system that reduces transaction costs by organizing decentralized information, matching supply and demand, and allowing for different forms of collaboration.[190] This broad definition of platforms that we have used in this article primarily encompasses the emergence of the digital platform business model which has disrupted the economy (*e.g.*, Airbnb disrupted the tourist

[184] Hannah A Posen, 'Ridesharing in the sharing economy: Should Regulators impose Über regulations on Uber?' (2015) 101(1) Iowa Law Review 405.
[185] M Jae Moon, 'Government-driven sharing economy: lessons from the sharing city initiative of the Seoul metropolitan government' (2017) 33(2) Journal of Developing Societies 223.
[186] Hofmann, Sæbø, Braccini and Za, fn 47.
[187] Irina Marica, 'Local producer ready to launch smart bike sharing system in Bucharest' (*Romania Insider*, 10 May 2018) <https://www.romania-insider.com/pegas-smart-bike-sharing-system-bucharest>.
[188] For an overview of this concept, see Bruce Schneier, 'Public-Interest Technology Resources' (*Public-Interest Tech*, 30 September 2019) <https://public-interest-tech.com>.
[189] Lin Li, Philip Hookon Park and Sung-Byung Yang, 'The role of public-private partnership in constructing the smart transportation city: A case of the bike sharing platform' (2018) Asia Pacific Journal of Tourism Research 1.
[190] See for instance, Annabelle Gawer (ed), *Platforms, Markets and Innovation* (Edward Elgar Publishing 2009); Carlo M Rosotto et al., 'Digital Platforms: A Literature Review and Policy Implications for Development' (2018) 19 Competition and Regulation in Network Industries 93.

accommodation sector) and shifted firms' competition models to data-driven systems.[191] This concept also includes at least two types of platforms that are becoming increasingly visible in cities: digital platforms developed for smart cities and 'sharing-economy' platforms.

New digital platforms can be used for municipal management, public safety and environmental protection, as well as smart transportation, smart government, smart education, and smart agriculture. While the power of platforms has been comprehensively analyzed when it comes to its global impact, the legal literature has so far only superficially touched upon what this power means for the local values represented and implemented by local authorities. Sharing economy and smart city platforms are, to this extent, two telling examples.

After establishing the theoretical framework relating to the notion of value, in this paper we provided an overview of selected private values extracted from the ToS and marketing materials of four different platforms: Airbnb, Lime, Sidewalk Labs and IBM Smart Cities. We then scrutinized interdisciplinary academic scholarship as well as an illustrative number of documents compiled by local authorities, to define and exemplify public values, and to critically address the potential conflict between the public and private value divide, with a specific emphasis on the interests of local communities.

We argued that regardless of the value-creation benefits produced by digital platforms, public authorities should be aware of the risks of technocratic discourses and potential conflicts between platform and local values. In this context, we suggested a normative framework focusing on two points: departing from values shared by platforms and authorities, in order to shape a new kind of knowledge-service creation, namely local public-interest technology; and addressing the digital enforcement issue driven by the functional sovereignty role of platforms, by proposing a negotiated contractual system that seeks to balance platform values with public values.

While the example of digital cities provides a resourceful starting point in its furtherance, the concept of public-interest technology is in its infancy and more research is necessary to determine its meaning, scope and implications for society. The same can be said for the need to elaborate on new models of negotiated regulation and co-regulation that can bring together platforms and local authorities on the basis of their shared values and guarantee a closer alignment of platform and public values.

[191] Aneesh Zutshi and Antonio Grilo,'The Emergence of Digital Platforms: A Conceptual Platform Architecture and impact on Industrial Engineering' (2019) 136 Computers & Industrial Engineering 546. On the competition impact of digital platforms, see also Ariel Ezrachi and Maurice Stucke, *Virtual Competition: The Promises and Perils of the Algorithm-Driven Economy* (Harvard University Press 2016).

Sanctions on Digital Platforms : Balancing Proportionality in a Modern Public Square

Enguerrand Marique and Yseult Marique

Abstract

This paper asks which legal tools digital operators could use to manage colliding rights on their platforms in a digitalised and transnational space such as the Internet. This space can be understood as a "modern public square", bringing together actions in the digitalised world and their interactions with actual events in the physical world. It is then useful to provide this space with a discursive framework allowing for discussing and contesting actions happening on it. In particular, this paper suggests that two well-known legal concepts, proportionality and sanctions, can be helpfully articulated within that discursive framework. In a first step, proportionality, a justificatory tool, is often used to suggest a way for managing colliding rights. This paper argues that for proportionality to be useful in managing colliding rights on digital platforms, its role, scope and limits need to be better framed and supplemented by an overall digital environment which can feed into the proportionality test in an appropriate way. This can be provided, thanks to a second step, namely labelling in law the actions digital operators take as sanctions. Sanctions are the reactions organised by digital operators to bring back social order on the platforms. The labelling of these reactions under the legal category of "sanctions" offers a meaningful tool for thinking about what digital operators do when they manage colliding rights by blocking or withdrawing contents and/or accounts. As different types of sanctions can be distinguished, differentiated legal consequences, especially in relation to managing colliding rights, can be identified. Here the role played by the proportionality test can be distinguished depending on the type of sanctions. In any case, for sanctions and proportionality to help address colliding rights on the modern public square, a discursive framework needs to be developed, which depends on the existence of relevant meaningful communities engaging in reflecting on the use of sanctions and proportionality.

Key words:

Digital platforms, transnational normativity, pluralism, regulation, sanctions, proportionality, subsidiarity

Authors details for editorial office

Enguerrand Marique

CRIDES – Institut pour la recherche interdisciplinaire en sciences juridiques – UCLouvain
Place Montesquieu, 2 bte L2.07.01/ C.351
B-1348-Louvain-la-Neuve
Belgium
Tél. +32 10 47 40 97

enguerrand.marique@uclouvain.be

Dr Yseult Marique
University of Essex (United Kingdom)
FÖV Speyer (Germany)

University of Essex
School of law
Wivenhoe Park
CO4 3SQ
Colchester
United Kingdom

ymarique@essex.ac.uk

1. Introduction

Made infamous by the El Paso shooting in the United States in the Summer 2019, 8chan was a far-right website which had developed its central identity around extremism. This site was "*modelled on another message board called 4chan. But in a key difference, 4chan's founder had the power to delete individual boards, while [8chan's founder] was committed to near absolute free speech. When 4chan banned the discussion of the misogynistic harassment campaign known as Gamergate in 2014, 8chan gained in popularity as a staging ground for the campaign*".[192] These different policies about acceptable behaviours on the platforms illustrate how the online world intensifies social, political or cultural offline claims. In the days following the shooting, 8chan was flagged up for encouraging hate speech. As a reaction to popular outcry against this practice, the internet infrastructure provider suspended its services to 8chan. Digital operators are at the interface between sources of legal and social norms shaping individual and collective behaviours. As a contrast to the 8chan story, Pinterest cut searches into anti-vaccines when a measles epidemic broke out in a range of countries following a controversial anti-vaccine campaign.[193] According to Pinterest, public health policy had to prevail over freedom of expression. This kind of reactions undertaken by digital operators are not neutral; they may have drastic consequences on social, political, cultural or economic interests of their targeted users.

Platforms react to users' behaviours using their power of coercion. They actively interfere for preventive, curative or punitive purposes in the interactions between users on the platforms. They limit, restrict, withdraw, curtail, adapt, blacklist, stop or block users' actions for a while or definitively. They seek to discipline some undesirable behaviours (negative) and ensure desirable interactions on the platform (positive). What constitutes desirable interactions or undesirable behaviour is left to them to appreciate. In general, platform operators seek to foster a sense of belonging to a shared community. They will thus seek to foster a sense of trust among the users, especially that the platform constitutes a safe environment for economic transactions.[194] Important values in the offline world such as truth, privacy, property or freedom of expression may be replicated or not so much. If reactions taken at an individual level may seem innocuous, akin to a traffic ticket for a minor speeding offence, they can, taken in an aggregate manner, direct and regulate interactions on digital platforms and their effects beyond the digital platforms in the offline world. The responses adopted by digital operators to undesirable behaviour in the online world are not merely the product of the "invisible (digital) hand". Digital operators create a social order and seek to preserve its integrity, challenging the benevolent picture of social, economic and political life in the online world.

This collective dimension of reactions taken by digital operators in the online world leads this paper to questioning how they operate in legal terms, especially in terms of their democratic and social values. This paper does not address issues pertaining to creating economic value, although these decisions may be taken with such purposes in mind. It asks which legal tools digital operators could use to manage colliding rights on their platforms in a digitalised and transnational world such as the Internet. This space can be understood as a "modern public square", linking together actions in the digitalised world and their interactions with actual events in the physical space. It is then useful to provide this modern public square with a discursive framework allowing for discussing and contesting actions happening on it. In particular, this paper suggests that two

[192] J Wong, « 8chan: the far-right website linked to the rise in hate crimes » *The Guardian*, London, 5August 2019 (available at HTTPS://WWW.THEGUARDIAN.COM/TECHNOLOGY/2019/AUG/04/MASS-SHOOTINGS-EL-PASO-TEXAS-DAYTON-OHIO-8CHAN-FAR-RIGHT-WEBSITE).

[193] C Newton, « Pinterest's work in public health shows the good a smaller social network can do » *The Verge*, 29 August 2019 (available at HTTPS://WWW.THEVERGE.COM/INTERFACE/2019/8/29/20837660/PINTEREST-VACCINE-INFORMATION-SEARCH-RESULTS-PUBLIC-HEALTH).

[194] R Botsman, *Who can you trust?: how technology brought us together–and why it could drive us apart* (London: Penguin 2017).

well-known legal concepts, sanctions and proportionality, can be helpfully articulated within that discursive framework.

First, labelling in law the actions digital operators adopt against undesirable behaviour in the modern public square as sanctions, *ie* reactions to ensure and bring back social order on digital platforms, offers a meaningful legal tool for thinking about what digital operators do when they manage colliding rights by blocking or withdrawing contents and/or accounts. As different types of sanctions can be distinguished, differentiated legal consequences can be identified, especially in relation to managing solutions when rights held by different users collide with each other. In following this approach, this paper departs from current analyses of digital platforms, often grounded in behavioural or regulatory perspectives: nudging and influencing users attract most of the academic attention for their apparent novelty.[195] "Soft" tools (online reputation system including reviews and ratings) may indeed play a specific role in policing platforms. Yet, sanctions as coercive responses to undesirable behaviours are very much part of the toolkit of digital operators. This shifts the focus to the practices of digital operators.

This phenomenon is even likely to increase as users become savvier, want to do more and test platforms' boundaries. This leads to sanctions appearing increasingly often on the radar. Users will see how far they can go. They may exit platforms when they have been punished, they may also want to stay on these platforms (maybe there are not that many alternatives) but seek to voice their discontent: for instance, through participatory structures where they may have their say about what is (or not) allowed on the digital platform and how behaviours should/could be monitored and policed. Sanctions would then be a catalyst for developing a bottom up form of organisation interested in how the collective interactions are regulated: here again sanctions and how they are reacted to may lead away from soft law regulation and the "invisible hand" approaches on digital platforms. Coordination of colliding rights may – at least partly – be analysed through classic legal lenses, such as sanctions, even if this concept may not encompass all actions available to digital operators. In using these lenses, this paper flags up how "hard law" and techniques remain relevant when it comes to adjudicating conflicts among users and between users and platform operators.

Secondly, proportionality, a justificatory tool, is often used to suggest a way for managing collisions between rights. Since Lessig's seminal work,[196] the Internet governance has been analyzed through the paradigm of constitutional law, *ie* an institutional (governance) approach[197] (who is competent to act? what are the decision-making processes?) or perspectives focused on human rights[198] (what is the extent of individuals' entitlement to the protection of their person or

[195] E Carolan and A Spina, « Behavioural Sciences and EU Data Protection Law: Challenges and Opportunities » in A Alemanno and AL Sibony (eds), *Nudge and the Law* (Oxford: Hart Publishing 2015) 8; F Zuiderveen Borgesius, « Behavioural Sciences and the Regulation of Privacy on the Internet » in A Alemanno and AL Sibony (eds), *Nudge and the Law* (Oxford: Hart Publishing 2015); L Belli and J Venturini, « Private Ordering and the Rise of Terms of Service as Cyber-Regulation » (2016) 5:4 *Internet Policy Review* 1-17; T Büthe, « Private Regulation in the Global Economy: A (P)Review » (2010) 12:3 *Business and Politics* 1-38; D Baron, « Private Ordering on the Internet: The Ebay Community of Traders » (2002) 4:3 *Business and Politics* 245-74.

[196] L Lessig, *Code 2.0*. (Basic Books 2006).

[197] See *eg* C Petersen, V Ulfbeck and O Hansen, « Platforms as Private Governance Systems - the Example of Airbnb » (2018) *Nordic Journal of Commercial Law* 38-61; M Finck, « Digital Co-Regulation: Designing a Supranational Legal Framework for the Platform Economy" (2018) 43 *European Law Review* 47-68; B Cannon and H Chung, « A Framework for Designing Co-Regulation Models Well-Adapted to Technology-Facilitated Sharing Economies » (2015) 31 *Santa Clara Computer and High Technology LJ* 23-96; C Reed, *Making Laws for Cyberspace* (Oxford: Oxford University Press 2012); D Koukiadis, *Reconstituting Internet Normativity. The Role of State, Private Actors, Global Online Community in the Production of Legal Norms* (Baden-Baden: Nomos 2015); C Marsden, *Internet Co-Regulation: European Law, Regulatory Governance and Legitimacy in Cyberspace* (Cambridge: Cambridge University Press 2011).

[198] See *eg* C de Terwangne and Q Van Enis, *L'Europe des droits de l'homme à l'heure d'Internet* (Brussels: Bruylant 2019) ; J Venturini, L Louzada, M Maciel, N Zingales, K Stylianou, and L Belli, *Terms of Service and Human Rights: An Analysis of Online Platform Contracts* (Editora Revan, Rio de Janeiro 2016); S Hick, E. Halpin and E. Hoskins, *Human Rights and the Internet* (Palgrave Macmillan UK 2016); R Fisman, and L Michael, *Fixing Discrimination in Online Marketplaces,* HBR, no. December 2016; C Geiger, and E Izyumenko, « The Role of Human Rights in Copyright Enforcement Online: Elaborating a Legal Framework for Website Blocking » (2016) 32:1 *Am U Int'l L R* 43; A Savin, *EU Internet Law* (Cheltenham: Edward Elgar 2013); J Glaser, and K B Kahn, « Prejudice and Discrimination and the Internet » in Y Amichai-

belongings?). Under a classic constitutional paradigm pertaining to the offline space, the proportionality test is often relied on to adjudicate interferences in the rights held by citizens. This paper argues that a transposition of the proportionality test from this usual offline setting to the online world may provide for answers when digital operators are confronted with colliding rights held by platform users or when they wonder whether they should or could interfere with the rights of users. Yet, for this transposition to provide a meaningful solution, the role, scope and limits of the proportionality test need to be better framed and supplemented by an overall digital environment feeding the proportionality test in an appropriate way.

This paper is structured as follows. It first locates sanctions within the conceptual framework of transnational hybrid governance and especially within the "modern public square" (Section 2), before revisiting proportionality, a traditional principle underpinning the balancing of rights and freedoms and the imposition of sanctions (Section 3). It then looks at the practical application of the proportionality test that digital operators need to consider when enforcing sanctions on their platforms (Section 4). In order to go beyond the limits of the proportionality test, this paper then suggests two avenues – one institutional and one community-based to address gaps in the discussions triggered by relying on the proportionality test to address colliding rights on the modern public square (Section 5). Section 6 concludes on further research avenues.

2. Framing Coordination of Differences in the Online World

To analyse the tools available to digital operators faced with colliding rights of users in the offline world, we need to proceed in two steps: first, one needs to understand what the digital space is in terms of interactions, *ie* a modern public square with gate-keepers and umpires, the digital operators, entrusted with specific "warden" functions (**2.1**); secondly, if we see the digital world as an extension of the offline space, under the form of a modern public square, one needs to consider how the law would label digital operators' interferences with colliding rights if we were in offline space, in particular one needs to examine if these interferences may be called "sanctions". One may then test whether it is possible to extent this label of "sanctions" to the reactions taken by digital operators to coordinate colliding rights in the online world, with all the legal consequences attached to this label (**2.2**).

2.1 The modern public square: different voices in a transnational space

Platform operators illustrate perfectly transnational hybrids: they operate across many jurisdictions, providing services in many countries, with key nodes located in strategically chosen countries in order to enjoy favourable legal rules pertaining to contracts, data, taxation, intellectual property for examples. The platform users are only vaguely aware that Uber processes worldwide payments to drivers (except in the US) through a Dutch company resident of the Bahamas. Similarly, Amazon.co.uk has no permanent establishment in the United Kingdom: it is incorporated in Luxembourg. These average same users are even less aware of the technological infrastructure leading some digital platforms to file taxes for their users in Ecuador or in Estonia when a contract is concluded. These features are key economic components in digital platforms however. Digital platforms play multiple roles however: they enable economic transactions between peers and are also vehicles for social and political interactions or behaviour. They do not only facilitate communication and exchanges; they also control communication between the parties to a transaction and facilitate payment, and thus extract a commission or a fee as a price for their intermediation.[199]

Hamburger (ed), *The Social Net. Human Behavior in Cyberspace* (Oxford: Oxford University Press 2005) 247-274. For an analysis under Intellectual Property perspective, N Tusikov, *Chokepoints: Global Private Regulation on the Internet* (Univ of California Press 2016). For an early analysis under freedom of expression, M Siegel, « Hate Speech, Civil Rights, and the Internet: The Jurisdictional and Human Rights Nightmare Comment » (1998-99) 2 *Albany Law Journal of Science & Technology* 375-98. For an analysis under non-discrimination, see A Chander, « The Racist Algorithm? » (2017) 115:6 *Mich L Rev* 1023-45. With regards to online access to justice, E Katsh and O Rabinovich-Einy, *Digital Justice: Technology and the Internet of Disputes* (New York: Oxford University Press 2017).

[199] D McKee, « The platform economy: natural, neutral, consensual and efficient? » (2017) 8:4 *Transnational Legal Theory* pp. 455-495.

This specific situation entails three main features for digital platforms; they are digitalised, transnational and pluralist. First, digital platforms are *digitalised*, *ie* an extension of the offline space into the online world. Hence actions taken by digital platforms may cause significant harm to individuals in the offline space. For instance, when a social media influencer's account or an Uber *driver* is deactivated, it limits their freedom of expression, but it also prevents them from conducting their professional activity. In the same vein, when Uber deactivates one of its *riders*' account because of her misconduct, it impacts on her mobility. These offline consequences need to be included in any assessment of the online space. Secondly, digital platforms are *transnational*, *ie* technological structure facilitating social and economic activity across people and borders or regardless of borders. Thirdly, they are *pluralist,*[200] in the sense that users come with different expectations and values when they operate on the digital platforms, some visible to the other users, some not so visible. In particular, digital users harbour different expectations, attitudes and understanding of how to behave. This is not only due to their different cultural and spatial attachments. This is also due to the changed visibility of their actions. Their visibility is increased as the Internet helps reach and put in contact users in an extensive way. Visibility is also modified because when posting or acting on digital platforms, the users may not know to which audience they will become visible (in time and space).

This paper considers that the conjoined effect of these three features of digital platforms lead to the development of a specific space of political, social, economic and technological interactions, a space best encapsulated under the expression of a "modern public square".[201] This expression reflects that digital platforms constitute an environment where goods, services and data are purchased or exchanged as well as news, opinions, ideas or creative expressions through digital interactions among people from potentially widely different backgrounds and who potentially know each other very well or not at all, who may repeat their interactions in the long term or never again. Relationships and exchanges between users on the modern public square may be conducted with very different strategies in mind. The modern public square cannot expect these relationships and exchanges to be harmonious by themselves. Collisions between rights held by users happen.[202] Doubts arise about their legal solutions. For instance, what is legally, politically or morally acceptable somewhere may not be elsewhere. Identifying where to look for a legal solution may not be an easy task. Overall, the modern public square has to rely on a certain level of organisation,[203] to ensure the coordination of exchanges and relationships.

Digital platforms challenge the classical relationships between power, law and territory as developed in classic international public law. They are part of global cross-border phenomena, with colliding public and private bodies claiming authority to organise social relationships and economic exchanges. States adopt laws applying to the whole of society while functional actors such as digital platforms are hyper-specialised (technically and functionally); they accumulate technical and social capital in a limited range of issues. However digital platforms combine extensive powers in their hands (through the combination of hard law and self-regulatory techniques such as the setting of terms of services): in a way, they concentrate quasi-normative, quasi-executive and quasi-judicial powers because they both create the applications, networks and things under their control and regulate their functioning.[204]

This leads to two questions when comparing the modern public square and the offline space of economic and political actions. One question pertains to the overall organisation of this modern public square in general. To answer this question one may venture to say that this modern public square is not a mere virtual construction with no supporting structure, underpinning organisation and normativity. This paper suggests that this modern public square, as a concept spanning the

[200] *Cfr* M Delmas-Marty, *Ordering Pluralism – A Conceptual Framework for Understanding the Transnational Legal World* (Oxford: Hart 2009).
[201] This notion refers here to these three main features, although there may be other relevant features connected with this notion. These three main features are relevant for analysing sanctions and proportionality as explored within the limits of this paper.
[202] D Harvey, *Collisions in the Digital Paradigm: Law and Rule Making in the Internet Age* (Oxford/Portland: Bloomsbury Publishing 2017).
[203] McKee (above (199)) explains how the "market" is not natural but has required historically structures to ensure that free exchange happens.
[204] L Belli and C Sappa, « The Intermediary Conundrum: Cyber-Regulators, Cyber-Police or Both? » (2017) 8 *JIPITEC* 183 para 7-8; E Marique and Y Marique, « Sanctions on digital platforms – beyond the public/private divide » (2019) 8:2 *Cambridge Journal of International Law* (forthcoming).

online and offline space, needs to be equipped with the necessary systems, institutions and procedures allowing for identifying problematic behaviour threatening its (economic, social or political) integrity and for providing for argumentation and discussion to happen in response to users' behaviour. There should be an accepted reflexive framework for calling actors to accountability, giving them an opportunity to make their case, to be listened to and properly replied to.[205] This type of framework would thus be a way to incentivise functional actors, such as digital platforms, to factor in the potential externalities of their decisions in their decision-making processes.[206]

Another question pertains to the role (powers and duties) that digital operators play on this modern public square. Due to the specific positions of digital operators on the modern public square as regulator and controller, this question asks whether digital platforms do have specific duties with regards to its functioning and in particular identifying and addressing problematic behaviour on the modern public square. In the online world, digital platforms are entities entrusted with functions akin to administrative policing, which puts them in charge of securing the general preservation of public order and morality.[207] There is a need for systems to guarantee the respect of the established rules, and prevent undue inconvenience for the integrity of life in the community. In the offline space, entities entrusted with such functions may be public or private actors, which may lead to distinction between their priorities in discharging their functions.[208] An increasing amount of empirical and socio-legal research shows that public law requirements, controls and accountability mechanisms are currently extended to private actors.[209] It is suggested here that digital operators have a duty to take the appropriate measures to ensure the integrity of users' interactions on the modern public square, subject to suitable accountability mechanisms. This duty has been connected to their role of gate-keeper of the platform[210] or cyber-police.[211] It will be referred to here as their "warden" function, because digital operators can exclude users from the modern public square as much as they can take a range of measures in relation to behaviours threatening the integrity of the modern public square. We turn to labelling these measures in the following lines.

2.2 Digital operators' (re)actions in case of colliding rights: sanctions

What constitutes sanctions has been discussed for centuries. Orbediek provides a good analytical starting point. For him, "*a sanction, […] is any threatened, promised, instituted or declared response on behalf of a group or institution attached to the breach or neglect of a recognized norm, policy, order, law or command done with the implicit or explicit intent of discouraging or preventing any such breach or neglect.*"[212] Sanctions bear interconnected features.

Sanctions are such a response, relying on the exercise of power and taken by digital operators towards undesirable behaviour on the modern public square. Sanctions react to a specific problematic situation or behavior defined as such by a socially recognized rule.[213] Therefore,

---

[205] P Kjaer, « Why justification? The structure of public power in transnational contexts » (2017) (8:1) *Transnational Legal Theory* 8-21.

[206] G Vilaca, « Transnational legal normativity » in M Sellers and S Kirste (eds), *Encyclopedia of the philosophy of law and social policy* (Springer 2017)

[207] L Belli and C Sappa, « The Intermediary Conundrum: Cyber-Regulators, Cyber-Police or Both? » (2017) 8 *JIPITEC* 183 para 18.

[208] In particular, private entities may seek profit maximisation over general well-being (*cfr* recurring issues in contracting out of public services in Europe). L Belli and C Sappa, « The Intermediary Conundrum: Cyber-Regulators, Cyber-Police or Both? » (2017) 8 *JIPITEC* 183 para 19.

[209] A Benish and J Pelisse, « Private companies and administrative justice » in M Hertogh, R Kirkham, R Thomas and J Tomlinson (eds), *Oxford Handbook of Administrative Justice* (Oxford: Oxford University Press forthcoming); J Bell, « Judicial Review in the Administrative State » in J de Poorter, E Hirsch Ballin and S Lavrijssen (eds), *Judicial Review of Administrative Discretion in the Administrative State* (Springer 2019) 3-26; J Freeman, « Extending public law norms through privatization » (2003) 116:5 *Harvard Law Review* 1285-1352.

[210] M Cian, "Online Platforms as Gatekeepers to the Digital World – a Preliminary Issue on Business Freedom, Competition and the Need for a Special Market Regulation" (2018) *Journal of European Consumer and Market Law* 209-10; R Van Loo, "Rise of the Digital Regulator" (2017) *Duke Law Journal* 1317.

[211] L Belli and C Sappa, « The Intermediary Conundrum: Cyber-Regulators, Cyber-Police or Both? » (2017) 8 *JIPITEC* 183.

[212] H Oberdiek, « The Role of Sanctions and Coercion in Understanding Law and Legal Systems » (1976) 21:1 *The American Journal of Jurisprudence* 71-94 at 75.

[213] Such social recognition can be problematic on digital platforms when it comes to users "accepting terms of services" without reading the fine prints. While platform operators take a formal stance on what constitutes a "recognized rule", they will face the criticism of consumers' associations that users are not aware of the substance of the rule. These associations

unintended actions of the platform operators, such as accidental exclusions by the algorithm (*eg* repeated error messages, non-technical recognition of information not included in the system) do not count as sanctions in this respect. This reactive nature of the sanction is important: they constitute a consequence of a behaviour rather than a condition for action. They can be automatized, or not, as a consequence of the behaviour, but cannot be considered *a priori,* without preliminary users' actions or problematic behaviour.

Although their functions can be discussed, sanctions usually have two main purposes, one negative and one positive. On the negative side, sanctions aim to ban a behavior or an action arising from social relationships and economic exchange. They aim to discourage users from adopting certain problematic behaviors, such as preventing illegal contents and harassment. On a more constructive side, sanctions can be retributive (where the individual needs to "pay back" to the community for the infringement), reparative (where the individual is isolated to place the community back in a state of peace as if the violation had never taken place),[214] or pedagogic (when they help users to identify, learn and transmit the core values the platform wishes to promote).[215] For practitioners (and academics), sanctions can thus express how amount digital operators defend certain values over others and counterbalance the marketing narratives with their actual practices. Such approach is indeed defended (with more or less success) by labor lawyers trying to characterize the sanctioning power of Uber and Deliveroo as a supervision and control mechanism giving rise to an employment relationship rather than an independent "partnership" agreement.[216] Conversely, failure to act against certain behaviors, despite recognizing them as unwanted in terms of services or marketing practices, amount to a policy choice that the underpinning value is not worth enough fighting for.

In the offline world, these actions would fall within the definition of "sanctions", with all the legal consequences attached to that legal qualification, including the availability of a judicial or independent review mechanism.[217] Principles of review need thus to be established on the modern public square. This paper proposes to examine the principle of proportionality under these new lenses.

3. A Test for Elucidating the Human Rights in Conflict in a Modern Public Square

Once the reactions from digital operators are labelled as a known legal category, that of sanctions, the legal consequences usually attached to this legal category can be investigated further.[218] Here the requirement of proportionality takes a special role for its principled use in the case of colliding rights (**3.1**) and for the modalities that have been suggested for this test in the digital space (**3.2**).

3.1 Proportionality in Principle

Faced with colliding rights on the online world, such as freedom of expression *vs* the right to privacy or freedom of expression *vs* public health policies,[219] platforms have a range of possible options: from doing nothing and letting users act as they wish to giving priority to specific rights or specifying a clear hierarchy between the rights at stake for instance. There is a preliminary question here: should digital operators interfere with users' interactions at all? It may be argued

would thus challenge that platform "can even take such measures". However, the formal approach chosen by the platform enable lawyers, litigators and practitioners to characterize platform behaviors as sanctions and therefore to find the need for adequate protection as sanction.

[214] *Eg* M van de Kerchove, « Les fonctions de la sanction pénale. Entre droit et philosophie » (2005) 127:7 *Informations sociales* 22-31, 27-30.

[215] For such an interpretation of Durkheim, see J Feinberg, *Doing and Deserving* (Princeton: Princeton University Press 1970) 102.

[216] C Wattecamps, A-G Kleczewski and E Marique, « Challenges related to law for the platform economy: A fresh look at some important dichotomies » (2017/3) *Reflets et perspectives de la vie économique* 57-95, 66-71.

[217] Following that line of reasoning, see *eg* SL Kalėda, « The Role of the Principle of Effective Judicial Protection in Relation to Website Blocking Injunctions » (2017) 8 *JIPITEC* 216 para 35.

[218] Other legal aspects than proportionality, such as legality or procedural guarantees, are not investigated within the limited remit of this paper.

[219] See Introduction the examples of 8chan and Pinterest.

that any interferences with users' activities on the digital platforms would be akin to censorship.[220] However, it may be advanced that the general interest of the "modern public square" requires that at least some level of policing and monitoring of users' activities on the platforms is organised. This allows to maintain a good environment for social interactions and exchanges on the digital platforms. The question shifts then from the principle of interference on the modern public square to the modalities of this interference, and especially how digital operators collect the information needed to carry out their minimal desirable monitoring, process it, organise their decision-making on their basis and implement it in practice. If one accepts this position, the proportionality test may be a tool to ensure that digital operators follow a reasoning process that can be called for account.

The proportionality test of public action is widely accepted as a way for controlling public power[221] and sanctions[222].[223] This test may be used at different stages of the sanctioning process, *ie* when setting the normative framework, monitoring compliance, imposing sanctions, implementing them and/or adjudicating between colliding rights in specific or general circumstances; or when reviewing them. Normally, the proportionality test implies that decision-makers should only follow a course of action if 1) their objective is legitimate; 2) their means is necessary to achieve their objective; 3) no means would entail a lighter encroachment of the right at stake; 4) the means is proportional (*sensu stricto*) to the objective to be achieved. The conceptual foundations and modalities for this test are endlessly discussed.[224]

Proportionality has been much criticized for its apparent neutrality. It would allow judges to review administrative actions without imposing their own values and priorities about which course of action is preferable in a specific case. However value judgements may be hidden at each stage of the reasoning process,[225] at the levels of both administrative and judicial decision-making. This challenges the uniform application and the predictability of the proportionality test. Yet, it is often said that the proportionality test provides a tool for requiring decision-makers to explain their reasoning process. In so doing, the proportionality test helps foster a culture of justification and persuasion.[226] In order to allay these criticisms, Alexy suggested an abstract "weight formula" assessing the interferences in the rights at stake in a more objective way.[227] In a modern public square, an additional challenge arises, that of identifying the main actors of this culture of justification and persuasion.

---

[220] Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, 16 May 2011, Human Rights Council, 17th session, para 40-43.

[221] For an attempt to reconcile the American and the European approaches to balancing rights: K-H Ladeur, « A critique of balancing and the principle of proportionality in constitutional law – a case for 'impersonal rights'? » (2016) 7:2 *Transnational legal theory* 228-256.

[222] *Eg* for administrative sanctions: C.E. fr., 19 mai 1933, *Benjamin*. A comment is available on the French Conseil d'Etat's website: http://www.conseil-etat.fr/Decisions-Avis-Publications/Decisions/Les-decisions-les-plus-importantes-du-Conseil-d-Etat/19-mai-1933-Benjamin.; for discussions in relation to criminal sanctions: *eg* A von Hirsch, « Proportionality in the philosophy of punishment » (1992) 16 *Crime and Justice* 55-98; J Deigh, « Punishment and proportionality » (2014) 33:3 *Criminal Justice Ethics* 185-199; J Deigh, « Punishment and proportionality: Part 2 » (2016) 35:1 *Criminal Justice Ethics* 21-38.

[223] A Stone Sweet and J Mathews, « Proportionality balancing and global constitutionalism » (2008) 47 *Colum J Transnat'l L* 72-164.

[224] *Eg*: F Urbina, « Is it really that easy – A critique of proportionality and "balancing" as reasoning » (2014) 27:1 *Canadian Journal of Law and Jurisprudence* 167-192; M Klatt and M Meister, « Proportionality – A benefit to human rights – Remarks on the I-Con controversy » (2012) 10 *Int'l J Const L* 687-708.

[225] *Eg*: S Greer, « 'Balancing' and the European Court of Human Rights: A Contribution to the Habermas-Alexy Debate » (2004) 63:2 *Cambridge Law Journal* 412-434.

[226] D Dyzenhaus, « Proportionality and deference in a culture of justification » in G Huscroft, B Miller and G Weber (eds), *Proportionality and the rule of law – Rights, Justification, Reasoning* (Cambridge: Cambridge University Press 2014) 234-258.

[227] R Alexy, « The construction of constitutional rights » (2012) 91:3 *Revue française de droit constitutionnel* 465-477; R Alexy, « Proportionality, constitutional law, and sub-constitutional law: A reply to Aharon Barak » (2018) 16:3 I•*CON* 871–879.

3.2 A specific proportionality test for the digital space: The internet balancing formula[228]

In the context of colliding rights in the online world, Susi builds on Alexy's weight formula to devise in concrete and practical terms a test that platform operators could use when making decisions pertaining to response to problematic behaviors and interactions happening on their platforms. While Alexy's balancing formula is expressed at a high level of abstraction to balance the intensity of interferences between human rights,[229] Susi gives an operational and concrete translation of the formula in the case of conflicts between the right to privacy and freedom of expression. His approach is best summarized as follows.

The first step of the reasoning is to exclude any balancing in case of hate speech, as hate speech should always be banned from digital platforms. The second step is to calculate the "value" of the right to privacy, which is the addition of the following factors: the vulnerability of the individual due to internet technologies, the interference in privacy (calculated in taking the perspective from an neutral onlooker), and the element of time with time passing being assumed to be a decreasing factor in terms of interference.[230] The third step in the reasoning pertains to the calculation of the numerical value of the freedom of expression. This is based on the addition of three elements:

> 1) the level of public interest in the matter (*ie*: minor public interest is public interest measured in terms of the local community affected by the matter; medium public interest refers to cases when the larger community is affected but with no immediate direct impact on the lives of the majority; and significant public interest to cases when matters affect the entire nation or have a direct effect upon the lives of the majority);
>
> 2) the determination as to whether the information concerns a public figure;
>
> 3) the ways in which information has been obtained (either legally or not, either morally or not).[231]

In addition, on each side of the formula, a further factor needs to be added, which Susi calls "empathy". The exact meaning of this concept is not clear, but its function is to ensure that human agency has to intervene in some cases at least, namely when there is a break even between the two sides of the equation and when moral reasons command it.[232] Finally, Susi reserves a specific treatment to cases where divulgation of the information trumps privacy due to its contribution to historical truth.[233]

Susi stresses that this formula should give a rational answer to cases where the right to privacy and the freedom of expression collide, a formula that would allow anybody, the "citizen journalist" included, to decide if an information should or could be published on a digital platform.[234] Anyone – even without legal training – would be able to use it.[235] This would thus prevent abuse and censorship from digital platforms in refusing to publish online information.[236] A higher degree of

[228] M Susi, « The Internet balancing formula » (2019) 25 *European Law Journal* 198-212.
[229] R Alexy, « Mart Susi's internet balancing formula » (2019) 25 *European Law Journal* 213-220.
[230] Susi above (228) 205.
[231] Susi above (228) 205-207.
[232] Susi above (228) 207.
[233] Susi above (228) 208.
[234] Susi above (228) 199.
[235] Susi above (228) 204.
[236] This discussion becomes especially relevant when examining the current European Union reforms on copyright in the Digital Single Market (Directive (EU) 2019/790). This reform might indeed change the current scope of online freedom of speech and other human rights. It will also shift the liability standards for digital operators' actions and thus alter their

transparency would indeed be gained. The formula would provide standards to carry out the assessment and the argumentation behind the decision.

4. Operationalizing Proportionality in Case of "sanctions": Testing Limits

In developing their platform architecture, policies and implementation tools, digital operators make a range of choices pertaining to the best strategy to address colliding rights on an on-going basis. In their search for finding a technique that could be accommodated with their technologies, they might be interested to turn to the internet balancing formula. If this approach may be a positive addition in their toolkit, there are however areas where caution is required (**4.1**). Finding means for supplementing the internet balancing formula might be a way forward. Here a more nuanced understanding of the reactions taken by digital operators is offered to distinguish cases where a proportionality test may be used from cases where this may be less the case (**4.2**).

4.1 Potential and Challenges to the Internet Balancing Formula

Susi's formula translates the proportionality test into a duty for digital platforms to ensure that their reactions are tailored to each of their users, according to the (relevant) data available to the platform in question following a balancing test between the right to privacy and freedom of expression.[237] As this approach is appealing to regulate users' behaviours on platforms, it deserves further attention. Its operationalisation faces challenges even if taken on its own terms. As Susi calls for further empirical and philosophical discussions, this paper asserts that a strength of the internet balancing formula is to launch an argumentative and discursive process on the balancing of colliding rights on the modern public square by private actors such as digital platforms. One key issue is however who will be active contributors to this process. Although Susi invites "citizens journalists" and lay people to use the formula and fine-tune it, such a formula is more likely to offer a reflexive framework and argumentative scheme for the use of an epistemic community, made up of digital operators in the first place, some key users, and maybe some social groups especially equipped for this role:[238] using the formula will indeed require a range of data and skills (including investigation and research skills), especially as some of this data will not be available online. Indeed, if we accept the interactions between the online and the offline worlds, some of the interferences will take place in the offline world, and it is not clear how the information about it will be captured by the Internet (more importantly even, whether it should be captured by the internet in the first place). In addition, using the formula will require a good knowledge of the law and the case law across various jurisdictions, a good grasp of the empirical reality[239] and access to a range of statistical tools about the ways in which platforms operate (their market, audience, countries of operations, specific groups or objectives). Here a technical support will be needed for fostering the concrete use of the formula.

In addition to questions related to the intended users of the internet balancing formula, one may argue whether human rights are ever quantifiable as a matter of principle or about the nitty-gritty aspects of the formula,[240] but suffice here to mention four general challenges for operationalizing the formula: values; human factors; individual dimension and argumentative space for using the formula. In a way all these challenges turn around one key question: the neutrality of the judgement exercised by the entity operating the formula. This neutrality is not guaranteed. The formula rests on a series of normative choices which could be contested or further argued about in order to test it.

---

willingness to commit money in compliance costs. This directive is an opportunity for digital operators to improve their review procedures in order to limit their exposure to liability in the face of undue removal of content.

[237] On the individual tailoring of decisions by algorithms, see for instance E Marique and A Strowel, « Gouverner par la loi ou les algorithmes: de la norme générale de comportement au guidage rapproché des conduites » (2017) 10 *Dalloz IP/IT* 517-521.

[238] See below Section 5.2.

[239] Susi mentions the need to base the formula on empirical experience, but the methods, project designs and the ways in which the findings would feed the formula are not explained (Susi above (228) 204).

[240] Professor Alexy flagged up some problems in his answer (R Alexy, « Mart Susi's internet balancing formula » (2019) 25 *European Law Journal* 213-220).

Firstly, the internet balancing formula has two major limits factored into it: first, hate speech is always banned, so that the formula does not apply when hate speech has been identified; secondly, historical truth can never be suppressed, so that the formula is not relevant either in that case. If one can fully agree with these two limits, they are more complicated in practice. Does hate speech have a universal definition? What is historical truth and how is it supposed to be ascertained? What if different groups have different claims about what constitutes historical truth? Is it really the function of digital operators to adjudicate this? In addition, these two limits assume minimal values that cannot be undermined by digital platforms on the modern public square. This leads to questioning whether these values are the only relevant ones or whether other values need to be included in a form of minimal core, a kind of "rule of law" for ensuring the integrity of the modern public square, and how they could be identified.

Secondly, the internet balancing formula seeks to recognise the human dimension of the modern public square, through the factors of empathy and internet vulnerability. This approach may be a way to address a subjective assessment of colliding rights by digital operators. There are problems here. In practice, it is difficult to justify why internet vulnerability is given a constant value of "1" in the formula and why empathy is given the same value in the two sides of the formula. On the one hand, the constant value "1" allocated to internet vulnerability in the formula gives more weight to the right to privacy compared to freedom of expression while different users or individuals may be more or less vulnerable to such exposure.[241] On the other hand, empathy is supposed to be used in the formula when moral grounds justify it. Yet, managing colliding rights will nearly always have a moral dimension of some kind, which renders unclear whether it is supposed to be included across the board or in borderline cases that remain to be identified.

Thirdly, the internet balancing formula is geared to address conflicts between the right to privacy and freedom of expression, and their individual variations. Yet, other rights – including collective rights and freedoms – may also have to be included in any balancing exercise as well as the longer and shorter terms effects of these balancing exercises.[242] It is not clear how the internet balancing formula can be expanded to take these aspects into account.

Finally, that the internet balancing formula contributes to confidence and transparency on the modern public square platforms needs to be acknowledged as important. Two points can be made here. The first one relates to the discursive and argumentative framework within which proportionality is used by judges for adjudicating offline colliding rights. Judges are acting within specific argumentative constraints: they need to convince a range of players that their decisions are the right ones. These constraints may work in different ways depending on the judiciaries, but they usually have two key features. Firstly, judges usually work within tight procedural constraints where both parties have the opportunity to make their case, to be heard, to be listened to and to be responded to, which are important components for recognising the human dignity of the players and to respect them.[243] Secondly, judges normally work within a constitutional context where the legislature can react to judicial decisions if they disagree with them. This step may or may not be activated, or only activated in marginal cases, but it does exist. This has two key consequences. The first consequence is that judges often act incrementally, often steeped in pragmatism.[244] The second consequence is that judges work within a community of legal professionals, dialoging, resisting and communicating with them. These key features of the discursive and argumentative framework underpinning the use of proportionality in offline conflicts of rights are not guaranteed in the case of the internet balancing formula. At least Susi does not suggest ways in which they would be replicated. One may be forgiven to think that the formula is meant to be included in an algorithm of some kind and that contestations and challenges may arise, although Susi does not spell out the grounds, space, processes or institutions which would host these challenges. It is thus not clear whether digital platforms think of themselves as being held to convince peers and institutional actors (regulators? investors? stakeholders?) and

[241] Alexy above (227) 216-17.
[242] We cannot agree with the assumption made by Prof. Susi that the passing of time always decreases the interferences with rights.
[243] J Waldron, « The rule of law and the importance of procedure » (2011) 50 *Nomos*, Getting to the rule of law, 3-31.
[244] We leave aside here the discussions about judicial activism and "*gouvernement des juges*", although this very discussion illustrates that judges need to be persuasive in their judgments.

according to which criteria (economic performance? corporate social responsibility in some form? something else?).

Overall, the modern public square does not include an institutional or procedural framework similar to what exists in the offline world for judges. In particular, the modern public square – with its key features of being digitalized, transnational and pluralistic – comes with a highly disaggregated audience. If and when digital operators want to rely on a proportionality test such as the internet balancing formula (or a possible variation thereof) to exercise their warden functions, the proportionality test may be used as a focal point of attention, drawing the interests of users likely to be affected by its use, but the whole modern public square will need to be equipped with further structures, institutions and processes so as to ensure that the components of the formula, its outcomes and modalities are effectively subject to scrutiny, discussions and accountability. In this developmental process, it may be important to fine-tune the cases where a proportionality test carried out by digital operators may be relevant. At an abstract general, a proportionality test can happen at four stages or moments: when a norm identifying a problematic behaviour is set (with the principle and its consequences), when a decision is taken in relation to a concrete problematic behaviour, when the specific reaction to this concrete problematic behaviour is chosen, and when an independent review is carried out into this reaction. Here, distinguishing between types of reactions by digital operators about colliding rights can be made can help better understand when digital operators may carry out a proportionality test for ensuring the integrity on the modern public square. A typology of these reactions is offered in the next sub-section.

4.2 Typology of Sanctions in the Modern Public Square

In the online world, digital platforms sanctioning processes can be broken down into four main types: 1) platforms acting on behalf of public bodies to sanction illegal behaviors (under EU/international or domestic law); 2) platforms using a discretionary power to implement policies or legal obligations adopted by public bodies; 3) platform operators imposing sanctions for behaviors they have identified themselves as undesirable; and 4) sanctions imposed following a quasi-judicial process organized by platforms to adjudicate between users.

In the first category of sanctions, operators may have to take action to take down content on the platforms, so as to avoid becoming liable themselves for hosting or curating illegal content. Good illustrations of this can be found in the secondary liability system protecting intellectual property rights[245] or in the famous *Yahoo!* case, where *Yahoo!* had to restrict access to French users on the part of its platform where nazi memorabilia were put to auction, so as to comply with French law.[246] In this case, platforms seem to act as delegatees or arms of the state and public bodies: they are in a position to take the necessary actions to ensure that legal obligations or interdictions are complied with. Here a private actor (the operator) extends the reach of public bodies onto the platforms. The private assists and enhances/strengthens the capability (action) of public bodies at a practical/material level.

In this first case, a statute clearly puts an obligation or a duty on digital operators[247] or a judgment enjoins them to comply with the law.[248] In such scenario where a legal duty to act (by statute or by court order) is clearly established, the platform operator does not have to make any proportionality assessment with regards to the principle of an interference in users' rights nor in its application. It merely has to assess whether the response triggered by the violation is proportionate to the infringement (to the extent that the expected response is not dictated by statute or court decisions). For instance, hate speech (or conversely, historic truth) is *ipso facto* outside of the scope of the Susi's formula. In such a scenario, states take the principled decision

[245] R Drath, « Hotfile, Megaupload, and the Future of Copyright on the Internet: What Can Cyberlockers Tell Us About DMCA Reform » (2012) 12 J *Marshall Rev Intell Prop L* 205.
[246] Paris Trial court, *Ligue Contre le Racisme et l'Antisémitisme v Yahoo!*, 20 November 2000; *Yahoo! Inc., v La Ligue Contre Le Racisme et l'Antisémitisme*, 379 F 3d 1120 (9th Ct, 23 August 2004).

[247] See for instance J Riordan, *The Liability of Internet Intermediaries* (Oxford: Oxford University Press 2016) 114–16
[248] Paris Trial court, *Ligue Contre le Racisme et l'Antisémitisme v Yahoo!*, 20 November 2000; *Yahoo! Inc., v La Ligue Contre Le Racisme et l'Antisémitisme*, 379 F 3d 1120 (9th Ct, 23 August 2004).

to forbid (or conversely, authorize) the publication of such content. Here, states assess the proportionality, not the digital operators who have no scope left to exercise a proportionality test.

In the second category of sanctions ("co-regulation"), the digital operator acts on behalf of public bodies but the exact link and its legal nature can be difficult to pin down. The digital operator may seem in a position where it can easily implement legal obligations and enforce them on the platforms. Yet, in practice, legal obligations do not get implemented in a void, through a magic wand or in a mechanical fashion. Some norms provide a large leeway for the operators to decide whether to respond or not to the violation of some legal duties and the extent of this violation. Digital operators have to develop processes and techniques to detect infringements and to decide what to do with them. For instance, Facebook hired teams of content moderators/reviewers in Germany to ensure that users acting on its platform comply with the German Network Enforcement Law and German criminal law. But this team has several options in front of them when examining content. Removal of certain sets of content uploaded is not the only possible response. Other solutions are available for platforms, such as retrograding the rank of content, making warnings or disclaimers about the content (as for website allowing classified ads for escort services or pornographic video sharing) or restricting comments (as online news website aware that some comments might go quickly off-topic because of political sensitivity). The platform, as an actor of the modern public square, is required to examine the proportionality of the reaction (the sanction) to the user's actions (users' misbehavior). This means that this passing of compliance monitoring functions on digital operators has a transaction cost for them. This also means that this implementation system can be flawed in many ways – maybe just because the scope of the obligations may be difficult to ascertain, because there may be conflicting obligations (*eg* with different jurisdictions claiming the right to regulate some actions) or because the operators decide to be over-inclusive in their systems in order to avoid their own liability.

In the third category of sanctions ("self-regulation"), digital operators develop their own ordering and discipline: they may decide which actions and behaviors are allowed on digital platforms, banning other actions as undesirable.[249] Digital operators have here a wide scope to exercise a two-prong proportionality test, with regards first the principle of interference and secondly the modalities of the sanction.

At the level of principle, digital operators, acting in their capacity of warden on the modern public square,[250] decide what the code of conduct and the social norms on the platforms are. They police behavior so as to ensure a specific ethos on the modern public square. An illustration of this kind of self-regulation is the decision taken by Facebook to ban nudes on pictures.[251] Here, digital operators develop their own sovereign spheres of power: they decide the norms that have to be complied with, the ways in which they want to monitor compliance and their enforcement. This may all seem to be the case of private operators exercising private powers under contractual agreement. Yet, platform power imbalance means users become part of the contractual ordering through adhesion, which is more like a social organization where the rules of the games have the power to include and exclude users, and which may discipline the membership against their will. In that case, the essence of this process seems to be close to that of a sovereign's authority, not based on the freedom of will core to private entities. In addition, the policing of the platform is deemed, under economic theories, to be done "in the public interest", as to attract as many users as it can.[252] The organ may be private in its legal form, but the very essence of the relationship is public.[253] In this warden function, digital operators may be in a position to apply the internet

[249] L Belli and J Venturini, « Private Ordering and the Rise of Terms of Service as Cyber-Regulation » (2016) 5:4 *Internet Policy Review* 1-17.
[250] See above Section 2.1.
[251] Facebook Community Standards, '12. Violence and Graphic Content' <https://www.facebook.com/communitystandards/graphic_violence>; see also Facebook Community Standards, '13. Violence and graphic content' <https://perma.cc/KZ39-CZUA> both sources accessed 22 July 2019.
[252] Przemyslaw J Palka, « Terms of Service are not Contracts — Beyond Contract Law in the Regulation of Online Platforms » in S Grundmann (ed) *European Contract Law in the Digital Age*, vol 3 (European Contract Law in the Digital Age, Intersentia 2018) 136–61. *Add* E Marique and Y Marique, « Sanctions on digital platforms – beyond the public/private divide », (2019) 8:2 *Cambridge Journal of International Law* (forthcoming). Even platforms identifying a niche (such as hate-propagating platforms such as 4chan and 8chan) try to increase the number users within the defined niche, eliminating or decreasing the reasons for users to access competitor websites.
[253] E Marique and Y Marique, « Beyond the Public and Private Divide on Digital Platforms? Revisiting Power Relationships » in E Bani, E Rutkowska-Tomaszewska and B Pachuca-Smulska (eds), *Public Law and the Challenges of New Technologies and Digital Markets* (CH Beck 2019 forthcoming).

balancing formula to its fullest extent (taking into accounts all the caveat discussed in Section 4.1).

At the level of modalities, platform operators select the most appropriate sanction amongst a large scale of possible options following the proportionality test, on the same model as already developed in the discussion on the second category of sanctions.

In the fourth category of sanctions, digital operators do not act on their own initiative or on behalf of public bodies: they act as judges in adjudicating disputes between users on the platforms. Operators then develop dispute resolution procedures which may take a range of modalities, some being mostly automated, some relying on human judgment.[254] They need to design a system able to cope with a large number of issues and yet give users confidence that it is impartial / not too much biased. As the dispute is of a private nature (between two users), the adjudicating platform has little leeway to examine proportionality because the dispute is in the hand of the parties and respond to framework of relationships under private law, which solve most issues. If any difficulties arise, the proportionality test may be used as a default option, would one of the three first sets of circumstances described above emerge.

5. Plugging Gaps in the Proportionality Test: Towards more Subsidiarity?

The previous sections point that digital operators may use the proportionality test in a series of cases when they police undesirable behaviours in the modern public square; yet, there reactions (or lack of reactions) have to be called to account. Here, we see that the proportionality test calls for some kind a justificatory structure to be developed. Such accountability may be grounded in a subsidiarity principle for borderline cases. This section explores institutional mechanisms to support such an accountability. On the one hand, this paper suggests that judicial control should complement the application of the proportionality test by digital platforms (**5.1**). On the other hand, epistemic communities should develop the argumentative framework to adapt it to the necessities of the evolution of digital operators' business models (**5.2**).

5.1 Proportionality Test in Reviewing Sanctions

Digital platforms sanctioning process can go wrong. From abusive content removal to the lack of removal of hate speech through the undue automatic imposition of penalties[255] or fines[256], the sanctioning process is subject to errors and mistakes. Users who feel betrayed by the digital operators, suffer losses or want to obtain a remedy against these decisions could consider several legal avenues: breach of contract and extracontractual liability are at the core of the discussions. Contractual claims are often unfit to respond to user's need because of the large discretion attributed to platform operators in the terms of use. Extracontractual claim raise similarly issues in terms of legal base for establishing a ground of liability. Because of the cost and the inappropriate character of these responses, other options need to be reviewed. Self-regulatory industry-wide review bodies have been considered,[257] but they have been found to be failing, so that they cannot not be relied on too heavily.[258] Facing the hurdles of these three first avenues, this paper proposes to consider independent review procedures for decisions undertaken by digital platforms on the model of what happens in administrative law, a legal field that developed techniques to control power. In particular, the *Global Administrative Law* scholarship recognizes that private bodies carrying out regulatory functions at a transnational level should be submitted

[254] See C Rule, « Designing a Global Online Dispute Resolution System: Lessons Learned from eBay» (2017) 13:2 *U St Thomas LJ* 354–364. Compare with Aide Youtube, « Qu'est-ce Qu'une Revendication Content ID? » <https://support.google.com/youtube/answer/6013276> (FR) accessed 22 July 2019.
[255] *Eg* financial gains from a video posted can be reoriented to other users, despite the video amounting to original content.
[256] *Eg* penalties for misbehaving during a ride-sharing, for instance dirtying the car typically include the cost of cleaning the car but also a compensation for the driver who will not be able to take passengers during the time of cleaning as well as some incentive to prevent this behavior to take place again in the future.
[257] *Eg* Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, 6 April 2018, Human Rights Council, 38th session, at 58.
[258] See *eg* https://ec.europa.eu/digital-single-market/en/news/second-monthly-intermediate-results-eu-code-practice-against-disinformation; https://ec.europa.eu/digital-single-market/en/news/fourth-intermediate-results-eu-code-practice-against-disinformation. See also T Schulz, « Does Online Dispute Resolution Need Governmental Intervention? The Case for Architectures of Control and Trust » (2004) 6:1 *North Carolina Journal of Law & Technology* 71-106.

to control similar to these existing in administrative law. Indeed, "*due to the lack of international public institutions, they often have great[..] power and importance*".[259]

Independent review should thus take place, to control digital operators' decision, and if needed to overturn/quash them and grant adequate compensation to users. This idea could, at first sight, seem far-fetched. However, in practice, such administrative review already exists. Public authorities already review and sanction companies (mis)behaviors. This control goes indeed together with all the rights and duties attached to decision-making by public authorities. Data protection agencies can receive complaint with regards to inappropriate data practice in relation to users' data or inadequate privacy policies[260]; competition law authorities can review business decisions; the English CMA can receive consumers' complaints on platform handling of users;[261] in the United States, the FCC also has similar powers.[262] As it stands, thus, this administrative review is *not* centralized, either in term of territorial-jurisdiction or in term of substance-matter. It is diffuse across a number of review bodies.

While judicial review under contractual standard may amount to a strict interpretation of the (breached) duties, extracontractual and administrative reviews are about controlling the powers and the abuses of platforms in their decision making. Additionally, while contractual and extracontractual claims do not have a recognized standard of proportionality review, judicial review of power has developed techniques to carry out a proportionality test. Indeed, amongst other principles applicable to such organizations, such as participation, transparency and reasoned decisions, the *Global Administrative Law* establishes the necessity for independent or judicial review. Standards considered include the respect for legitimate expectations, means-end rationality, avoidance of unnecessarily restrictive means but also significantly proportionality.[263] It is therefore possible that, in practice and subject to peculiarities of domestic legal systems, digital operators' power could be included in the existing administrative review process.

In practice, the independent reviewer needs to assess whether the principles of interference, the principle of a sanction and its casuistic modalities is proportional will be the usual proportionality test, as currently developed in the legal scholarship. While Susi's formula would constitute an element to be considered, parties could advance a series of other arguments. If this proposal of independent review becomes ever part of positive law at a large scale/global level, this new set of case-law will also have to be integrated in the platform operators decision-making process. This will hopefully feed a virtuous loop in order to improve the proportionality test for a modern public square.

5.2 Proportionality test in epistemic communities

While the paper has outlined the assessment of proportionality in a modern public square by lawmakers, digital operators and independent reviewers, this principle needs to evolve and be discussed in order to ensure its evolution and uniform use. Epistemic communities should develop around the argumentative framework to adapt the proportionality to the necessities of the

[259] B Kingsbury, N Krisch and R B Stewart, « The Emergence of Global Administrative Law » (2005) 68 *Law and Contemp Probs* 15, 23. *Add* D Oliver, *Common Values and the Public-Private Divide* (Cambridge: Cambridge University Press, 1999) also advocated controlling private powers on the model of controls existing for public powers.
[260] See *eg* in France, Commission national de l'informatique et des libertés Délibération n°SAN-2018-011 du 19 décembre 2018 prononçant une sanction pécuniaire à l'encontre de la société Uber France SAS (available on the official website : legifrance).
[261] See *eg* https://www.gov.uk/cma-cases/online-travel-agents-monitoring-of-pricing-practices .
[262] See *eg* https://consumercomplaints.fcc.gov/hc/en-us .
[263] B Kingsbury, N Krisch and R B Stewart, « The Emergence of Global Administrative Law » (2005) 68 *Law and Contemp Probs* 15 at 38-41.

evolution of digital operators' business models.[264] As already pointed in Section 5.1, only key users will be equipped with the skills, knowledge, data and tools to operate the formula.

Therefore, to ensure full accountability of digital platform operators, and of their users, a community of "proportionalists" need to emerge. Such community would establish regular, public forums where data available on the (1) forms of interferences and behaviors prohibited, (2) modalities of sanctions, (3) complaints by users and (4) the result of the review process would be shared by the different actors. Interpretation of these results would be subject to discussions, as well as means to improve the existing assessment.

Publicity of these discussions is absolutely necessary. Indeed, the proportionality test as currently framed lack an important component: the public interest factor. While the exclusion for "historic truth" amounts to one form of public interest, the assessment as proposed under the Internet balancing formula is an individualist approach, with little space for collective, social and cultural rights. This leads to asking the question whether one needs to understand conflicts on platforms from the perspective of balancing rights, freedoms and interests at the level of the beneficiaries (adopting then a subjective approach) or of balancing norms at the point of their sources and authors (adopting then a more objective approach). The contribution of a pluralistic approach may be to accept that these two approaches do not exclude each other automatically, but that a method to coordinate them (at an aggregate level) may offer a way forward. Indeed, power on individuals stands at the interface between these two questions.

In addition, publicity is required to go against the private ordering powers that platform operators enjoy and would limit their discretion, facilitate the functioning of a modern public square, akin to an assembly where groups in their diversity/pluralism can be express their opinions and be heard.

Last but not least, such discussion would ensure the credibility of the proportionality assessment, and its openness to criticism and feedback. The methodology could therefore be refined after discussions with NGOs as well as with professional organizations (*eg* ethics boards for web developers[265]). The discussions would also be an opportunity to educate users and make them learn discursivity and alterity, and to accept differences.

6. Conclusion

This paper asked how digital operators could manage colliding rights on the modern public square. Suggesting that digital operators have special duties to maintain the integrity and social order as warden of this modern public square, this paper argues that reactions from digital operators towards undesirable behaviours threatening the proper functioning of the community can be labelled as "sanctions". In choosing the principles and the modalities of their reactions, digital operators may rely on a proportionality test when exercising their discretion.

So in the case of 4chan, the digital operators ruled out harassment, hate speech and similar in order to protect users and third parties. Their banning of some abusive behaviours constitutes a reaction showing the core priorities to the platform, *ie* a strict commitment to comply with the legal framework and to avoid any legal liability. As a by effect, this reaction restored social peace within the community of users. In reaction to this limitation in free speech, a break way group left 4chan to set up 8chan. 8chan's policy was to guarantee freedom of expression as radically as possible.

[264] S Quack, « Expertise and authority in transnational governance » in R Cotterrell and M del Mar (eds), *Authority in Transnational Legal Theory – Theorising Across Disciplines* (Cheltenham: Edward Elgar 2016) 361-386.

[265] I Sample, « Maths and tech specialists need Hippocratic oath, says academic » *The Guardian*, London, 16 August 2019 (available at https://www.theguardian.com/science/2019/aug/16/mathematicians-need-doctor-style-hippocratic-oath-says-academic-hannah-fry).

This led eventually to the platform shutdown. These illustrations are extreme: the proportionality test has been exercised in neither cases. However, they show the possible scope and opportunities for a proportionality test. In order to avoid situations as critical, digital operators have interests to adopt moderating techniques, *ie* to balance colliding values and rights with a proportionality test. However, the long-term sustainability of such a strategy relies on developing a relevant and shared justificatory framework.

Digital platforms should indeed be accountable for their reactions to undesirable behaviours and for breaching their commitments. However, they should keep a certain level of discretion in exercising their warden functions on the modern public square. This rather liberal approach does not exclude that the functioning of the proportionality test to assess sanctions generates ethical behavioural norms for fostering diversity and integrity in the modern public square.

Overall, the use of the proportionality test to assess sanctions on the modern public square should benefit from further analysis into its contribution to the rule of law and to discussions into issues related to global justice on the modern public square.

7. References

R Alexy, « Mart Susi's internet balancing formula » (2019) 25 *European Law Journal* 213-220

R Alexy, « Proportionality, constitutional law, and sub-constitutional law: A reply to Aharon Barak » (2018) 16:3 I•*CON* 871–879

R Alexy, « The construction of constitutional rights » (2012) 91:3 *Revue française de droit constitutionnel* 465-477

D Baron, « Private Ordering on the Internet: The Ebay Community of Traders » (2002) 4:3 *Business and Politics* 245-74

J Bell, « Judicial Review in the Administrative State » in J de Poorter, E Hirsch Ballin and S Lavrijssen (eds), *Judicial Review of Administrative Discretion in the Administrative State* (Springer 2019) 3-26

L Belli and C Sappa, « The Intermediary Conundrum: Cyber-Regulators, Cyber-Police or Both? » (2017) 8 *JIPITEC* 183

L Belli and J Venturini, « Private Ordering and the Rise of Terms of Service as Cyber-Regulation » (2016) 5:4 *Internet Policy Review* 1–17

A Benish and J Pelisse, « Private companies and administrative justice » in M Hertogh, R Kirkham, R Thomas and J Tomlinson (eds) *Oxford Handbook of Administrative Justice* (Oxford: Oxford University Press forthcoming 2021)

R Botsman, *Who can you trust?: how technology brought us together–and why it could drive us apart* (London: Penguin 2017)

T Büthe, « Private Regulation in the Global Economy: A (P)Review » (2010) 12:3 *Business and Politics* 1-38

B Cannon and H Chung, « A Framework for Designing Co-Regulation Models Well-Adapted to Technology-Facilitated Sharing Economies » (2015) 31 *Santa Clara Computer and High Technology LJ* 23-96;

E Carolan and A Spina, « Behavioural Sciences and EU Data Protection Law: Challenges and Opportunities » in A Alemanno and AL Sibony (eds), *Nudge and the Law* (Oxford: Hart Publishing 2015) 8

A Chander, « The Racist Algorithm?" (2017) 115:6 *Mich L Rev* 1023-45

M Cian, « Online Platforms as Gatekeepers to the Digital World – A Preliminary Issue on Business Freedom, Competition and the Need for a Special Market Regulation » (2018) *Journal of European Consumer and Market Law* 209-10

Conseil d'Etat français, 19 mai 1933, *Benjamin*

Couderc and Hachette Filipacchi associés v France [GC], 40454/07 (ECtHR 10 November 2015)

Cour de Cassation française, Civ. 1re, 11 July 2018, nr 17-22.381

C de Terwangne and Q Van Enis, *L'Europe des droits de l'homme à l'heure d'Internet* (Brussels: Bruylant 2019)

J Deigh, « Punishment and proportionality: Part 2 » (2016) 35:1 *Criminal Justice Ethics* 21-38

J Deigh, « Punishment and proportionality » (2014) 33:3 *Criminal Justice Ethics* 185-199

M Delmas-Marty, *Ordering Pluralism – A Conceptual Framework for Understanding the Transnational Legal World* (Oxford: Hart 2009)

R Drath, « Hotfile, Megaupload, and the Future of Copyright on the Internet: What Can Cyberlockers Tell Us About DMCA Reform » (2012) 12 J *Marshall Rev Intell Prop L* 205

D Dyzenhaus, « Proportionality and deference in a culture of justification » in G Huscroft, B Miller and G Weber (eds), *Proportionality and the rule of law – Rights, Justification, Reasoning* (Cambridge: Cambridge University Press 2014) 234-258

J Feinberg, *Doing and Deserving* (Princeton: Princeton University Press 1970) 102

M Finck, « Digital Co-Regulation: Designing a Supranational Legal Framework for the Platform Economy » (2018) 43 *European Law Review* 47-68

R Fisman, and L Michael, *Fixing Discrimination in Online Marketplaces*, HBR, no. December 2016

J Freeman, « Extending public law norms through privatization » (2003) 116:5 *Harvard Law Review* 1285-1352

C Geiger, and E Izyumenko, « The Role of Human Rights in Copyright Enforcement Online: Elaborating a Legal Framework for Website Blocking » (2016) 32:1 *Am. U. Int'l L. R.* 43

J Glaser, and K B Kahn, « Prejudice and Discrimination and the Internet » in Y Amichai-Hamburger (ed), *The Social Net. Human Behavior in Cyberspace* (Oxford: Oxford University Press 2005)

S Greer, « 'Balancing' and the European Court of Human Rights: A Contribution to the Habermas-Alexy Debate » (2004) 63:2 *Cambridge Law Journal* 412-434

D Harvey, *Collisions in the Digital Paradigm: Law and Rule Making in the Internet Age* (Oxford/Portland: Bloomsbury Publishing, 2017)

S Hick, E. Halpin and E. Hoskins, *Human Rights and the Internet* (Palgrave Macmillan UK 2016)

SL Kalėda, «The Role of the Principle of Effective Judicial Protection in Relation to Website Blocking Injunctions » (2017) 8 *JIPITEC* 216

E Katsh and O Rabinovich-Einy, *Digital Justice: Technology and the Internet of Disputes* (New York: Oxford University Press 2017)

B Kingsbury, N Krisch and R B Stewart, « The Emergence of Global Administrative Law » (2005) 68 *Law and Contemp Probs* 15

P Kjaer, « Why justification? The structure of public power in transnational contexts » (2017) (8:1) *Transnational Legal Theory* 8-21

M Klatt and M Meister, « Proportionality – A benefit to human rights – Remarks on the I-Con controversy » (2012) 10 *Int'l J Const L* 687-708

D Koukiadis, *Reconstituting Internet Normativity. The Role of State, Private Actors, Global Online Community in the Production of Legal Norms* (Baden-Baden: Nomos 2015)

K-H Ladeur, « A critique of balancing and the principle of proportionality in constitutional law – a case for 'impersonal rights'? » (2016) 7:2 *Transnational legal theory* 228-256.

L Lessig, *Code 2.0*. (Basic Books 2006)

E Marique and Y Marique, « Sanctions on digital platforms – beyond the public/private divide » (2019) 8:2 *Cambridge Journal of International Law* (forthcoming)

E Marique and Y Marique, « Beyond the Public and Private Divide on Digital Platforms? Revisiting Power Relationships » in E Bani, E Rutkowska-Tomaszewska and B Pachuca-Smulska (eds), *Public Law and the Challenges of New Technologies and Digital Markets* (CH Beck 2019 forthcoming)

E Marique and A Strowel, « Gouverner par la loi ou les algorithmes: de la norme générale de comportement au guidage rapproché des conduites » (2017) 10 *Dalloz IP/IT* 517-521

C Marsden, *Internet Co-Regulation: European Law, Regulatory Governance and Legitimacy in Cyberspace* (Cambridge: Cambridge University Press 2011)

D McKee, « The platform economy: natural, neutral, consensual and efficient? » (2017) 8:4 *Transnational Legal Theory* pp. 455-495

H Oberdiek, « The Role of Sanctions and Coercion in Understanding Law and Legal Systems » (1976) 21:1 *The American Journal of Jurisprudence* 71-94

D Oliver, *Common Values and the Public-Private Divide* (Cambridge: Cambridge University Press 1999)

Paris Trial court, *Ligue Contre le Racisme et l'Antisémitisme v Yahoo!*, 20 November 2000

Przemyslaw J Palka, « Terms of Service are not Contracts — Beyond Contract Law in the Regulation of Online Platforms » in S Grundmann (ed) *European Contract Law in the Digital Age*, vol 3 (European Contract Law in the Digital Age, Intersentia 2018) 136–61

C Petersen, V Ulfbeck and O Hansen, « Platforms as Private Governance Systems - the Example of Airbnb » (2018) *Nordic Journal of Commercial Law* 38-61

S Quack, « Expertise and authority in transnational governance » in R Cotterrell and M del Mar (eds), *Authority in Transnational Legal Theory – Theorising Across Disciplines* (Cheltenham: Edward Elgar 2016) 361-386

C Reed, *Making Laws for Cyberspace* (Oxford: Oxford University Press 2012)

Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, 6 April 2018, Human Rights Council, 38th session

Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, 16 May 2011, Human Rights Council, 17th session

J Riordan, *The Liability of Internet Intermediaries* (Oxford: Oxford University Press 2016)

C Rule, « Designing a Global Online Dispute Resolution System: Lessons Learned from eBay » (2017) 13:2 *U St Thomas LJ* 354–364

A Savin, *EU Internet Law* (Cheltenham: Edward Elgar 2013)

T Schulz, « Does Online Dispute Resolution Need Governmental Intervention? The Case for Architectures of Control and Trust » (2004) 6:1 *North Carolina Journal of Law & Technology* 71-106

M Siegel, « Hate Speech, Civil Rights, and the Internet: The Jurisdictional and Human Rights Nightmare Comment » (1998-99) 2 *Albany Law Journal of Science & Technology* 375-98

A Stone Sweet and J Mathews, « Proportionality balancing and global constitutionalism » (2008) 47 *Colum J Transnat'l L* 72-164

M Susi, « The Internet balancing formula » (2019) 25 *European Law Journal* 198-212

N Tusikov, *Chokepoints: Global Private Regulation on the Internet* (Univ of California Press 2016)

F Urbina, « Is it really that easy – A critique of proportionality and "balancing" as reasoning » (2014) 27:1 *Canadian Journal of Law and Jurisprudence* 167-192

M van de Kerchove, « Les fonctions de la sanction pénale. Entre droit et philosophie » (2005) 127:7 *Informations sociales* 22-31

R Van Loo, « Rise of the Digital Regulator » (2017) *Duke Law Journal* 1317

J Venturini, L Louzada, M Maciel, N Zingales, K Stylianou, and L Belli, *Terms of Service and Human Rights: An Analysis of Online Platform Contracts* (Editora Revan, Rio de Janeiro 2016)

A von Hirsch, « Proportionality in the philosophy of punishment » (1992) 16 *Crime and Justice* 55-98

G Vilaca, « Transnational legal normativity » in M Sellers and S Kirste (eds), *Encyclopedia of the philosophy of law and social policy* (Springer 2017)

J Waldron, « The rule of law and the importance of procedure » (2011) 50 *Nomos*, Getting to the rule of law, 3-31

C Wattecamps, A-G Kleczewski and E Marique, « Challenges related to law for the platform economy: A fresh look at some important dichotomies » (2017/3) *Reflets et perspectives de la vie économique* 57-95

Yahoo! Inc., v La Ligue Contre Le Racisme et l'Antisémitisme, 379 F 3d 1120 (9th Ct, 23 August 2004)

F Zuiderveen Borgesius, "Behavioural Sciences and the Regulation of Privacy on the Internet" in A Alemanno and AL Sibony (eds), *Nudge and the Law* (Oxford: Hart Publishing 2015)

Author Biography

Enguerrand Marique is a PhD candidate at the CRIDES (Center for business law) at the UCLouvain School of Law. His doctoral research focuses on the legal tools to build trust in the digital platform economy. His project analyses the legal protections in the management of identification and e-reputation and big data under both private and public law scholarship. Enguerrand was a guest editor for a special issue of the *Revue Internationale de Droit Economique* on the regulation of digital platforms (publication: September 2019). His main interests pertain to taxation, intellectual property as well as the regulation of algorithms, including in relation with robots. He holds law degrees from the UCLouvain and the University of California at Davis.

Yseult Marique is Senior Lecturer at the University of Essex (UK) and research associate at the FÖV Speyer (Germany). She is the author of *Public private partnerships and the law* (Edward Elgar 2014) and the co-editor of *Access to Justice: Beyond the Policies and Politics of Austerity* (with E Palmer, T Cornford, and A Guinchard) (Bloomsbury 2016). Her main research interests focus on the rule of law and its practical implementation in relation to issues of access to limited resources in Western societies (e.g. access to justice, to energy, to public services, to administrative documents, to public participation etc.) and more generally on administrative enforcement and ethics of care in pluralistic Western societies both today and in the past. She holds a law degree from the Université libre de Bruxelles (ULB) and from the Vrije Universiteit Brussels (VUB) and a PhD from Cambridge University.

Acknowledgements

This paper builds on comments and discussions following a presentation made at the 8th Annual Conference of International Law, Cambridge (20–21 March 2019). We would like to thank for their generous comments offered by Professor Benedict Kingsbury, Thomas Streinz, and the two anonymous referees of this special number for their detailed comments, which have opened new avenues for reflection in our work.

A New Framework for Online Content Moderation

Ivar A. Hartmann[a, 266*]

[a] *Center for Technology & Society (CTS) at FGV Law School, Rio de Janeiro, RJ, Brazil.*

Abstract

I wish to provide a description of context and practical changes in the institutions, places and tools of speech moderation before and after the internet. This description revolves around legally and institutionally relevant aspects of how excesses in speech were identified and countered in order to provide support for a normative claim about what online speech moderation should look like today and in the future. The article starts with a list of elements of content moderation up until three decades ago and the follow with a list of elements of content moderation today. The primary goal is to contrast the two scenarios in order to highlight the inconvenience of certain assumptions lawmakers, lawyers and judges make on how communication works in a networked society. I do not intend to provide alternative descriptions to the characteristics of this phenomenon in order to dispute prevailing descriptions. My point is merely to uncover certain aspect that usually remain unnoticed or underestimated in the legal debate about content moderation. The third part of the article will then propose the outline of a new procedural legal framework for moderation of online speech without dwelling too deep on considerations of substantive legal standards for balancing speech.

1. Introduction

The goal of balancing free speech and countervailing interests or rights[267] [268] [269] such as privacy, copyright and honor on the Internet is widely considered to be as complex as it is central to a well-functioning global internet. It is naturally an international objective as much as a national issue, but in this article I wish to make a contribution that is focused on the national context and legal possibilities - even if conceptually my descriptive and normative propositions could be applied to international law in many respects.

It is certainly necessary to discuss legal standards for reviewing and moderating speech, in the sense of substantive rules about what can and cannot be said online. Many landmark studies have been dedicated to this purpose and my goal here is not to provide any advancements in this front. Rather, I wish to provide an overview of context and practical changes in the institutions, places and tools of speech moderation before and after the internet. This analysis will revolve around legally and institutionally relevant aspects of how excesses in speech were identified and countered in order to provide support for a normative claim about what online speech moderation should look like today and in the future.

I will therefore start with a list of elements of content moderation up until three decades ago and then follow with a list of elements of content moderation today. The primary goal is to contrast the

[266] Praia de Botafogo, 190 - 13° andar, Rio de Janeiro – RJ, Brazil. E-mail address: <ivar.hartmann@fgv.br>.

[267] Eduardo Andrés Bertoni, 'The Inter-American Court of Human Rights and the European Court of Human Rights: a dialogue on freedom of expression standards' (2009), 03 European Human Rights Law Review.

[268] Jean-François Flauss, 'The European Court of Human Rights and the Freedom of Expression' (2009), 84 (03) Indiana Law Journal.

[269] Malcolm D Evans, 'From Cartoons to Crucifixes: Current Controversies Concerning The Freedom of Religion and The Freedom of Expression Before The European Court of Human Rights' (2010), 26 Journal of Law & Religion.

two scenarios in order to highlight the inconvenience of certain assumptions lawmakers, lawyers and judges make on how communication works in a networked society[270]. I do not intend to provide alternative descriptions to the characteristics of this phenomenon in order to dispute prevailing descriptions. My point is merely to uncover certain aspect that usually remain unnoticed or underestimated in the legal debate about content moderation.

The third part of the article will then propose the outline of a new procedural legal framework for moderation of online speech without dwelling too deep on considerations of substantive legal standards for balancing speech. That is to say: I do not wish to offer an alternative for balancing[271] [272] [273] [274] as a method of resolving clashes between speech and other constitutional rights, including personality rights[275], nor do I hope to provide new elements or criteria for balancing expression[276] [277] [278]. On the contrary, the third part of this article takes the current balancing literature and practice for what it is and only suggests changes in terms of the *who* and *when*.

I therefore intend to offer a proposal about *roles* of different stakeholders and not about what the legal standard for hate speech or defamation is or should be in a specific country. This means I am not concerned with the specific legal standards of one or a group of countries and would rather suggest changes for the roles of the Judiciary, the Executive and Legislative branches of government, as well as for private social media platforms and users.

2. Content Moderation in the Age of Mass Media

Content moderation is a complex issue that deserves to be considered from various angles. The first aspect of content moderation in societies where speech is manifested primarily through mass media is **(a)** the concentration of the decision to disseminate content. As I stated previously, the goal here is not to offer surprisingly new accounts of what communication was like three decades ago or is now. Rather, it is to highlight certain elements that are commonly unappreciated by legal operators and in doing so to try painting a picture that is not necessarily more accurate, but more adequately nuanced and therefore offers a sharper support for normative claims on content moderation and the desired roles of the institutions involved.

The fact that very few players were responsible for almost all judgement calls on what speech was seen, read or heard is one of these commonly unappreciated elements of the type of society in which the theory of constitutional rights balancing was developed. I do not mean by this that the framers of the theory did not think carefully about these elements, rather that those who employ balancing today do not pay enough attention to them.

Mass media means one-to-many because although the people providing opinions and personal accounts were numerous, they only took up the scarce space in mass media as a result of the decisions of a few. Regular people never had in their lifetime an opportunity to decide whether or

[270] Manuel Castells, 'Infomationalism, Networks, and the Network Society: A Theoretical Blueprint', *The network society: a cross-cultural perspective* (Edward Elgar, 2004).
[271] Robert Alexy, *Teoria de los Derechos Fundamentales* (Centro de Estudios Constitucionales, 1993).
[272] José Joaquim Gomes Canotilho, *Direito constitucional e teoria da constituição* (7th edition, Almdedina, 2003).
[273] Ingo Wolfgang Sarlet, *A Eficácia dos Direitos Fundamentais na Constituição de 1988. Uma Teoria Geral dos Direitos Fundamentais na Perspectiva Constitucional* (11th edition, Livraria do Advogado, 2012).
[274] Alec Stone Sweet, Jud Mathews, 'Proportionality balancing and Global Constitutionalism' (2008), 47 Columbia Journal of Transnational Law.
[275] Fábio Siebeneichler de Andrade, 'A tutela dos direitos da personalidade no direito brasileiro em perspectiva atual' (2013), 24 Revista de Derecho Privado.
[276]Fábio Carvalho Leite, 'Liberdade de Expressão e Direito à honra: novas diretrizes para um velho problema', Clèmerson Merlin Clève; Alexandre Freire, D*ireitos fundamentais e jurisdição constitucional: análise, crítica e contribuições* (1st edition, Revista dos Tribunais, 2014).
[277] Leonardo Martins, 'Direitos Fundamentais à intimidade, à vida privada, à honra e à imagem (Art. 5o, X da CF): Alcance e substrato fático da norma constitucional (intervenção estatal potencialmente violadora)' (2016), 07 (01) Ius Gentium. Curitiba.
[278] Daniel Sarmento, *A ponderação de interesses na Constituição Federal* (Lumen Juris, 2003).

not to say something to tens or hundreds of thousands of other people. The average person never had a choice of manifesting a preference for a political party, a position in a relevant social debate or even of sharing creative works to an audience of potentially millions. That power was concentrated in the hands of an extremely small number of individuals.

A second, resulting aspect, is **(b)** the severe imbalance in the power to disseminate information. While many had no such power or had at most the ability to have their opinions made known to the low dozens of people who were physically or geographically close to them, others owned private means of mass communication or managed public avenues of mass media[279]. This is not merely about private media ownership concentration[280] and its problems, it is also about people in office who, through licensing and other kinds of decisions, could sway one way or another the content of messages widely transmitted. And it is about public figures who knew and exercised an unwritten prerogative of being broadcast by media whenever they decided to say something to the public. It is about private people who became celebrities, and as public figures were able to voice concerns, preferences or criticism through a mass media that was eager to give more and more space and airtime to those who were famous. Whether mass media is private or state-owned, it makes no difference for this latter aspect of the imbalance.

This imbalance in the power to disseminate information, in the ability to reach a wide audience, is a key characteristic of the archetypal case of the private person who has their reputation irremediably ruined by a newspaper running a poorly-fact-checked story. In this scenario, judicial moderation of speech in order to impose not only content removal but also compensation for immaterial damages to the private person's honour and image was a vital element because there were no alternatives for mediation or for calling the newspaper on its mistake. If not for courts, there were no alternatives to punish the speech abuse of those few who concentrated the power to disseminate information. Accountability necessarily meant judicial review. Libel laws find an authoritarian and aristocratic foundation in the medieval culture of the show of respect[281], but in modern liberal democracies[282] their best possible justification is the pervasive imbalance in the capacity to reach and audience.

Another characteristic of speech before the Internet is that **(c1)** the places available for unimpeded, autonomous political or artistic manifestations were public spaces. Because such spaces played a pivotal role in enabling freedom of expression, constitutional law in different countries took into account the need to ensure certain limits to state censorship of speech in public spaces. The public forum doctrine developed by the United States Supreme Court is one example of this. It was almost natural that public authorities should yield some of their control over public spaces in order to allow for spontaneous speech and expression that did not have to wait for prior consent.

In parallel, **(c2)** any private space or media with a large audience necessarily entailed evaluation and authorization before speech could be relayed forward or published. For those who controlled the media, third party speech in their waves, tubes and pages was always opt-in. It always entailed a case-by-case decision to proactively pass forward that message.

Furthermore, **(d)** the number of choices made specifically about content by those who decided on or influenced the speech that was widely disseminated was small. A newspaper editor only made a handful of decisions each day, including on what columns or letters from readers to publish. The managers of a television station made some decisions on content by proxy in

[279] Owen M Fiss, 'Free Speech and Social Structure' (1986), 71 Iowa Law Review.
[280] Ellen P Goodman, 'Media Policy Out of the Box: Content Abundance, Attention Scarcity, and the Failures of Digital Markets' (2004), 19 Berkeley Technology Law Journal.
[281] James Q Whitman, 'Enforcing civility and respect: three societies' (2000), 109 Yale Law Journal.
[282] John Raws, *Political Liberalism* (Expanded edition, Columbia University Press, 2005).

selecting a few reporters, show runners, presenters and so on to take up the scarce airtime. Alternatively, authorities responsible for public fora did not have a say on what content speech was allowed and therefore made no decisions at all. They were not editors.

In a world where the number of decisions about content is small, it is viable for courts to exercise review based on the content of speech in order to classify each instance as legal or illegal, lawful or libel, acceptable or defamation. In such a scenario, where (b) and (d) are true, the Judiciary not only *de facto* can perform content-based review of each case, it must do so as there are no other accountability alternatives.

Lastly, **(e)** the tools available to enforce such review are prior restraint, removal, compensation and right of reply. Because any of these tools are forced upon the media controllers by another institution, there is little room for nuance or flexibility. Compensation means a large spectrum of different possible amounts of money, but it does not affect the reviewed speech itself – and neither does the right of reply. Prior restraint is a yes or no question with at most some wiggle room on the starting and ending date of the censorship. Removal is the same. The only two mechanisms for review that interfere on the expression at hand are binary.

In addition, the few people making decisions on what content to broadly disseminate have no technical capacity to fine tune the profile of their audience. They can choose to sell their newspapers at newsstand a or b, but they cannot keep track of who buys it, when and for what purpose. Once the newspaper is sold, they have no information on who read it, for how long or on which parts of the paper each person gave more attention to. A radio station can choose to broadcast in region a or b, but once there it has no idea of who is listening or for how long. And it certainly cannot prevent certain demographics in region a to be able to tune in while authorizing others.

Therefore, content moderation traditionally means individual decisions about the merits of a particular instance of speech because the speaker exercised a limited number of binary decisions to disseminate or not to disseminate on the merits of that speech. This is further simplified by the fact that the decisions were to publish or not and the counteractions by courts are to censor or not. Those are the limitations of what content moderation could be in a modern democracy and free speech balancing at the behest of courts was designed to fit those limitations.

3. What Content Moderation can be

In this section I will describe the changes to elements (a) through (e) presented above as we inch towards a networked society.

There is far less concentration in the power to disseminate information. Whereas before there was a very small number of people capable of making judgement calls on the circulation of specific instances of information that would reach a wide audience, now billions of people make decisions on what to share, forward or give visibility to on a daily basis. There is worrying market concentration of digital platforms, such that three people with voting stock at Google and Facebook make decisions about the architecture of speech media that certainly influences public debate all over the world[283]. This does not change the fact, however, that much more people make individual decisions about the dissemination of content today than those who made editorial choices in the mass media past. Page space and airtime are no longer constraints on how much information can widely circulate. There are multiple avenues online, from a cheap web page or blog to a social media profile and or email account. Indeed, the very many-to-many logic of the

[283] Luigi Zingales, Filippo Maria Lancieri, 'Committee on Digital Platforms: Policy Brief. Chiago Booth' (2019), Stigler Center for the Study of the Economics and the State.

Internet as a decentralized network guarantees anyone connected has the ability to reach and audience verbally or in writing. This has been widely interpreted as an improvement over traditional mass media limitations, although some authors would describe this development negatively[284]. Whatever criticism might be directed at what Umberto Eco described as a "legion of idiots"[285], the most comprehensive empirical study on trolls concludes they only repeat the discourse vices of mainstream media[286]. That is to say: the sophistication of speech today might be low, especially among groups that perpetuate harmful stereotypes and hurtful messages. But this does not change the fact that traditional media always had similar flaws. Therefore, if problems with the quality of speech arise as a collateral damage of the Internet, we should not hurry to blame it on the fact that access to media is more democratic.

This means a change in **(a)** (the concentration of the decision to disseminate content), but does not in itself guarantee a change in (b) (the severe imbalance in the power to disseminate information). There is still scarcity of attention[287], which maintains the precondition imbalance in the ability to reach an audience. Because the concentration of the decision to disseminate content changed, technically communication in society could flow in a perfectly uniform fashion. My point is not that it should or that it currently effectively does, but rather that it could from a technological viability perspective.

The scarcity of attention however sustains the incentive for some to employ resources to have more communicative capacity[288] than others. Communicative capacity can be understood as "a social and situated ability that emerges and exists in the interactions between people while being engaged in the process (…) of participatory practice."[289]. Some public policy and private enterprise choices throughout the development of the Internet have to some extent capped the possibility of power in the financial sphere to translate into power in the public opinion and information sphere, which advances a progressive conception of justice[290]. For example, the number of visits to a website or how far away the readers are does not affect the host's price to keep the website up. Domain name registration costs an objectively low amount of money. Social media rarely charges users for joining and posting content, rather relying on other sources of revenue that come with perverse incentives. This means that at least the barrier to entry in terms of speech is extremely low. However, **(b)** (the severe imbalance in the power to disseminate information) has not completely altered and there were early warnings that market choices could indeed reverse the initial trend of decoupling money from media power[291].

There is still imbalance[292], but what is important to note is that it has decreased substantially[293]. Many relationships that previously showed fundamental imbalance in communicative capacity, such as the case of the newspaper wrecking the arduously cultivated reputation of a private person in one fell news piece, now allow for different outcomes. That person could post a complaint on social media that goes viral and turns the newspaper's audience wildly against it in

[284] Brian Leiter, 'Cleaning Cyber-cesspools: Google and Free Speech', Marth Nussbaum, Saul Levmore, *The Offensive Internet: Speech, Privacy, and Reputation* (Harvard University Press, 2010).
[285] Umberto Eco La Stampa 'Con i social parola a legioni di imbecilli', *La Stampa* (10 june 2015) <https://www.lastampa.it/2015/06/10/cultura/eco-con-i-parola-a-legioni-di-imbecilli-XJrvezBN4XOoyo0h98EfiJ/pagina.html> acessed 21 jan. 2018.
[286] Whitney Phillips, *This Is Why We Can't Have Nice Things: Mapping the Relationship between Online Trolling and Mainstream Culture* (MIT Press, 2016).
[287] Zeynep Tufekci, 'Algorithmic Harms Beyond Facebook and Google: Emergent Challenges of Computational Agency' (2015), 13 Colorado Technology Law Review.
[288] Vesna Bagarić, Jelena Mihaljević Djigunović, 'Defining Communicative Competence' (2007), 08 (14) Metodika.
[289] Koen Bartels, 'Communicative Capacity: The Added Value of Public Encounters for Participatory Democracy' (2014), 44 (06) American Review of Public Administration.
[290] Michael Walzer, *Spheres of Justice. A defense of pluralism and equality* (Basic Books, 1984).
[291] Howard Rheingold, *The virtual community: homesteading on the electronic frontier* (The MIT Press, 2000).
[292] Matthew Hindman, 'What is the Online Public Sphere Good For?', Joe Turow, Lokman Tsui, *The Hyperlinked Society* (University of Michigan Press, 2008).
[293] Yochai Benkler, The Wealth Of Networks: How Social Production Transforms Markets and Freedom (Yale University Press, 2006).

a couple of days. Traditional media outlets are in fact facing direct criticism from civil society and politicians enabled by social media and unfortunately have their work - even serious investigative journalism - many times discredited by swarms of regular people, a phenomenon that is key to understanding post-truth[294].

Public figures such as celebrities and politicians, who historically sued media outlets for defamation now have big followings on social media and become as influential as those media outlets, if not more. More importantly, the path to being considered a public figure with high communicative capacity is different and less selective than it was before. Many times it does not follow the same logic or criteria. In order to adapt public figure doctrine in a given defamation case, a court should consider, for instance, whether "a social media user plaintiff has greater access to the media than other users on the plaintiffs social media network"[295].

The average person in a networked society can reach a vastly larger audience than the average person in the age of mass media, although the ways in which such reach will happen in practice are increasingly intermediated by algorithmic and human moderation within major digital platforms. Whether public figure or not, if the defendant (speaker) in a defamation lawsuit does not have higher communicative capacity than the plaintiff (offended party), the ruling on the merits should not even go into a rights balancing phase and instead recognize that there is no possibility of actionable damage[296]. Speech regulation that aims at content and fails to take into account relationships of power is intrinsically inefficient[297], as opposed to the misguided view that government censorship is the only constitutionally-relevant impediment to free speech[298].

It is precisely because more people have more communicative capacity and there is less imbalance that the Judiciary is no longer the only alternative for accountability of speech. While before racist speech could only really be dissuaded by courts, a much more inclusive and dynamic court of public opinion now renders a trial on a significantly larger number of cases and deals a quicker and possibly harsher punishment[299].

The changes to **(c)** are also decisive. Not only are the main fora for communication no longer public **(c1)**, raising several problems for the lack of traditionally open and free public spaces for expressions[300], their private nature does not necessarily mean all speech is subject to prior approval **(c2)**. The Internet is said to be private in the sense that the infrastructure layer and the logical layer (benkler) are owned by private companies. A web page is hosted in a private server. A social media profile only exists inside a private social media network.

To GoDaddy, Facebook or YouTube, third-party speech is opt-out, not opt-in. Although content platforms exercise control, they were not built with individual prior restraint in mind[301] - on the contrary, their business model is frequently based in the monetization of user-produced content or user personal data. Slowly but surely, however, they are being forced to change the law of the code under pressure from traditional lawmakers[302]. They can do one of two things in order to opt-

[294] Silvio Waisbord, 'Truth is What Happens to News. On journalism, fake news, and post-truth' (2018), 19 Journalism Studies.
[295] Matthew, Lafferman, 'Do Facebook and Twitter Make You a Public Figure: How to Apply the Gertz Public Figure Doctrine to Social Media' (2012), 39 Santa Clara Computer & High Tech.
[296] Ivar A Hartmann, 'Liberdade de Expressão e Capacidade Comunicativa. Um novo critério para resolver conflitos entre direitos fundamentais informacionais' (2018), 12 (39) Revista Brasileira de Direitos Fundamentais e Justiça.
[297] Kathleen Sullivan, 'Free speech and unfree markets' (1994), 42 UCLA Law Review.
[298] Ronald A Cass, Melville B Nimmer, 'The Perils of Positive Thinking: Constitutional Interpretation and Negative First Amendment Theory' (1987), 34 UCLA Law Review.
[299] Jon Ronson, *So you've been publicly shamed* (Picador, 2015).
[300] Molly Sauter, *The Coming Swarm: DDOS Actions, Hacktivism, and Civil Disobedience on the Internet* (Bloomsbury Academic, 2014).
[301] Geoffrey Parker, Marshall W. Van Alstyne, Sangeet Paul Choudary, *Platform Revolution: How Networked Markets Are Transforming the Economyand How to Make Them Work for You* (W. W. Norton & Company, 2016).

[302] Lawrence Lessig, *Code. Version 2.0* (Basic Books, 2006).

out of specific instances of third-party speech. The first possibility is to perform manual case-by-case content moderation based on the merits of each instance of speech. If they do so, however, that would mean a deluge of choices on content moderation that platforms would have to handle each second. It is the total opposite of **(d)** (when the number of choices made specifically about content by those who decided on or influenced speech that was widely disseminated was small). Early on after the popularization of online speech it was already clear that protecting speech was decreasingly about rulings on individual instances of speech and increasingly about decisions on architecture of information systems[303].

Content platforms are not equipped to perform millions of content evaluations everyday and arguably they should not even be allowed to play this part. At best, they do not take seriously the task of manually reviewing expression and make sloppy decisions always erring on the side of censorship. The documented frequency of false positives in notice and take down systems impose monumental chilling effects[304] on speech online[305] [306]. At worst, these private companies fully engage in a role for which they utterly lack legitimacy. Case-by-case speech review by content platforms in a context where almost all news, opinion and creative artwork flows through them raises a significant private censorship concern[307].

If the tools of review today were the same as in the age of mass media, we might be left with no alternative to this first option that private content platforms can dispose of in order to opt-out of specific instances of third-party speech. Fortunately, the content moderation tools available today are not all blunt and binary as before - **(e)** (the tools available to enforce such review are prior restraint, removal, compensation and right of reply). There are several new mechanisms that allow for extremely sophisticated fine tuning in terms of timing, audience profile and content.

A post that is deemed prejudicial can be red-flagged without being removed[308]. This means it can be marked with a sign of disapproval without any direct action by the platform to reduce its dissemination or visibility. After this, all viewers will necessarily be made aware of this branding. There was nothing remotely close to a tool like red-flagging in mass-media content moderation. A paper could be forced to print a right of reply, but it would not come in the same edition as the original news story.

A red flag might be used as a first step taken by the platform that opens up a procedure whereby further red-flags on the same posts by viewers will trigger temporary or permanent removal. Temporary removal was also unavailable in as much as suspending a ban on an edition of a magazine would not result in the company republishing the exact same issue. Suspending a ban on a television news story would also not cause the station to run the exact same story to the same or other audiences.

Another important tool that the framers of balancing rights theory could not have imagined is modulation of views. Without removing a picture, a social media company can make it appear much *less* often to users with profile “a” while allowing it to continue just as visible as before to users with profile “b”. Modulation of views can be combined with red-flagging such that a picture flagged as type x will appear much *more* often to certain specific demographics. These complex

[303] Jack Balkin, ‘Digital speech and democratic culture: a theory of freedom of expression for the information society’ (2004), 79 (01) New York University Law Review.
[304] Frederick Schauer, ‘Fear, Risk and the First Amendment: Unraveling the Chilling Effect’ (1978), 58 Boston University Law Review.
[305] Wendy Seltzer, ‘Free Speech Unmoored in Copyright’s Safe Harbor: Chilling Effects of The DMCA on The First Amendment’ (2010), 24 Harvard Journal of Law and Technology.
[306] Gerald Spindler, ‘Internet Intermediary Liability Reloaded The New German Act on Responsibility of Social Networks and its (In-) Compatibility with European Law’ (2017), 8 JIPITEC.
[307] Laura Denardis, ‘Hidden Levers of Internet Control’ (2012), 15 Information, Communication & Society.
[308] Kate Crawford, Tarleton Gillespie, ‘What is a flag for? Social media reporting tools and the vocabulary of complaint’ (2016), 18 (03) New media & Society.

mechanisms are especially useful for moderating fake news, where adding new layers of information to suspicious posts is a better first step than outright removing or simply red-flagging[309]. Fake news is not a phenomenon of mass media but of networked communication: "fabricated information that mimics news media content in form but not in organizational process or intent"[310] and attempts to moderate it with the old system are futile. This new layer of information can be something apparently as simple as the number of times that specific post has been forwarded[311].

The very trigger of moderation mechanisms is evolved in the sense that it does not have to be the result of a decision by the social media company, but rather by its users. Wikipedia is an example of an intricate ecosystem of crowd-sourced moderation[312] [313] where different hierarchies of editors exercise what can be called decentralized gatekeeping[314] [315].

4. A New Framework for Online Content Moderation

Profound changes in the conditions of information flow as well as the tools for its moderation require a new framework of regulation. I will lay out its characteristics by describing the changes in the roles of each stakeholder. The plurality of the types of stakeholders makes this model resemble the multistakeholder model, a paradigm that online regulation scholarship has been developing for a long time[316] [317] [318]. Describing in detail a new framework for online content moderation in one section of a short paper is naturally an impossible task and one that I do not aim to accomplish. The purpose of this paper is to describe very basic underpinnings of such a model, as I have already started doing in the previous section and will continue here. The reason such overview is even possible is that I have developed some elements of this new framework in other works and many of the pieces of the puzzle have already been placed by other scholarly work. My goal is not to create a framework with entirely novel elements, it is to make a few adjustment suggestions and tie up loose strings.

4.1 Courts

An important premise is that courts have the starring role in rights balancing speech review systems[319], among other reasons, because balancing means general rules of when speech should prevail are normally out of question and answers can only be reached after examination

[309] Hanna Kozlowska, 'Facebook is ditching its own solution to fake news because it didn't work', *Quartz* (22 December 2017, <https://qz.com/1162973/to-fight-fake-news-facebook-is-replacing-flagging-posts-as-disputed-with-related-articles/> accessed 21 jan. 2018).
[310] David M. J Lazer et all, 'The science of fake news' (2018), 359 (6380) Science.
[311] Marcia Sekhose, 'WhatsApp's new feature tells you how many times your message has been forwarded', *Hindustan Times* (22 March 2019, <https://www.hindustantimes.com/tech/whatsapp-s-new-feature-tells-you-how-many-times-your-message-has-been-forwarded/story-12S1flWBiomgF9jTazSTeM.html> accessed 21 jan. 2018).
[312] Dariusz Jemielniak, *Common Knowledge?: An Ethnography of Wikipedia* (Stanford University Press, 2014).
[313] Aniket Kittur, Robert E. Kraut, 'Harnessing the Wisdom of Crowds in Wikipedia: Quality Through Coordination' (2008), *CSCW'08* (San Diego, California, USA, 08-12 november, 2008).
[314] Karine Barzilai-nahon, 'Toward a Theory of Network Gatekeeping: A Framework for Exploring Information Control' (2008), 59 Journal of The American Society For Information Science and Technology.
[315] Aaron Shaw, 'Centralized and Decentralized Gatekeeping in an Open Online Collective' (2012), 40 Politics & Society.
[316] Wolfgang Kleinwächter, 'Internet co-governance. Towards a multilayer multiplayer mechanism of consultation, coordination and cooperation (M3C3)', Robin Mansell, *The information society v. III (Democracy, governance and regulation)*, (Routledge, 2009).
[317] Milton Mueller et al., 'The Internet and Global Governance: Principles and Norms for a New Regime' (2007), 13 Global Governance.
[318] Luca Belli, 'A heterostakeholder cooperation for sustainable internet policymaking' (2015), 04 Internet Policy Review.
[319] Ilton Robl Filho, Ingo W. Sarlet, 'Estado Democrático de Direito e os Limites da Liberdade de Expressão na Constituição Federal de 1988, com Destaque para o Problema da dsua Colisão com outros Direitos Fundamentais, em Especial, com os Direitos de Personalidade' (2016), 08 (14) Constituição, Economia e Desenvolvimento: Revista da Academia Brasileira de Direito Constitucional.

of the characteristics of each specific case[320] [321] - even if constitutional courts that exercise balancing can be described as erring on the side of speech more often[322] [323] [324]. Very little in terms of guidelines can be or is provided by law and it is up to judges to exercise merit-based review on each individual instance of expression. Furthermore, courts were the only possible institutional deterrent to defamation, hate speech[325] or even bullying. There were no other workable mechanisms for evaluating and pushing back on widely broadcast expression that society deemed negative.

In a networked society, courts are no longer the only institution capable of socially punishing speech, as private content platforms establish and enforce content rules[326] while users themselves criticize, red-flag, shun and boycott speech or its authors providing a quicker and usually more effective response than judges ever could for most cases. The perils are now also of an excess of user push back on content, with doxing campaigns and online bullying sometimes relentlessly targeting users with disproportionate punishment.

More importantly, the number of single decisions on expression that courts would have to issue in order to remain the sole or main reviewers of expression is simply unthinkable. And the portion of cases that they do find a way to decide puts them largely in a position of enablers of private censorship. Defamation lawsuits have 40%[327] and 20%[328] success rates in the United States and China, respectively. In Brazil, for example, the Google Transparency Report shows that Judicial removal requests are more numerous than Executive Branch removal requests in almost every year in the last decade. In the last report from June, 2018, judicial requests made up 63% of the total, against 43% from the Administration. In that period, defamation was the major cause for removal requests, with 46% of the total. The situation is similar in Germany, where court orders predominate over Executive branch removal requests almost every year since 2010. In the first half of 2018, defamation-based court removal requests were 57% of the total, while requests with grounds on privacy and security were a distant second at 22%[329]. As the imbalance in communicative capacity decreases, there are gradually less reasons for courts to arbitrate defamation disputes between private people.

This article argues that it is time judges move away from reviewing single instances of speech on the merits of content and turn to evaluating the procedural elements of content moderation systems created and managed by platforms. If a plaintiff sues the social media platform alleging they sustained immaterial damages due to a post by another person, as is possible in most countries, it is not sustainable for the judge to insist on the role of deciding whether that specific post caused harm and reply with an order of removal and financial compensation for damages. Rather, they should evaluate to what extent the private platform provided means for the plaintiff

[320] Ana Paula de Barcellos, 'Intimidade e Pessoas Notórias. Liberdades de Expressão e de Informação e Biografias. Conflito entre Direitos Fundamentais. Ponderação, Caso Concreto e Acesso à Justiça. Tutelas Específica e Indenizatória' (2014), 05 Revista Direito Público.
[321] Luís Roberto Barroso, 'Colisão entre liberdade de expressão e direitos da personalidade. Critérios de ponderação. Interpretação constitucionalmente adequada do Código Civil e da Lei de Imprensa' (2004), 235 Revista de Direito Administrativo.
[322] Klaus Stern, *Das Staatsrecht Der Bundesrepublik Deutschland. Band IV/1. Die einzelnen Grundrechte* (C.H. Beck, 2006).
[323] Christian Starck, *Kommentar zum Grundgesetz. Band I. Band 1, Präambel, Artikel 1 bis 19* (Franz Vahlen GmbH, 2005).
[324] Josef Isensee, Paul Kirchhof, *Handbuch des Staatsrechts. Band IV – Freiheitsrechte* (C.F. Müller Juristischer Verlag, 1989).
[325] Danielle Keats Citron, *Hate crimes in cyberspace*, (Harvard University Press, 2014).
[326] Ian Brown, Christopher T Marsden, Regulating Code. Good governance and better regulation in the information age (MIT Press, 2013).
[327] David Unwin, 'Defamation Litigation Patterns Across the United States, England, and Australia' (2013, <https://works.bepress.com/david_unwin/1/download/> accessed 21 jan 2018).
[328] Xin He, Fen Lin, 'The Losing Media? An Empirical Study of Defamation Litigation in China' (2017), 230 The China Quarterly.
[329] Google Inc. 'Government requests to remove content' (2018, <https://transparencyreport.google.com/government-removals/overview?hl=en> accessed 21 jan. 2018).

to reply online, to submit a complaint against the original post, one that is actually taken into account. Courts should check if the content moderation scheme set in place respects minimum due process rules, especially in legal systems where such constitutional rights bind private parties and not only the State.

The idea that public virtual decision making systems - fully or partly automated - need to respect basic technological due process has been pioneered at least a decade ago by Danielle Citron[330]. Virginia Eubanks[331], for example, has performed extensive qualitative research documenting the profound damaging effects of local government automated decision-making that disregards due process guarantees in the access to welfare and housing. In the context of private social media platforms, Rebecca Tushnet criticized the role ascribed to these companies by safe harbor liability standards that resulted in a misalignment of incentives towards legally protected private censorship. A more desirable alternative would be "some type of procedural due process, democratic self-governance, or nondiscrimination rule (...) [which should include] at a minimum, alternatives for empowering users of major ISPs substantively and procedurally (...)." [332].

This is the role that courts must turn to: enforcing procedural rules for content moderation and reviewing the architecture and basic rules of information flows in such platforms. As such, courts would be the vector of implementation of a system of regulated self-regulation – where, ideally, legislators will provide guidelines of procedure for content self-regulation by platforms, allowing judges to verify whether these basic rules were followed and, if they were, abstaining from intervention by finding against the plaintiff. David Kaye proposes elements of this new system which would include, from a procedural perspective, decentralized decision-making and radically better transparency - both rulemaking and decisional transparency[333].

4.2 Users

Who then should exercise review of each single instance of speech on its merits? This paper argues that this is the new role of users. For the first time in history, many societies are close enough to balance of communicative capacity in order to permit the realization of the logic of the free marketplace of ideas. That is still not to say that every country is already fully there, only that the scales have moved significantly in terms of the equilibrium of the ability to reach an audience. Posts with racial hate speech, for example, cannot be left for judges to punish, as courts show bias when deciding about speech[334] and are themselves much less representative of social diversity than social media due to, among other things a lack of gender inclusion in higher courts[335] [336]. Furthermore, the example of Brazil, a country that chose the path of criminalizing racism, shows that judges simply play down racist speech as humor[337] and rarely every punish it[338]. Private platforms, through teams of outsourced employees who spend hours looking at specific posts, pictures or videos, lack legitimacy to exercise this role.

[330] Danielle Keats Citron, 'Technological Due Process' (2008), 85 WASH. U. L. REV.
[331] Virginia Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor* (St. Martins Press, 2018).
[332] Rebecca Tushnet, 'Power Without Responsibility: Intermediaries and the First Amendment' (2008), 76 The George Washington Law Review.
[333] David Kaye, 'Speech Police: The Global Struggle to Govern the Internet', (2019), Columbia Global Reports.
[334] Lee Epstein, Christopher M Parker, Jeffrey A. Segal, 'Do Justices Defend the Speech They Hate? In-Group Bias, Opportunism, and the First Amendment' (2013), American Political Science Association Annual Meeting <http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2300572> accessed 05 may 2015.
[335] David A Cotter. et al., 'The glass ceiling effect' (2001), 80 (02) Social Forces.
[336] Maria Gloria Bonelli 'Profissionalismo, gênero e significados da diferença entre juízes e juízas estaduais e federais' (2011), 1 Contemporânea - Revista de Sociologia da UFSCar.
[337] Luvell Anderson, 'Racist humor' (2015), 10 (08) Philosophy Compass.
[338] Adilson Moreira, *Racismo Recreativo* (Pólen Livros, 2019).

The merits of content must then be evaluated in a decentralized system of self-regulation, where the tools discussed earlier are made available to users by social media companies[339]. Users should be the filters[340] because the vast majority of speech flows between agents without especially dominant communicative capacity such as large newspapers against regular people. It is regular people speaking in an environment that is not a perfect free market of ideas, but one that resembles that ideal scenario more than any other in history. A perfectly objective, static set of substantive norms on what expression is or is not allowed is not necessarily a precondition for such self-government to occur online[341], as spontaneous self-regulation of online worlds has shown for decades[342]. This solution offers the additional advantage of guaranteeing compatibility with local or regional norms of speech. If German users make content merit-based decisions on what constitutes hate speech in posts made by German users or specifically targeted at them, then German users will not be subjected to an American company's concept of hate speech. If Indian users make content merit-based decisions on what constitutes defamation in posts made by Indian users or specifically targeted at them, then Indian users will not be subjected to an European court's concept of defamation.

The conditions that warrant self-regulation[343] of speech are present and such a model is arguably more in line with liberal democracy ideals in that it means more people influencing public debate[344]. The central participation of users themselves in content regulation would avoid some of the conflict of interest problems typical to traditional regulation and diminished in self-regulation[345].

4.3 Platforms

The role of platforms is not to exercise review of individual instances of speech, but to enable and empower users to do so. Private media platforms cannot claim that they akin to the editor in charge of individually selecting the reader letters that make it to print. They have no free speech interest because they are not editors, but conduits. A legal system's freedom of expression guarantees should not be interpreted to warrant private platforms complete autonomy in deciding what speech they filter[346]. Quite to the contrary: "To the extent dominant search engines, social networks, and other new media simply enable connections between audiences and content, they fall even further on the 'less expressive'" end of a spectrum that includes traditional press vehicles, radio, television and cable networks[347].

Legislators should enact basic decentralized gatekeeping procedural obligations that private platforms need to comply with and courts should enforce such procedural rules. Private speech platforms should gradually be pushed to be less like editors and more like Wikimedia, managing "highly uneven geographies of participation"[348] and focusing on the intricate architecture of a

[339] Cliff Lampe et al., 'Follow the Reader: Filtering Comments on Slashdot', (28 april-03 may, 2007), ACM Conference on Human Factors in Computing Systems (CHI'07).
[340] Ivar A Hartmann, 'Let the Users be the Filter? Crowdsourced Filtering to Avoid Online Intermediary Liability' (2017), 2017 (01) Journal of the Oxford Centre for Socio-Legal Studies.
[341] Michael Risch, 'Virtual Rule of Law' (2009), 112 West Virginia Law Review.
[342] Sal Humphreys, 'Ruling the Virtual World. Governance in Massively Multiplayer Online Games' (2008), 11 European Journal of Cultural Studies.
[343] Anthony Ogus, 'Rethinking self-regulation' (1995), 15 Oxford Journal of Legal Studies.
[344] C. Edwin Baker, Media, Markets, and Democracy: Communication, Society and Politics (Cambridge University Press, 2001).
[345] Joseph Stiglitz, 'Regulation and failure', David A Moss, John A Cisternino, *New perspectives on regulation* (The Tobin Project, 2009).
[346] Jack Balkin, 'Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation' (2018), 51 UC Davis Law Review.
[347] Frank Pasquale, 'Platform Neutrality: Enhancing Freedom of Expression in Spheres of Private Power' (2016), 17 Theoretical Inquiries L.
[348] Mark Graham, Ralph K. Straumann, Bernie Hogan, 'Digital Divisions of Labor and Informational Magnetism: Mapping Participation in Wikipedia' (2015), 105 (06) Annals of the Association of American Geographers.

sophisticated system of user self-regulation that employs moderation tools never available to courts in the age of mass media. Such a system involves user hierarchy profiles that reflect each individual's contribution to moderation - have they often voted to downgrade posts that a majority of others found to be acceptable expression? Or were they one of a few users in their group within the ideology spectrum to denounce fake news that indulges the group's views? Is it the first or the hundredth time this user weighs in on hate speech?

The current model of safe harbor from liability has enabled the commercial incentives of private platforms to operate rampantly, causing the companies to filter more speech than is legally necessary and assuming a position of a quasi-judicial institution arbitrating what speech is acceptable and what is not in most societies[349]. There is already a rich literature describing the fallout of the predominance of such a private censorship system subjecting billions of people worldwide. To the lack of legitimacy we can today also add the perverse mental health effects on a large contingent of privately hired content moderators, who resemble factory workers due to the precarious working conditions and compensation, as has been documented by Sarah Roberts[350] and others.

Under a new framework for content moderation, it is up to the platform managers to ensure that nothing is removed without being evaluated by enough users - more than one, at the very least, as all things equal, a decision by several people is better than a decision by only a couple[351] [352]. Research on the quality of the results of online content moderation efforts shows quality is more often correlated with the effectiveness of collective coordination tools than with the number of overall editors[353].

Platforms should, in any event, seek active user moderators with different backgrounds, avoiding filter bubbles[354] and radicalization of certain positions[355]. The companies' targeted-advertising efforts allow them, for example, to pick users from different ethnicities or religions to review posts flagged as terrorism, in order to both curb its spread to global audiences[356] and avoid culturally-biased review. Likewise, social media companies should also ensure fake news[357] is not reviewed solely based on the opinion of one specific demographic and instead leverage a cross-ideological debate that might be small, but not nonexistent[358]. The platform should also permit users to self-select as red-flaggers or content reviewers, as more often than not - especially in the context of information goods - a central authority does not select as accurately for a task as the ones who will perform the task themselves[359]. The issue then is: how to create or allow the continuation of incentives for users to engage in such a task of moderation[360]? Uber and other gig economy platforms have implemented systems of review that stimulate detailed input and prioritize feedback that indicates the grounds for a negative or positive evaluation, as opposed to reviews that consist solely on a single click. Private media platforms should design their decentralized

[349] Kate Klonick, 'The New Governors: The People, Rules, and Processes Governing Online Speech' (2018), 131 Harvard Law Review.
[350] Sarah Roberts, *Behind the Screen: Content Moderation in the Shadows of Social Media* (Yale University Press, 2019).
[351] Krishna K. Ladha, 'The Condorcet Jury Theorem, Free Speech, and Correlated Votes' (1992), 36 (03) American Journal of Political Science.
[352] David Austen-smith, Jeffrey S. Banks, 'Information Aggregation, Rationality, and the Condorcet Jury Theorem' (1996), 90 (01) American Political Science Review.
[353] Aniket Kittur, Robert E. Kraut, 'Harnessing the Wisdom of Crowds in Wikipedia: Quality Through Coordination' (8-12 November, 2008), CSCW'08.
[354] Seth Flaxman, Sharad Goel, Justin M. Rao, 'Filter Bubbles, Echo Chambers, and Online News Consumption' (2016), 80 (s1) Public Opinion Quarterly.
[355] Cass R. Sunstein, *Republic.com 2.0.* (Princeton University Press, 2009).
[356] Daniel Weimann, 'Terror on Facebook, Twitter, and Youtube' (2010) V. XVI (II) Brown Journal of World Affairs.
[357] David O. Klein, Joshua R. Wueller, 'Fake News: A Legal Perspective' (2017), 20 (10) Journal of Internet Law.
[358] Eszter Hargittai et al., 'Cross-ideological discussions among conservative and liberal bloggers' (2008), 134 Public Choice.
[359] Yochai Benkler, 'Coase's Penguin, or, Linux and The Nature of the Firm' (2002), 112 Yale Law Journal.
[360] Sheizaf Rafaeli, Yaron Ariel, 'Online Motivational Factors: Incentives for Participation and Contribution in Wikipedia', Azy Barak, *Psychological Aspects of Cyberspace. Theory, research, applications* (Cambridge University Press, 2008).

moderation systems to also incentivize moderating users to provide detailed, even written grounds for their assessments. Input on moderation should include users who are close to the author of the post. The findings of a study with 1910 Facebook users "suggest that people derive benefits from receiving online communication, as long it comes from people they care about and has been tailored for them." [361]

Algorithms which weigh in on elements of content moderation that have been automated need to be developed and improved to reduce bias[362] but also to remain accountable[363] [364]. Content moderation AI, more than in most other fields, needs to produce evidence of what elements were given more weight for specific decisions[365] in order to be explainable AI. The current use of automated decision making by platforms to filter copyright violations is fundamentally flawed because of its lack of explainability and transparency of the grounds for each decision, denying the uploader as well as society in general the capacity to evaluate the system's performance[366]. Legally denying platforms the role of direct automated decision-maker on the merits of each post produced by its users would be a step towards dismantling, at least in the field of content moderation, a scenario accurately described by Frank Pasquale as a black box society[367].

4.4 Administration

Lastly, this new framework warrants a new role for the Administration as well. This is seemingly at odds with the traditional wisdom of modern democracies that the Executive branch should be nowhere near speech review. However, "protecting free speech values in the digital age will be (...) more and more a problem of technology and administrative regulation"[368]. To be sure, government should remain prohibited from exercising content moderation based on the merits of expression, but as I have attempted to show here, that is not the only piece of that puzzle that requires private and public action. While legislators establish the general procedural guidelines for content self-regulation systems where private platforms and users operate and courts review the extent to which platforms have respected such guidelines, the Administration should play a role in between[369]. Its task is to permanently audit such private, crowdsourced governance schemes in order to review procedural elements enshrined as architecture choices, which regulate *ex post ante*[370]. The source code cannot always be fully open and technical oversight is needed to make sure private platforms are not using code to avoid regulation[371]. The Administration is the best stakeholder to review the use of AI in content moderation, as courts are generally unable to gather all of the evidence and decide on matters of algorithm bias[372], a role some have already said requires the creation of an agency[373].

5. Final Remarks

[361] Moira Burke, Robert Kraut, 'The Relationship Between Facebook Use and Well-Being Depends on Communication Type and Tie Strength' (2016) 21 Journal of Computer-Mediated Communication.
[362] Thomas Davidson et al., 'Automated Hate Speech Detection and the Problem of Offensive Language' (2017), Proceedings of The 11th International AAAI Conference on Web and Social Media, <https://arxiv.org/abs/1703.04009>, acessed 20 jan. 2018.
[363] Mike Annany, Kate Crawford, 'Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability' (2013) New Media & Society.
[364] Nicholas Diakopoulos, 'Accountability in Algorithmic Decision Making' (2016), 50 (02) Communications of the ACM.
[365] Deven Desai, Joshua Kroll, 'Trust but Verify: A Guide to Algorithms and the Law' (2017), 31 (01) Harvard Journal of Law and Technology.
[366] Maayan Perel, Niva Elkin-Koren, 'Accountability in Algorithmic Copyright Enforcement' (2016), 19 Stanford Technology Law Review.
[367] Frank Pasquale, *The Black Box Society* (Harvard University Press, 2015).
[368] Jack Balkin, 'The Future of Free Expression in a Digital Age' (2009), 36 Pepperdine Law Review.
[369] Jack Balkin, 'Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation' (2018), 51 UC Davis Law Review.
[370] Lawrence Lessig, 'The zones of cyberspace' (1996), 48 Stanford Law Review.
[371] Tim Wu, 'When code isn't law' (2003), 89 Virginia Law Review.
[372] James Grimmelmann, 'Speech Engines' (2014), 98 Minnesota Law Review.
[373] Andrew Tutt, 'An FDA for Algorithms' (2017), 69 (01) Administrative Law Review.

My goal with this paper was first to provide a reformulated account of what the activity of separating legitimate from illegitimate expression can be in light of profound changes in the information flow in current societies. Second, it was to bring together concerns found in the literature about online speech and add a few innovative propositions in order to present an overview of a new framework for the protection of expression on the internet.

The descriptive part covers characteristics of balancing, judicial activity and online content moderation today that are similar in most constitutional democracies. On the other hand, the framework that I propose is agnostic to rules on the merits of content - what exactly constitutes hate speech in one country or what configures defamation or copyright violations in another are considerations that are not germane to the main theoretical elements of this framework.

These elements are focused on procedural fairness and a rearrangement of institutional roles. My first core new proposition is that courts should gradually refrain from review of the merits of specific instances of expression and limit their analysis to whether platforms respect certain legally established due process standards such as decentralized gatekeeping and decision making transparency. The second central proposition is that platforms must avoid systems where moderation of individual posts is performed by platform employees or AI and slowly design and implement systems where this task is taken up exclusively by users themselves. The third central proposition is that the Administration has a new and up until now unlikely role of regulating details of these due process requirements that private platforms must follow, in a way that is viewpoint neutral. Oversight of private content moderation by, for example, a regulatory agency, is especially needed to ensure accountability of automated elements of online content moderation, a task that requires technical expertise not available to courts and where total transparency is not only impossible but undesirable.

Naturally, a comprehensive proposal such as this cannot be fully described in one article. This is merely an outline that presents the basic underpinnings. Nevertheless, it already indicates certain paths down which to advance and the ones that should be should not be threaded as countries struggle to find proper solutions for the adequate protection of expression online today.

Author Biography

Ivar Hartmann hold an MSc in public law from the Catholic University of Rio Grande do Sul (Brazil), an LL.M. from Harvard Law School and an SJD in public law from Rio de Janeiro State University. He is a professor and research project coordinator at FGV Law School in Rio de Janeiro since 2012 and is currently head of the Center for Technology and Society (CTS/FGV). Ivar teaches courses at the bachelors, masters and PhD level at FGV Law School on topics related to regulation of technology as well as programming and data science. His main research areas are cyberlaw - especially online speech - and judicial politics.

Socio-Ethical Values and Legal Rules on Automated Platforms: The Quest for a Symbiotic Relationship

Rolf H. Weber

University of Zurich, Zurich, Switzerland

Keywords:

Compliance, impact assessment, rule-making, standard terms, trust

Abstract:

The deployment of artificial intelligence on automated platforms needs to go hand in hand with the development of a legal framework safeguarding socio-ethical values as well as fundamental rights, particularly the self-determination and the non-discrimination principle. A trust-based approach focused on human values can mitigate a potential clash between a solely market- and technology-oriented use of artificial intelligence and a more inclusive multistakeholder approach. The regulatory tools are to be designed in a manner that leads to a symbiotic relationship between ethics and law.

1. Introduction

Artificial intelligence (AI) offers new business opportunities which have an impact on platforms and markets. The use of intelligent "devices" and the availability of algorithms on platforms include the potential to replace human activities by software and/or machines. Instead of a human intervention, the programming of the code, which executes the tasks, becomes important; an automation of platforms can take place for manifold business models. This fact calls for the implementation of fundamental socio-ethical values into the AI within an appropriate legal framework.

Artificial intelligence allows implementing a "regime" of automated decision-making being conducted in a very timely and effective manner. Such kind of automation is mainly feasible in situations not requiring a specific human input, for example in case of an algorithm-driven search or in case of a standardized exchange platform. However, the automated decision-making can cause many socio-ethical and legal challenges. Hereinafter, this contribution is going to conceptualize the value dimension in respect of automated platforms and to analyze possible regulatory tools that could help implementing the appropriate safeguards for its practical realization.

AI-driven platforms do have an impact on civil society as well as on the competitive environment. In order to reconcile socio-ethical values with legal rules, the following questions derived from a normative concept of society are to be addressed:[374]

- Do the AI processes comply with fundamental principles such as human rights and non-discrimination?

[374] So also Rolf H. Weber, Dürfen Maschinen über Menschen entscheiden? Eine rechtliche Auslegeordnung im Lichte neuer Technologien, Schweizer Monat, Februar 2019, 72 et seq.

- Is the automated decision-making based on a sufficient legal basis, at least in respect of governance-related matters?
- Does an automated decision-making comply with all the applicable requirements of data protection laws?
- Who is responsible for the monitoring of socially responsible activities and liable in case of a failure caused by the algorithms?

The following contribution mainly discusses the first question; it assesses the possibilities of embedding socio-ethical values into the AI systems by way of a trust-oriented framework with regulatory tools being suitable to minimize technological risks and attempting to place the human being in the center of AI deployment. The third question is now subject to the application of article 22 of the General Data Protection Regulation (GDPR) having been intensively debated during the last years. The second and the fourth question are already subject to literature on administrative law and tort law.

2. Conceptualization of Values

Value conceptualizations can be done in various ways. In view of this contribution's target to reconcile socio-ethical and legal issues related to automated platforms in a potentially symbiotic relationship, the respective two dimensions need to be analyzed in more detail at the outset.

2.1 Socio-ethical Dimension

The development of automated platforms must be embedded into a broader socio-ethical environment. The inherent complexities associated with artificial intelligence and algorithms require the assessment of new technological advances and systems as a whole. As an example, the following substantive principles developed by the United States Association for Computing Machinery (as a technologically oriented association)[375] are worth to be taken into account with regard to the implementation of automated platforms:

- *Awareness:* All stakeholders should become aware of the possible biases involved in the design and implementation of algorithms.
- *Access and redress:* The adoption of mechanisms that enable questioning and redress for individuals and groups should be encouraged.
- *Accountability:* Institutions must explain how the algorithms produce their results and should become responsible for decisions made by algorithms.
- *Explanation:* The procedures followed by the algorithms and the specific decisions made should be properly explained.
- *Data provenance:* A description of the collection and processing of data is necessary; the principles can be drawn from data protection law.
- *Auditability:* Models, algorithms, data, and decisions should be recorded so that they can be audited if needed.

[375] USACM, Statement on Algorithmic Transparency and Accountability, updated May 25, 2017, 2.

- *Validation and testing:* Rigorous methods to validate the models and algorithms must be implemented (with routinely performance tests).

These substantive principles must become part of an ethically aligned framework of AI processes. According to the IEEE as another technologically oriented organization,[376] the ethical and values-based design, development, and implementation of autonomous and intelligent systems should be guided by the following general principles: human rights, well-being, data agency, effectiveness, transparency, accountability, awareness, and competence. These principles are not easy to implement, particularly in case of contradictions. But compliance with the values developed by the ethics discipline remain a desirable objective.[377]

In addition, the expertise and the knowledge of civil society must be improved: People should have the ability to better assess the consequences of AI processes; societal implications of AI are to be exposed, for example algorithmic biases and de-anonymization.[378] Based on this understanding, policies and regulations helping the society to adapt more easily to the use of AI can be developed.[379] Similar processes should also be implemented in the public administration; since automated decision-making systems do have a big impact on individuals and society, public administration must ensure the appropriate deployment of AI processes.[380]

2.2 Legal and Economic Dimension

Law as a structural system that expresses legal norms in a linguistic form gives guidance about the desired behavior.[381] Thereby, normative expectations of civil society can be stabilized. The functions of law crystalize in rules and institutions that underpin civil society, facilitate orderly interaction and resolve disputes and conflicts arising in spite of such rules.[382] Law is in a position to allow people and businesses in a community to determine the limits of what can and cannot be done in their collective interest.[383] Thereby, the rule of law helps to achieve a high degree of certainty and predictability of legal norms; correspondingly, the authorities have to employ their discretion within the limits of the implemented rules.[384]

At first instance, artificial intelligence is a technology; nevertheless, the discussions about automated decision-making processes should not be limited to technology, for example the issues of data security, data accuracy and data quality.[385] Moreover, it is important to assess how AI is procured and finally deployed. Thereby, the human agency also plays a role; social systems cannot only be conducted by machines.[386] An important part of the legal framework enshrines fundamental rights; as a consequence, the ongoing dialogue regarding the ethics of AI should

[376] IEEE, Ethically Aligned Design, Version 2, July 2018, 20-32, available at https://ethicsinaction.ieee.org.
[377] Partly a skeptical assessment that such compliance will be possible is made: see for example Guido Noto La Diega, Against the Dehumanisation of Decision-Making, Algorithmic Decisions at the Crossroads of Intellectual Property, Data Protection, and Freedom of Information, JIPITEC 9 (2018), 3, nos. 10 et seq.; in contrast, other authors plead for a demystification of AI: Philippe Lorenz/Kate Saslow, Demystifying AI & AI Companies, in: Stiftung Neue Verantwortung (Hrsg.), Berlin, July 2019.
[378] Algorithm Watch/Bertelsmann Stiftung, Taking Stock of Automated Decision-Making in the EU, January 2019, 14, available at https://algorithmwatch.org/wp-content/uploads/2019/01/Automating_Society_Report_2019.pdf.
[379] Rolf H. Weber, Digitalisierung und der Kampf ums Recht, in: A. Dal Molin Kränzlin/A. M. Schneuwly/J. Stojanic (Hrsg.), Digitalisierung – Gesellschaft – Recht, Analysen und Perspektiven von Assistierenden des Rechtswissenschaftlichen Instituts der Universität Zürich, Zürich/St. Gallen 2019, 3, 15 et seq.
[380] Algorithm Watch/Bertelsmann Stiftung (n. 378), 15.
[381] Rolf H. Weber, Realizing a New Global Cyberspace Framework, Zürich 2014, 33.
[382] Warren B. Chik, „Custormary Internet-ional Law": Creating a Body of Customary Law for Cyberspace. Part I: Developing Rules for Transitioning Customs into Law, CLSR 26 (2010), 3, 6.
[383] Weber (n. 381), 34.
[384] See also Chris Reed, Making Laws for Cyberspace, Oxford 2012, 70 et seq.
[385] Algorithm Watch/Bertelsmann Stiftung (n. 378), 15.
[386] Ben Wagner, Liable, but Not in Control? Ensuring Meaningful Human Agency in Automated Decision-Making Systems, Policy & Internet 11/1 (2019), 104.

expand to consider the human rights implications of these technologies[387] as well as the risks of discrimination.[388]

The term "value" can also have an economic meaning, mainly in the context of the value chain. Such a chain encompasses various steps:[389] (i) In the first phase an individual or an organizational entity (company, government) is creating a value, usually with the intention to exploit such value. (ii) Particularly in the case of data as value, the question follows whether the creator is entitled to build a private "data silo" or whether certain or all third parties should have access to the data; the applicable access rules have to lead to a balance of interest analysis weighing the incentives of the data creator and the needs of society. (iii) A regulatory interference into the value process or the value change appears to mainly be justified if the creator of data is earning an "economic rent" which exceeds the amount justified under economic and/or social considerations following the exploitation of an advantageous position ("extra returns that firms or individuals obtain due to their positional advantages"[390]). This situation is likely to happen in case of automated platforms since the its "owner" or "controller" often enjoys a market-dominant position; obviously, competition law can intervene but the respective instruments usually only have a delayed effect.[391] Notwithstanding the regulatory treatment of economic rents the general objective remains at stake that a trust-oriented framework must find a reasonable reconciliation between the value creation and the value extraction.

2.3 Symbiotic Relationship between Ethics and Law?

As outlined, both socio-ethical and legal elements play a role within a framework of automated platforms and of AI governance in general. However, even if both are necessary, neither is sufficient and neither can substitute the other.[392] Moreover, the two disciplines act in a complementary way, being able to inspire each other. Not surprisingly for many years already, the respective roles and territories are intensively discussed and analyzed.[393]

In view of this assessment, efforts must be strengthen to develop a comprehensive approach for the socio-ethical and the legal dimension of value conceptualizations moving into the direction of a potentially symbiotic relationship. The Council of Europe and the European Commission are in the process of undertaking respective efforts.[394] Academia is also called to develop interdisciplinary thoughts and studies; this contribution attempts to lay the foundation for such a comprehensive approach.

---

[387] Filippo A. Raso/Hannah Hilligoss/Vivek Krisnamurthy/Christopher Bavitz/Levin Kim, Artificial Intelligence & Human Rights: Opportunities & Risks, Berkman Klein Center for Internet & Society at Harvard University, Research Publication No. 2018-6, September 25, 2018, 4; Mathias Risse, Human Rights and Artificial Intelligence: An Urgently Needed Agenda, Carr Center for Human Rights Policy, Harvard Kennedy School, Cambridge Mass. 2018, available at https://carrcenter.hks.harvard.edu/files/cchr/files/ccdp_2018_002_hrandai.pdf.

[388] Frederik Zuiderveen Borgesius, Discrimination, artificial intelligence, and algorithmic decision-making, Study published by the Council of Europe, Strasbourg 2018, 10 et seq.

[389] For an overview see European Commission, Building a European Data Economy, COM(2017) 9 final, January 2017.

[390] This definition stems from Lucian Bebchuk/Jesse Fried, The managerial power perspective, in L. Bebchuk/J. Fried (eds.), Pay without performance: the unfulfilled promise of executive compensation, Cambridge Mass. 2004.

[391] The discussions about the so-called „economic rent" have been quite intensive during the last few decades; the details are not further analyzed in this contribution (for a general overview see Robert D. Tollison, Rent seeking: a survey, Kyklos 35/4 [1982], 575 et seq.).

[392] Nathalie A. Smuha, The EU-Approach to Ethics Guidelines for Trustworthy Artificial Intelligence, CRi 2019, 97, 101.

[393] See for example Gregory C. Shaffer/Mark A. Pollack, Hard vs. Soft Law: Alternatives, Complements, and Antagonists in International Governance, Minnesota Law Review 94 (2010), 706-799; Ryan Hageman/Jennifer Huddleston/Adam D. Thierer, Soft Law for Hard Problems: The Governance of Emerging Technologies in an Uncertain Future, Colorado Technology Law Journal, February 2018, available at https://ssrn.com/abstract=3118539.

[394] See below chapters 0.2 and 3.1.3.

3. Rule-making and Compliance in the Automated Platform Context

3.1 Legitimacy for Rule-making

3.1.1 Processes and Mechanisms of Rule-making

After having described the relevant socio-ethical and legal values, the processes and mechanisms being best suited to implement the respective values and social justice must be identified. This task does not only encompass substantive elements but also organizational and procedural factors. Subsequent to some general comments, the efforts of the Council of Europe and of the European Commission are briefly outlined hereinafter.

Rule-making issues can be addressed from the perspective of different disciplines; nevertheless, in private matters such as in case of online platforms the discussions must concentrate on the appropriate allocation of duties and responsibilities as well as the proper structuring of the concerned "organization" (offeror of the platform). In other words: rule-making, at whatever level of social organization it may take place, refers to setting norms for the conduct of the business in an appropriate way. In this context, some key questions are to be asked and answered:[395] (i) Who is entitled to set the rules?, (ii) in whose' interest?, (iii) by which mechanisms? and (iv) for which purposes? The need is given to develop overarching networks and negotiation systems between the difference stakeholders thus forming a cooperative approach to rule-making that includes the whole society, hence dividing responsibilities between public and private actors.[396] Thereby, governmental regulations must be supplemented by self-regulatory initiatives; in particular, guidelines designing the normative framework for the activities executed by the "owners" or "controllers" of automated platform appear to be an appropriate instrument.

3.1.2 Council of Europe's Efforts

In 2018 a special group of experts (Consultative Committee of the Convention for the Protection of Individuals with regard to Automated Processing of Personal Data) has worked out a Report on Artificial Intelligence with the title "Artificial Intelligence and Data Protection: Challenges and Possible Remedies". The more than 20 pages long Report was submitted to the Council of Europe on December 3, 2018. Notwithstanding the fact that the origins of this Report, written by Professor Alessandro Mantelero (University of Turin), was originally rooted in the field of data protection, its contents cover wide areas of artificial intelligence. As challenges and possible remedies, the Report addresses limitations to artificial intelligence use, transparency, risk assessment, participatory assessment, liability and vigilance as well as sector-specific issues.[397]

Based on this Report, the Committee of Ministers of the Council of Europe adopted the Declaration on the manipulative capabilities of algorithmic processes on 13 February 2019.[398] Amongst others, the Council of Europe encourages Member States to assume the responsibility to address their artificial intelligence threats by "taking appropriate and proportionate measures to assure that effective legal guarantees are in place against such forms of illegitimate interference" (lit. d) and by "empowering users by promoting critical digital literacy skills and robustly enhancing public awareness of how many data are generated and processed by personal devices, networks, and platforms through algorithmic processes" (lit. e). The Declaration also

[395] In general to legitimacy aspects Rolf H. Weber, Shaping Internet Governance: Regulatory Challenges, Zurich 2009, 105 et seq.
[396] Weber (n. 381), 112/13; see also chapter 0 below.
[397] Council of Europe, Report on Artificial Intelligence, T-PD(2018)09 Rev, Strasbourg, 3 December 2018, 11 et seq.
[398] Decl(13/02/2019)1 of 13 February 2019.

draws the attention to the necessity of critically assessing the need for stronger regulatory or other measures to assure adequate and democratically legitimized oversight over the design, development, deployment and use of algorithmic tools.

3.1.3 European Commission's Efforts

As the heart of its strategy in the artificial intelligence context and as a response to the increasing ethical questions raised by this technology, the European Commission, after having already published some ideas in 2018,[399] established an independent High-Level Expert Group in Artificial Intelligence (AI HLEG) in June 2018. This Group had the task to draft two instruments, namely AI Ethics Guidelines as well as Policy and Investment Recommendations. The AI HLEG was composed of neutral experts of different disciplines coming from academia and practice. The Group intensively worked for months and already published the Ethics Guidelines in April 2019 based on a fundamental rights approach and setting out a comprehensive framework to achieve "Trustworthy AI".[400] The notion of "Trustworthy AI" enshrines three components: actors and processes involved in AI systems should be lawful (complying with all applicable laws and regulations), ethical (ensuring adherence to ethical principles and values) as well as robust (both from a technical and social perspective).

The Ethics Guidelines contain four ethical principles, namely (i) respect for human autonomy, (ii) prevention of harm, (iii) fairness, and (iv) explicability (or explainability). From these principles, seven key requirements are derived which should be taken into account by AI systems,[401] namely (1) human agency and oversight, (2) technical robustness and safety, (3) privacy and data governance, (4) transparency, (5) diversity, non-discrimination and fairness, (6) societal and environmental wellbeing and (7) accountability (including auditability).

In the Policy and Investment Recommendations on Artificial Intelligence having been published in June 2019 the AI HLEG pleads for the establishment of an appropriate governance and regulatory framework expressed in a comprehensive approach.[402]

3.2 Towards a Broader Rule-making Approach

The experiences of the last few months and in particular the work of expert groups appointed by the Council of Europe and by the European Commission have shown that the traditional rule-making approach in international matters, namely the conclusion of multilateral treaties, does not fit the objectives of a regulatory framework setting guidelines for AI systems (and automated platforms). Moreover, other mechanisms have to play a more important role. A new rule-making approach having been developed (and partly also applied) in the Internet governance (as well as climate change/sustainability) context is the multistakeholder participation model.[403] If all concerned persons and organizations of the public and the private sphere are involved in the discussions and negotiations of the regulatory framework for AI processes, the chances are increasing that the developments are in the interest and to the benefit of the whole society.[404]

Practical experience transmits the lesson that some basic challenges need to be addressed in order to make the multistakeholder concept successful; mainly, if several forms of co-operation

[399] See for example Commission Communication "Artificial Intelligence for Europe", COM(2018) 237 final, 25 April 2018, and COM(2018) 795 final, 7 May 2018.
[400] Available at https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=58480.
[401] For more details see Smuha (n. 392), 99/100.
[402] For more details see Smuha (n. 392), 102.
[403] Report of the Working Group on Internet Governance, June 2005, available at: www.wgig.org/docs/WGIGREPORT.pdf.
[404] See also Rolf H. Weber, Legal foundations of multistakeholder decision-making, Zeitschrift für schweizerisches Recht 135 (2016) I 247, 249.

based on a variety of actors involved are be taken into account. Thereby, four fundamental questions must be tackled:[405] (i) How do governance groups best match challenges with the organizations and networks? (ii) How can governing bodies/entities be most able to help develop legitimate, effective, and efficient solutions? (iii) How should the flow of information and knowledge necessary for successful governance be structured? (iv) How can different governance groups approach a coordination between the available governance networks in order to avoid conflicting interest?

Answers to these questions need a differentiated thinking depending on the given environment. At any rate, however, in realizing an appropriate governance framework, civil society involvement should be strengthened in (automated) platform matters. As items of shortcomings the identification of adverse effects of automated decision-making in the relevant policy field, the facilitation of networking opportunities and the public support of civil society interventions are to be considered. The involvement of different stakeholders should also be achieved in the development of criteria for good design processes and audits.[406]

On the one hand, the AI era (incl. automated platforms) needs, as shown, a broader and more complex consideration of values exceeding a narrow perception of legal rights, and, on the other hand, the traditional legal instruments, particularly the multilateral treaties, do not suffice anymore to tackle the challenges in the digital world. Acting in compliance with ethical principles improves the reputation of automated platforms which in turn helps to gain the trust of users and make the offerors more attractive.[407] The mirroring of compliance in reality means to implement practical standards based on self-regulatory instruments and/or provide for contractual terms which are appropriate and fair deviating from an asymmetrical private regulation; in other words, the terms of service of platform operators need to take into account the interests of all involved stakeholders.[408]

3.3 Compliance on Automated Platforms

3.3.1 Substantive Issues

In respect of the compliance requirement with the applicable regulatory framework, an interdisciplinary approach appears to be appropriate. In general, the following issues related to algorithms/artificial intelligence applied on platforms might have to be taken into account:[409]

- *Property rights*: Algorithms could (also) be used to collect, aggregate, display, and share informational goods protected by intellectual property law.

- *Privacy rights*: Automated systems collecting personal data from users and processing them with algorithms cause concerns for data protection and privacy (taken now up by the new provision of Article 22 GDPR).[410]

[405] See Weber (n. 404), 249 and Urs Gasser/Ryan Budish/Sarah Meyers West, Multistakeholder as Governance Groups: Observations from case Studies, Berkman Center for Internet & Society Research Publications 2015, 1 et seq.
[406] See below chapter 3.3.2.
[407] Rolf H. Weber, Ethics in the Internet Environment, in: Global Commission on Internet Governance, Paper Series no. 39, July 2016, 4.
[408] Weber (n. 407), 4/5.
[409] OECD, Directorate for Financial and Enterprise Affairs, Competition Committee, Algorithms and Collusion – Background Note by the Secretariat, Paris, 9 June 2017, 43.
[410] The specific challenges caused by automated systems in the data protection context are not a topic of this contribution; for further information see Lilian Edwards/Michael Veale, Slave to the Algorithm? Why a «Right to an Explanation» is probably not the Remedy You are Looking For, Duke Law & Technology Review 16/1 (2017), 18 et seq.;

- *Censorship*: Algorithmic programs can introduce restrictions to control or block content that otherwise is accessible for users.

- *Discrimination*: Automated data-decision processes have the potential to lead to social discrimination based on the processed personal information.[411]

- *Abuse of market power*: Algorithms (artificial intelligence) could facilitate the application of exclusionary and exploitative measures.

- *Tacit collusion*: If algorithms coordinate competition parameters even without personal intent of market players, anti-competitive effects similar to the well know concerted practices can occur.

- *Manipulation*: Manipulated algorithms collect and select information in view of given business and political interests instead of its relevance or quality.[412]

The above list of substantive issues is neither comprehensive nor enumerative. Depending on the given environment, only some issues are relevant or additional issues merit attention. In the context of automated platforms, for example property rights, privacy rights, discrimination, abuse of market power, and manipulation can play a role.

Irrespective of its design, a technical system such as an automated platform should inspire trust; from an ethical perspective trust enhances cooperation and fosters reciprocal relations.[413] A proper compliance can support the building of trust enshrining more than legal standards but also socio-ethical values that have a broacher scope than the traditional justice.[414] Further, trust is also linked to "reliance"; if an individual is relying on something or someone to display a certain behavior, than confidence in the respective activities will increase.[415] Reliance can be based on standards or contractual terms, observed by the offeror of a service, leading to transparency and accountability.[416]

3.3.2 Organizational and Procedural Structures

(i) In order to assess compliance with the manifold objectives to be observed in the context of AI processes, the implementation of appropriate organizational and procedural structures is necessary. A good way forward is the establishment of ethics committees as an institutional measure, i.e. such committees often are a worthwhile mechanism.[417] The implementation of ethics committees can be done at various levels, for example in the form of a national committee or of an internal committee of the concerned businesses:[418]

Stefanie Hänold, Profiling and Automated Decision-Making: Legal Implications and Shortcomings, in: M. Corrales et al. (eds.), Robotics, AI and the Future of Law, Singapore 2018, 123 et seq.
[411] For more details see Zuiderveen Borgesius (n. 388), 10 et seq.
[412] To the democracy aspect of the manipulation see the Declaration of the Council of Europe (n. 398).
[413] Weber (n. 407), 7.
[414] Onora O'Neill, A Question of Trust, Cambridge 2002, 61 et seq.
[415] Weber (n. 407), 7.
[416] These two tools are discussed below in chapters 0 and 0; see also Onara O'Neill, Justice, Trust and Accountability, Cambridge 2005.
[417] As an example see the Guidelines of the IEEE (n. 375).
[418] Council of Europe (n. 397), 15; see also Alessandro Mantelero, Regulating Big Data. The Guidelines of the Council of Europe in the Context of the European Data Protection Framework, CLSR 33/5 (2017), 584 et seq.

- A national committee could implement general guidelines on issues of AI development or of AI deployment; this approach would contribute to a desirable standardization increasing the foreseeability of the normative ecosystem.

- An internal committee could support the responsible persons for the AI operations and processes in a focused way. Such experts who need to be independent from the corporate bodies might assume a broader role and act not only on ethical issues, but also on a more extended range of societal issues relating to AI, including the contextual application of fundamental rights.

Ethics committees may play an even more important role if transparency and participatory assessment are difficult to achieve; at any rate, the valuable support of the respective experts to AI developers in designing rights-based and socially-oriented algorithms leading to increased trust should not be underestimated.[419]

(ii) From a procedural perspective, following the concept of the privacy (data protection) impact assessment, legal doctrine has proposed to introduce the concept of a social impact assessment or a human rights impact assessment; these procedures could help to identify the societal consequences of AI and its impact on fundamental rights, collective values, public participation, individual and group empowerment as well as non-discrimination policies.[420] Such an instrument comes close to the notion of a general risk assessment (however, with specific elements), which entails the definition of the project risks, the implementation and monitoring of protection measures and possible mitigation mechanisms.[421] Impact assessments are an interdisciplinary task encompassing academics from different fields, civil society groups and potentially concerned individuals being able to contribute their experience to the systems' discussion.[422]

Finally, vigilance and liability remain an open issue for various reasons.[423] The existing legal framework with product liability rules and tort rules does not fully cover the needs of the risk management requirements and the uncertainties in the new technological fields.[424] New regulatory models and strategies are to be developed; amongst others, supervisory authorities could adopt forms of algorithm vigilance analogous to other risk-exposed market segments.[425]

4. Regulatory Tools

In view of the technological innovations, the legal order is confronted with the need to provide for appropriate instruments in the data-driven economy. Legal doctrine has looked at transparency and accountability for many years, not at least in the Internet governance context.[426] These regulatory tools should also be taken into account in the artificial intelligence environment, i.e. these principles being of importance in many segments of society are particularly relevant in the context of platforms. Furthermore, safety and robustness of platforms providing for trust need to be achieved. Therefore, additional regulatory interventions appear to be unavoidable in the AI context.

[419] Council of Europe (n. 397), 15/16.
[420] For a comprehensive discussion of such new forms of impact assessments see Alessandro Mantelero, AI and Big Data: A Blueprint for a Human Rights, Social and Ethical Impact Assessment, CLSR 34/4 (2018), 754 et seq.
[421] See also Zuiderveen Borgesius (n. 388), 28.
[422] A respective reference is contained in the provisions to the data protection impact assessment (article 35 para. 9 GDPR).
[423] Council of Europe (n. 397), 17.
[424] Rolf H. Weber, Liability in the Internet of Things, Journal of European Consumer and Market Law 2017, 207 et seq.
[425] Council of Europe (n. 397), 17/18.
[426] Weber (n. 395), 122 et seq.

4.1 Transparency

Transparency is usually defined as "easily seen through, … evident, obvious, clear".[427] Historically, with reference to Supreme Court Judge Louis Brandeis, the early promoter of privacy (in the form of a "right to be let alone", 1890) as well as of transparency ("sunlight is said to be the best of disinfectants", 1914), the term "transparency" can be understood as the prohibition of arbitrary and unforeseeable actions by the absolute sovereign in justice matters requiring the publication of the law in force.[428]

All involved stakeholders should promote a culture of transparency enshrining the disclosure of the used AI applications, a description of their logic as well as the access to the structure of algorithms and to the introduced datasets.[429] But notwithstanding the fact that transparency is important with a view to public scrutiny of AI processes, a generic statement on the use of AI does not allow to easily assess the all challenges and risks.[430] Concrete circumstances do play a role; therefore, a solution focused on disclosing the logic of algorithms may be the better option.

Transparency requires robust and general rules, not necessarily more regulation. The improvement of transparency does not mean to have a quantitative increase of information, but "more" in terms of higher information quality.[431] A future-oriented understanding of transparency helping to identify unfair value extraction must observe the following elements:[432]

- Existence of publicly reliable information, i.e. substantive quality standards related to information, supported by an adequate legal framework which influences the individuals' choices since a rational person would arguably organize his or her conduct in accordance with law.

- Definition of the information recipient as a holder of rights and an essential component for the perception of both information and transparency;

- Availability of disclosure procedures and observance of the time element, i.e. transparency implies constant visibility of information.

Giving information about the type of input data and the expected output, explaining the variables and their weight, and sharing light on analytics architecture usually contributes to transparency in respect of the logics of AI algorithms.[433] Transparency can be both, an *ex ante* or an *ex post* requirement for data-centered decision-making; transparency also means understandable and forward-looking information, appropriate to the context and the state of art, in order to make the various stakeholders aware of their interactions. [434]

[427] Oxford English Dictionary Online, 1989.
[428] For more details see Christine Kaufmann/Rolf H. Weber, The role of transparency in financial regulation, Journal of International Economic Law 13 (2010), 779, 782.
[429] Council of Europe (n. 397), 11/12.
[430] See Mike Ananny and Kate Crawford, Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability, New Media & Society 2016, available at https://doi.org/10.1177/1461444816676645.
[431] Christine Kaufmann/Rolf H. Weber, Transparency of Central Banks' Policy, in: P. Conti-Brown/R. M. Lastra, Research Handbook on Central Banking, Cheltenham/Northampton 2018, 518, 520.
[432] Weber (n. 395), 121 et seq. with further references, particularly 131.
[433] Council of Europe (n. 397), 12.
[434] Weber (n. 395), 122/23.

4.2 Accountability

All stakeholders being involved in datafication and artificial intelligence mechanisms should be accountable for the proper functioning of the used systems as well as for the respect of the regulatory environment.[435] In the data protection context Art. 22 GDPR now sets the regulatory requirements for the automated decision-making.[436] In general, accountability can be qualified as the acknowledgment and assumption of responsibility for actions, products, decisions and policies within the scope of the designated function.

Accountability consists in the obligation of a person to another, according to which the former must give account of, explain and justify his/her actions and decisions against criteria of the same kind.[437] Thereby, the proportionality principle inspiring an adequate and appropriate deployment of AI should apply.[438] Accountability also relates to good governance; the development of the respective concepts in public institutions and private enterprises are requiring publicly assessable accounts as a pre-condition for a sustainable society.[439]

The obligation to be accountable encompasses the task to disclose information about the actual "activities" of AI processes; in order to improve the respective foreseeability, standards should be developed and introduced which design the behavioral requirements in a more concise manner. Furthermore, the responsibility of the accountable person to keep those concerned individuals and businesses harmless from damages having suffered a detriment is to be legally outlined in a more precise way.[440]

4.3 Safety and Robustness

Apart from transparency and accountability, the safety and robustness of the platforms equally are of importance; these properties are the basis of trust.[441] The respective "infrastructure" and the related software programs must be safe and robust throughout their entire lifecycle so that the data-driven communications and transactions can overcome adverse conditions or foreseeable potential misuse. As a consequence, the traceability of the datasets, processes and decisions must be secured.

International instruments already state that the likely impact of intended AI processing and its broader ethical and social implications must be adequately taken into account in order to safeguard human rights and fundamental freedoms. For example, the Recommendation of the OECD Council on Artificial Intelligence, adopted on 22 May 2019, refers in part 1.4 to robustness, security and safety as follows:[442] "AI systems should be robust, secure and safe throughout their entire lifecycle so that, in conditions of normal use, foreseeable use or misuse, or other adverse conditions, they function appropriately and do not pose unreasonable safety risk. To this end, AI actors should ensure traceability, including in relation to datasets, processes and decisions made during the AI system lifecycle, to enable analysis of the AI system's outcomes and responses to inquiry, appropriate to the context and consistent with the state of art".

[435] Weber (n. 381), 78 et seq.
[436] See the respective references in note 410 above.
[437] Weber (n. 381), 78; Weber (n. 395), 133.
[438] Weber (n. 395), 137 et seq. with further references.
[439] Kaufmann/Weber (n. 428), 789.
[440] Weber (n. 395), 147.
[441] See above chapter 0.
[442] OECD, Recommendation of the Council on Artificial Intelligence of 22 May 2019.

Furthermore, use of AI by modern data processing techniques and the trend towards implementation of data-intensive processes require a more advanced risks' assessment understanding by individuals and businesses since possible adverse outcomes of data processes cannot be excluded.[443] In particular, the challenge must be tackled that automated decision-making may have an impact on fundamental rights including discrimination as well as the collective social and ethical values. The compliance with socio-ethical values (as well as their assessment) is more complicated in case of AI deployment than in case of a traditional data processing.

In this context it should not be overlooked that risks and compliance assessments are not only justified by the collective social and ethical values as well as by the nature of the fundamental rights and freedoms potentially affected by AI application, but they also represent an opportunity (if undertaken in a participatory environment) to better fostering public trust as key objective of the information society.[444]

5. Outlook

The development of artificial intelligence and automated decision-making on platforms has become a reality in the given technological environment. As a consequence, different approaches are emerging in the regulation of AI processes and AI deployment. Thereby, a key driver must consist in the attempt to find ways of safeguarding public trust, fundamental rights, the personal self-determination and the non-discrimination principle; these constitutional yardsticks should not only apply in the context of governmental activities, but also with respect of platform providers organized as private enterprises. Recently, the G20 Ministerial Statement[445] convincingly refers to the need of having implemented "human-centred Artificial Intelligence" and the mentioned AI Guidelines of the European Commission crystalize the "Trustworthy AI" in the terms "lawful, ethical and robust".[446] The OECD Recommendation on Artificial Intelligence of 22 May 2019 also promotes values and fairness in the interest of humans (dignity, autonomy, equality, non-discrimination, social justice, etc.).[447]

The adoption of a perspective that is focused on socio-ethical and legal values can mitigate a potential clash between a solely market- and technology-oriented development of AI and a more inclusive stakeholder model. Thereby, a trust-oriented approach complying with the basic values should strongly impact the AI development. The concrete implementation of these principles in practice can be facilitated by improved state-of-the-art standardization of platform business models and by more appropriate interest-balancing terms of service applied by platform providers. Such kind of standard terms must be lawful, ethical and robust; in addition, they have to meet the basic principles of a "Trustworthy AI", namely the respect for human autonomy, prevention of harm, fairness, and explicability.[448]

[443] Council of Europe (n. 397), 13.
[444] See above chapter 0 and the books of Onara O'Neill, cited in n. 414 and n. 416.
[445] Adopted in Tsukuba City, Japan, on 8/9 June 2019, nos. 17 et seq.
[446] See above chapter 0.
[447] OECD Recommendation (n. 442), no. 1.2.
[448] See chapter 0. above.

Democratising Online Content Moderation: A Constitutional Framework

Giovanni De Gregorio

Abstract

Freedom of expression is one of the cornerstones on which democracy is based. This non-exhaustive statement firmly clashes with the troubling evolution of the algorithmic society where artificial intelligence technologies govern the flow of information online according to opaque technical standards established by social media platforms. These actors are usually neither accountable nor responsible for contents uploaded or generated by the users. Nevertheless, online content moderation affects users' fundamental rights and democratic values, especially since online platforms autonomously set standards for content removal on a global scale. Despite their crucial role in governing the flow of information online, social media platforms are not required to ensure transparency and explanation of their decision-making processes. Within this framework, this work aims to show how the liberal paradigm of protection of the right to free speech is no longer enough to protect democratic values in the digital environment, since the flow of information is actively organised by business interests, driven by profit-maximisation rather than democracy, transparency or accountability. Although the role of free speech is still paramount, it is necessary to enhance the positive dimension of this fundamental right by establishing new users' rights in online content moderation based on transparency and accountability of online platforms.

Summary: 1. Introduction. – 2. From the Free Marketplace of Ideas… – 3. …To the Law of the Platforms in Online Content Moderation. – 4. Users' Rights in Online Content Moderation: The *Status Quo*. – 5. Injecting Democratic Values in Online Content Moderation. 5.1 Notice System. 5.2 Decision-making. 5.3 Redress. – 6. Conclusion.

Keywords: Democracy – Online Platforms – Content Moderation – Freedom of Expression – Accountability – Transparency – Constitutionalism

1. Introduction

Freedom of expression is one of the cornerstones on which democracy is based.[449] This non-exhaustive statement acquires a specific relevance in the digital environment.[450] Indeed, in the last twenty years, the Internet has become one of the primary means to exercise rights and freedoms. Thanks to the possibility to access online contents ubiquitously, the digital environment plays a crucial role in promoting the sharing of opinion and ideas on a global scale.

Nevertheless, this flourishing democratic framework firmly clashes with the troubling evolution of the algorithmic society where social media platforms govern the flow of information online by implementing artificial intelligence technologies to moderate online content.[451] The relevance of this concern can be understood by observing that more than 2 billion of users are today governed

[449] Cass Sunstein, *Democracy and the Problem of Free Speech* (The Free Press 1995).

[450] Jack M. Balkin, 'Digital Speech and Democratic Culture: a Theory of Freedom of Expression for the Information Society' (2004) 79(1) New York University Law Review 1.

[451] Sarah T. Roberts, *Behind the Screen. Content Moderation in the Shadows of Social Media* (Yale University Press 2019); Kate Klonick, 'The New Governors: The People, Rules, and Processes Governing Online Speech' (2018) 131 Harvard Law Review 1598; Kyle Langvardt, 'Regulating Online Content Moderation' (2018) 106 The Georgetown Law Journal 1353.

by Facebook's community guidelines,[452] and YouTube decide how to host and distribute billions of hours of video each week.[453]

Although these online spaces positively affect fundamental rights by increasing the opportunities to exercise individuals' rights such as freedom of expression,[454] serious concerns deserve to be taken into account. When looking at the digital environment, it is possible to underline how an oligopoly of private entities organises transnationally online information by using automated technologies,[455] imposing their functional sovereignty.[456] The organisation of social networks' news feed or the results provided by a search engine are only some examples of the role of automated decision-making systems in online content moderation.

Moderation can be defined as 'the screening, evaluation, categorization, approval or removal/hiding of online content according to relevant communications and publishing policies. It seeks to support and enforce positive communications behaviour online, and to minimize aggression and anti-social behaviour'.[457] According to Grimmelman, content moderation is 'the governance mechanisms that structure participation in a community to facilitate cooperation and prevent abuse'.[458] This activity can be implemented before content is actually published (ie pre-moderation) or after publication (ie post-moderation). In particular, post-moderation is usually implemented as a reactive measure to assess noticed content and as a proactive tool to actively monitor published content.

Moreover, content moderation decisions can be entirely automated, made by humans or a mix of them. While the activities of pre-moderation like prioritisation, delisting and geo-blocking are usually automated, post-moderation is usually the result of a mix between automated and human resources.[459] As observed by Gillespie, 'moderation is not an ancillary aspect of what platforms do. It is essential, constitutional, definitional. Not only can platforms not survive without moderation, they are not platforms without it'.[460] The moderation of online content is an almost obligatory step for social media not only to manage removal requests but also to prevent that their digital spaces turn into hostile environments for users due to the spread for example, of incitement to hatred. Indeed, the interest of platforms is not just focused on facilitating the spread of opinions and ideas across the globe but establishing a digital environment where users feel free to share information and data that can feed commercial networks and channels and, especially, attract profits coming from advertising.[461] In other words, the activity of content moderation is performed to attract revenues by ensuring a healthy online community, protect the corporate image and

[452] Ben Popper, 'A Quarter of the World's Population Now Uses Facebook Every Month' (*The Verge*, 3 May 2017) <https://www.theverge.com/2017/5/3/15535216/facebook-q1-first-quarter-2017-earnings> accessed 2 August 2019.
[453] Jack Nicas, 'YouTube Tops 1 Billion Hours of Video a Day, on Pace to Eclipse TV' (*Wall Street Journal*, 27 February 2017) <https://www.wsj.com/articles/youtube-tops-1-billion-hours-of-video-a-day-on-pace-to-eclipse-tv-1488220851> accessed 2 August 2019.
[454] Henry Jenkins, Convergence Culture: Where Old and New Media Collide (New York University Press 2006); Yochai Benkler, The Wealth of Networks: How Social Production Transforms Markets and Freedom (Yale University Press 2006).
[455] Jack M. Balkin, 'Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation' (2018) 51 University of California Davis 1151, 1.
[456] Frank Pasquale, 'From Territorial to Functional Sovereignty: The Case of Amazon' Law & Political Economy Blog (6 December 2017) <https://lpeblog.org/2017/12/06/from-territorial-to- functional-sovereignty-the-case-of-amazon/> accessed 24 July 2019.
[457] Terry Flew and others, 'Internet Regulation as Media Policy: Rethinking the Question of Digital Communication Platform Governance' (2019) 10(1) Journal of Digital Media & Policy 33, 40.
[458] James Grimmelmann, 'The Virtues of Moderation' (2015) 17 Yale Journal of Law and Technology 42, 47.
[459] Sarah T. Roberts, 'Content Moderation', in Laurie A. Schintler and Connie L. McNeely (eds), *Encyclopedia of Big Data* (Springer 2017).
[460] Tarleton Gillespie, *Custodians of the Internet. Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media* (Yale University Press 2018), 21.
[461] Tarleton Gillespie, 'Regulation of and by Platforms' in Jean Burgess, Alice E. Marwick and Thomas Poell (eds), *The SAGE Handbook of Social Media* (Sage 2018), 254–78.

show commitments with ethic values. Within this business framework, users' data are the central product of online platforms under a logic of accumulation.[462]

Notwithstanding several social media exploit rhetoric statements advocating to represent a global community enhancing free speech transnationally, however, online platforms need to moderate content to protect their business interests by avoiding users' escape because of the dissemination of content like terrorism and hate. This 'content moderation paradox' explains why, on the one hand, social media commit to protecting free speech, while, on the other hand, they moderate content regulating their communities for business purposes. Therefore, one of the primary concerns is the compatibility between their private interests and public values.[463]

This business logic can also be appreciated by looking at the use of artificial intelligence systems to moderate online content. Platforms rely on automated technologies to cope with the amount of content loaded by users whose non-automated management would require enormous costs in terms of human, technological and financial resources. If, on the one hand, content moderation constitutes an important resource for social media, on the other hand, the use of technologies (*e.g.* machine learning) for moderating content on a global scale challenges the protection of freedom of expression in the digital environment that extends far beyond domestic boundaries.[464] The information uploaded by users is processed by automated systems that define (or at least suggest to human moderators) content that must be removed in a bunch of seconds according to non-transparent standards and without providing the user access to any remedy against a specific decision.

Despite the fundamental role of online platforms in establishing the standard of free speech and shaping democratic culture on a global scale,[465] the information provided by these companies about content moderation is opaque or lawless threatening the rule of law.[466] Online platforms are free to decide how to show and organise online content according to predictive analysis based on the processing of users' data.[467] In other words, although, at first glance, social media foster freedom of expression by empowering users to share their opinion and ideas cross-border, however, the high degree of opacity and inconsistency of content moderation frustrates democratic values. Content moderation does not only constitute an autonomous set of technical rules to ensure a peaceful digital environment but also contributes to defining the standard of protection of fundamental rights in the digital environment.

This situation leads to the 'mathematisation of the law' since the concept of legality is defined by a mere algorithmic calculation. The power of online platforms to shape the scope of protection of rights lies mostly in their ability to mathematically materialize abstract notions through digital means. Since artificial intelligence technologies are always becoming more pervasive in online content moderation, the opacity of these technologies raises legal (and ethical) concerns for democracy.[468] Individuals are increasingly surrounded by technical systems influencing their decisions without the possibility to understand or control this phenomenon.[469] In other words,

[462] Shoshana Zuboff, 'Big Other: Surveillance Capitalism and the Prospects of an Information Civilization' (2015) 30(1) Journal of Information Technology 75.
[463] José van Dijck and others, *The Platform Society: Public Values in a Connective World* (Oxford University Press 2018).
[464] Jack M. Balkin, 'Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation' (2018) UC Davis Law Review 1149; James Grimmelmann, 'Speech Engines' (2014) 98 Minnesota Law Review 868.
[465] Marvin Ammori, 'The "New" New York Times: Free Speech Lawyering in the Age of Google and Twitter' (2014) 127 Harvard Law Review 2259.
[466] Nicolas Suzor, *Lawless: The Secret Rules That Govern Our Digital Lives* (Cambridge University Press 2019).
[467] For the purposes of this work, the term 'online platforms' refers to providers hosting and organizing third-party content.
[468] Brent D. Mittlestadt and others, 'The Ethics of Algorithms: Mapping the Debate' (2016) 3(2) Big Data & Society 1.
[469] Paul Nemitz, 'Constitutional Democracy and Technology in the age of Artificial Intelligence' (2018) Royal Society Philosophical Transactions A.

notwithstanding the Internet has allowed users to access different types of information, the mediation of automated technologies leads users to participate in what Cohen defines a 'modulated democracy'.[470]

This situation is the result of a lack of transparency in online content moderation since algorithmic technologies are programmed according to the economic and ethical values of private entities. Users cannot still rely on any legal right vis-à-vis online platforms concerning content moderation. Moreover, platforms do not usually implement transparent procedures to explain to users how their content is managed or provide explanations when removing or blocking online content. If content moderation plays a crucial role in influencing the information flow in the digital environment, it is worth focusing on how it is possible to remedy this lack of transparency and accountability. This asymmetry between users and platforms leads to discuss whether the traditional liberal feature of the right to freedom of expression can ensure democratic values in the algorithm era. Democratic States are open environments for pluralism. The expression 'liberal democracy' evokes values and principles such as liberty, equality, transparency and accountability. On the contrary, the activity of online platforms is based on business interests, private procedures and pragmatic decision-making. Unlike online platforms, which have a responsibility rather than a duty to guarantee the respect of fundamental rights and freedoms, democratic States are required to safeguard these interests to protect the entire democratic system. Such duty also encompasses a positive obligation to protect individuals against acts committed by private persons or entities.[471] Indeed, without protecting equality, freedom of expression or assembly, it would not be possible to enjoy a democratic society. This consideration shows why fundamental rights and democracy are substantially intertwined.

Within this clash between democratic public values and non-democratic business interests, this work argues that the vertical and negative nature of freedom of expression is no longer enough to protect democratic values in the digital environment, since the flow of information is actively organised by business interests, driven by profit-maximisation rather than democracy, transparency or accountability. Therefore, the primary goal is to propose a set of new legal rights empowering users vis-à-vis online platforms and fostering democratic values in the digital environment.

In order to achieve this aim, the first part of this work analyses the shift from a liberal economic narrative based on the metaphor of the free marketplace of ideas to the rise of online platforms power in moderating content online by comparing the EU and US experience. This part allows understanding how the development of the information society has challenged the liberal paradigm of free speech. The second part focuses on the current *status quo* underlining the lack of general legal rights on which the users can rely vis-à-vis online platforms. In particular, this work focuses on the European Union ('EU' or 'Union') as an example of a potential shift from a liberal to a constitutional approach to the right to freedom of expression in the digital environment. In the light of the previous sections, the third part supports the introduction of users' rights in the process of content moderation and, especially, in the phases of notice, decision-making and redress.

2. From the Free Marketplace of Ideas…

The right to freedom of expression in modern and contemporary history has liberal roots. Like other civil and political liberties arisen at the end of the XIX century, the right to free speech is

[470] Julie E. Cohen, 'What Privacy Is For' (2013) 126 Harvard Law Review 1904.
[471] UN Human Rights Committee (HRC), 'General comment no. 31 [80], The nature of the general legal obligation imposed on States Parties to the Covenant', 26 May 2004 <https://www.refworld.org/docid/ 478b26ae2.html> accessed 7 October 2019.

based on the idea that liberties and freedoms can be ensured by limiting interferences coming from public actors.

It is not by chance that that one of the most suggestive legal metaphors in this field is that of the 'free market place of ideas'.[472] This expression was coined for the first time by Justice Douglas in *United States v Rumely*.[473] This liberalist belief can be contextualised in the classical theory of market balance to the field of ideas. Since individuals act rationally, they can choose the best products and services in a free market. As in a competitive market where the best products or services prevail, the same mechanism applies to the best information resulting from market balance.

However, the grounds of this liberal theory are deeper. In the seventeenth century, John Milton, opposing to the English Parliament's Press Ordinance, which had introduced a system of censorship to punish the promoters of ideas considered illegal, argued that freedom of expression should not be limited to allow the truth to prevail thanks to the free exchange of opinion.[474] Milton compares the truth to a streaming fountain whose water constitutes the flow of information saving men from prejudice. According to this perspective, it is necessary to avoid any interference with the flow of information to lead men to the highest level of knowledge. Two centuries later, John Stuart Mill similarly shared a liberal approach to freedom of expression.[475] According to Mill, even falsehood could contribute to reaching the truth.[476] Otherwise, censoring falsehood would make meaningless the comparison between ideas and opinions with the risk of dogmatising the current truth.[477]

The scope of these liberal ideas opposing against public actors' interferences also emerged in the twentieth century from the Justice Holmes' dissenting opinion in *Abrams v United States* of 1919.[478] This dissenting opinion can still be considered the constitutional essence of freedom of expression in the United States as enshrined in the First Amendment.[479] In particular, the case concerned the distribution of leaflets calling for ammunition factories to strike to express a clear

[472] Oreste Pollicino, 'Fake news, Internet and Metaphors' (2017) 1(1) Rivista di diritto dei media 23; Daniel E. Ho and Frederik Schauer, 'Testing the Marketplace of Ideas' (2015) 90 New York University Law Review 1161; Eugene Volokh, 'In Defense of the Market Place of Ideas / Search for Truth as a Theory of Free Speech Protection' (2011) 97(3) Virginia Law Review 591; Joseph Blocher, 'Institutions in the Marketplace of Ideas' (2008) 57(4) Duke Law Journal 820; Paul H. Brietzke, 'How and Why the Marketplace of Ideas Fails' (1997) 31(3) Valparaiso University Law Review 951; Alvin I. Goldman and James C. Cox, *Speech, Truth, and the Free Market for Ideas* (Cambridge University Press 1996); Ronald Coase, 'Markets for Goods and Market for Ideas' (1974) 64(2) American Economic Review 1974.

[473] *United States v Rumely* [1953] 345 U.S. 41. 'Of necessity I come then to the constitutional questions. Respondent represents a segment of the American press. Some may like what his group publishes; others may disapprove. These tracts may be the essence of wisdom to some; to others their point of view and philosophy may be anathema. To some ears their words may be harsh and repulsive; to others they may carry the hope of the future. We have here a publisher who through books and pamphlets seeks to reach the minds and hearts of the American people. He is different in some respects from other publishers. But the differences are minor. Like the publishers of newspapers, magazines, or books, this publisher bids for the minds of men in the market place of ideas'.

[474] John Milton, *Aeropagitica* (1644). According to Milton: 'So Truth be in the field, we do injuriously, by licensing and prohibiting, to misdoubt her strength. Let her and Falsehood grapple; who ever knew Truth put to the worse, in a free and open encounter?'.

[475] John S. Mill, *On Liberty* (1859).

[476] Ibid, 'First, if any opinion is compelled to silence, that opinion may, for aught we can certainly know, be true. To deny this is to assume our own infallibility'.

[477] Ibid, 'Thirdly, even if the received opinion be not only true, but the whole truth; unless it is suffered to be, and actually is, vigorously and earnestly contested, it will, by most of those who receive it, be held in the manner of a prejudice, with little comprehension or feeling of its rational grounds. And not only this, but, fourthly, the meaning of the doctrine itself will be in danger of being lost, or enfeebled, and deprived of its vital effect on the character and conduct: the dogma becoming a mere formal profession, inefficacious for good, but cumbering the ground, and preventing the growth of any real and heartfelt conviction, from reason or personal experience'.

[478] *Abrams v United States* [1919] 250 U.S. 616.

[479] Ibid, 'Persecution for the expression of opinions seems to me perfectly logical. If you have no doubt of your premises or your power and want a certain result with all your heart you naturally express your wishes in law and sweep away all opposition [...] But when men have realized that time has upset many fighting faiths, they may come to believe even more than they believe the very foundations of their own conduct that the ultimate good desired is better reached by free trade in ideas. […] The best test of truth is the power of the thought to get itself accepted in the competition of the market, and that truth is the only ground upon which their wishes safely can be carried out'.

message of resistance against the US military intervention in Russia. According to Justice Holmes, although men try to support their positions by criticising opposing ideas, they must not be persuaded that their opinions are certain. Only the free exchange of ideas can confirm the accuracy of each position.[480]

This liberal approach has also been expressed, more recently, in the framework of the digital environment, at least in two landmark decisions of the US Supreme Court. In 1997, in *Reno v ACLU*,[481] the Supreme Court ruled that the provisions of the Communications Decency Act concerning the criminalisation of obscene or indecent materials to any person under 18 was unconstitutional. As observed by the Supreme Court, unlike traditional media outlets, 'the risk of encountering indecent material by accident is remote because a series of affirmative steps is required to access specific material'.[31] According to Justice Stevens, the Internet plays the role of a 'new marketplace of ideas' observing that 'the interest in encouraging freedom of expression in a democratic society outweighs any theoretical but unproven benefit of censorship'.[482] In the aftermath of the Internet, the optimist interpretation of the US Supreme Court was also reflected by the theories of those scholars who considered the Internet as a new place outside the interference of any public actor.[483] However, this approach has been generally questioned by whom had already underlined an increasing discretion of new actors in the cyberspace,[484] deriving also from the private enforcement of public policies online.[485]

Despite the passing of years and opposing positions, this liberal approach has been reiterated more recently in *Packingham v North Carolina*.[486] The case involved a statute banning registered sex offenders from accessing social networking services to avoid any contact with minors. The US Supreme Court placed the Internet and social media on the same layer of public places where First Amendment enjoy a broad scope of protection. In the words of Justice Kennedy: 'It is cyberspace – the "vast democratic forums of the Internet" in general, and social media in particular'.[487] Therefore, social media enjoy a safe constitutional area of protection under the First Amendment, which in the last twenty years, has constituted a fundamental ban on any regulatory attempt to regulate speech online.[488]

Nevertheless, it would be enough just to cross the Atlantic to understand how this general trust for a vertical paradigm of free speech is not shared worldwide by other democracies, especially when the right to freedom of expression is framed in the digital environment. While, in the US, the Internet and social media still benefit from the frame coming from the traditional liberal metaphor of the free marketplace of ideas as a safeguard for democracy, in Europe, the protection of freedom of expression online does not enjoy the same degree of protection.[489] In the European framework, the right to freedom of expression is subject to a multilevel balancing with other rights enshrined in the Charter of Fundamental Rights of the European Union ('Charter'),[490] the

[480] See, in particular, Sheldon Novick, *Honorable Justice* (Laurel 1990).

[481] *Reno v American Civil Liberties Union* [1997] 521 U.S. 844.
[482] Ibid.
[483] John P Barlow, 'A Declaration of Independence of the Cyberspace' (Electronic Frontier Foundation, 1996) <www.eff.org/cyberspace-independence> accessed 2 July 2019; David R Johnson and David Post, 'Law and Borders: The Rise of Law in Cyberspace' (1996) 48(5) Stanford Law Review 1371.
[484] Lawrence Lessig, *Code 2.0* (Basic Books 2006).
[485] Jack Goldsmith and Tim Wu, *Who Controls the Internet?* (Oxford University Press 2006); Joel R Reidenberg, 'States and Internet Enforcement' (2004) 1 University of Ottawa Law & Techonology Journal 213.
[486] *Packingham v North Carolina* [2017] 582 U.S. ___.
[487] Ibid.
[488] See, for example, *Reno* (n 33); *Ashcroft v Free Speech Coalition* [2002] 535 U.S. 234; *Aschroft v American Civil Liberties Union* [2002] 535 US 564.
[489] Oreste Pollicino and Marco Bassini, 'Free Speech, Defamation and the Limits to Freedom of Expression in the EU: A Comparative Analysis', in Andrej Savin and Jan Trzaskowski (eds), *Research Handbook on EU Internet Law* (Edward Elgar 2014).
[490] Charter of Fundamental Rights of the European Union [2012] OJ C326/12. Art 52.

European Convention of Human Rights ('Convention'),[491] and national constitutions. In particular, unlike the US Supreme Court, the Strasbourg Court has shown a more cautious approach to the protection of the right to freedom of expression in the digital environment, perceived more like a risk rather than an opportunity for the flourishing of democratic values.[492] Such a cautious approach in Europe does not only aim to balance different constitutional interests but also to avoid that granting absolute protection to one right could lead to the destruction of other fundamental rights undermining *de facto* their constitutional relevance.[493]

This non-exhaustive framework is one of the most important reasons to understand why the EU has paved the way towards the regulation of online content moderation. Despite the difference in the protection of the right to freedom of expression in the EU and the US, this fundamental right is still the pre-requisite to live in a democratic society. However, in the digital environment, the protection of the right to freedom of expression is no longer a matter of quantity but quality because of the crucial role of online platforms in determining the standard of protection of free speech and other fundamental rights on a global scale. In other words, the primary challenge for democracies is no longer protecting extensively freedom of expression by granting access to new digital channels and avoiding public actors' interferences, but ensuring that users can effectively enjoy their rights and freedoms in a democratic digital environment.

3. …To the Law of the Platforms in Online Content Moderation

At the World Summit on the Information Society, Lessig underlined the significant potentialities afforded by the digital environment: '[f]or the first time in a millennium, we have a technology to equalize the opportunity that people have to access and participate in the construction of knowledge and culture, regardless of their geographic placing'.[494] Likewise, Shapiro stated: 'Hierarchies are coming undone. Gatekeepers are being bypassed. Power is devolving down to "end users" [...] No one is in control except you'.[495] Unlike in the atomic marketplace of ideas, information sources have spread online. The new online communication channels have enabled users to potentially reach a global audience without relying any longer on the traditional channels of communications where editorial decisions are in the hand of publishers like newspapers and televisions.[496]

Although the rise of information pluralism should generally be welcomed for the development and maintenance of a democratic environment, the characteristics of the information flow online raise serious concerns in terms of pluralism for at least two reasons. First, from a quantitative perspective, in the last twenty years, a high degree of concentration of the online platforms' market has characterised the digital environment. As foreseen by Zittrain,[497] the characteristics of the information society have led to the creation of monopolies,[498] linked to the platformization of the Internet, [499] which Srnicek would call the era of 'platform capitalism'.[500] This market concentration empowers a limited number of platforms to set the conditions on which vast amount

[491] European Convention on Human Rights [1950]. Art 10(2).
[492] Oreste Pollicino, 'Judicial Protection of Fundamental Rights in the Transition from the World of Atoms to the World of Bits: The Case of Freedom of Speech' (2019) 25(2) European Law Journal 155.
[493] Charter (n 42), Art 54; Convention (n 43), Art 17.
[494] Lawrence Lessig, 'An Information Society: Free or Feudal' (2004) World Summit on the Information Society (WSIS), <http://www.itu.int/wsis/docs/pc2/visionaries/lessig.pdf> accessed 4 August 2019.
[495] Andrew L. Shapiro, *The Control Revolution: How the Internet is Putting Individuals in Charge and Changing the World we Know* (Public Affairs 1999), 11, 30.
[496] Jack M. Balkin, 'Old-School/New-School Speech Regulation' (2014) 127 Harvard Law Review 2296.
[497] Jonanthan Zittrain, *The Future of the Internet and How to Stop It* (Yale University Press 2008).
[498] Robin Mansell and Michele Javary, 'Emerging Internet Oligopolies: A Political Economy Analysis' in Arthur S Miller, Warren J Samuels (eds), *An Institutionalist Approach to Public Utilities Regulation* (Michigan State University Press 2002)
[499] Anne Helmond, 'The Platformization of the Web: Making Web Data Platform Ready' (2015) 1(2) Social Media + Society 1.
[500] Nick Snricek, *Platform Capitalism* (Polity Press 2016).

of content and data flow online. Notwithstanding, at first glance, the digital environment has empowered users to access new channels to share ideas and access sources of information, however, the aforementioned digital convergence dangerously affects media pluralism.

Second, from a qualitative standpoint, pluralism is based on different manifestations of thinking and promotes heterogeneous ideas. Instead, in the digital environment, the use of artificial intelligence for online content moderation mitigates this positive effect. The organisation of content aims to engage users based on their data and preferences, leading to the polarisation of the debate due to the creation of 'filter bubbles' or 'information cocoons'.[501] The personalization of online content leads to the creation of echo chambers where each user is isolated and marginalised from opposing positions as resulting from a mere algorithmic calculation. In other words, users are encouraged to interact only with information inside the area of their preferences. This situation leads to the debasement of information pluralism in the digital environment. Within this framework, public actors are no longer the only source of concern in the (digital) free marketplace of ideas. Instead of a democratic and decentralised society as defined at the end of the last century, an oligopoly of private entities has emerged, controlling information and determining how people exchange it.[502] As such, the platform-based regulation of the internet has prevailed over the community-based model. Furthermore, the lack of transparency and accountability in online content moderation processes frustrate the quality of information pluralism. The lack of information pluralism can be considered one of the primary failures of the digital marketplace of ideas.[503] As a result, these considerations would explain why considering public actors as the only threat to freedom of expression online could seem anachronistic today. Indeed, a further challenge raised by the information society concerns how to address the discretion of private actors freely influencing the limits of freedom of expression on a global scale without any publicity guarantee.

This situation can be primarily considered the result of the system of online intermediaries' liability based on a liberal regulatory approach adopted by the US and EU at the end of the last century. When the US Congress passed Section 230 of the Communication Decency Act ('CDA') in 1996,[504] the primary aim was to encourage the sharing of free expression and development of the digital environment.[505] In order to achieve this objective, the choice was to exempt computer services from liability for merely conveying third-party content. Before the adoption of the CDA, some cases had already made clear how online intermediaries would have been subject to a broad and unpredictable range of cases concerning their liability for editing third-party content.[506] Since this risk would have slowed down the development of new digital services in the aftermath of the Internet, online intermediaries have been encouraged to grow and develop their business under the protection of the Good Samaritan rule.[507] Similarly, the Digital Millennium Copyright Act ('DMCA') introduced in 1997 allows online intermediaries not to be held liable for hosting unauthorised copyright works.[508] Nevertheless, unlike the CDA, the DMCA does not provide an absolute exemption but shield online intermediaries from liability according to certain conditions.[509]

---

[501] Eli Pariser, *The Filter Bubble: What the Internet is Hiding from You* (Viking 2011); Cass Sunstein, *Republic.com 2.0* (Princeton University Press 2007).
[502] Martin Moore and Damian Tambini (eds), *Digital Dominance. The Power of Google, Amazon, Facebook, and Apple* (Oxford University Press 2018).
[503] Annemarie Bridy, 'Remediating Social Media: A Layer-Conscious Approach' (2018) 24 Boston University Journal of Science & Technology Law 193.
[504] Communication Decency Act [1996], Section 230.
[505] Klonick (n 3).
[506] Cubby, Inc. v CompuServe Inc. [1991] 776 F. Supp. 135 (S.D.N.Y.); Stratton Oakmont, Inc. v Prodigy Services Co. [1995] WL 323710 (N.Y. Sup. Ct. May 24).
[507] Zeran v Am. Online, Inc. [1997] 129 F.3d 327, 330 (4th Cir.). Davis S. Ardia, 'Free Speech Savior or Shield for Scoundrels: An Empirical Study of Intermediary Immunity under Section 230 of the Communications Decency Act' (2010) 43 Loyola of Los Angeles Law Review 373.
[508] Digital Millenium Copyright Act [1997].
[509] Ibid, 17 U.S. Code § 512(c).

Likewise, in the EU, the e-Commerce Directive, adopted in 2000, exempts hosting providers (*e.g*. social network or search engine) from liability for third-party content, provided that they remove or disable online content once they become aware of its unlawful nature.[510] The online platforms' awareness, which can result, for example, by the notice submitted by public actors or users, triggers the responsibility of online platforms to remove content. Therefore, even within the EU framework, online platforms are not liable for third-party content, provided that they perform their activities in a passive way and comply with the conditions applying to the exemption of liability.[511] Several scholars have underlined how this political choice has led platforms to exploit the legal framework to their advantage. According to Pasquale, online platforms try to avoid regulatory burdens by relying on the protection recognised by the First Amendment, while, at the same time, they claim immunities as passive conduits for third-party content.[512] Likewise, Citron and Norton observe how social media 'not only are free from First Amendment concerns as private actors, they are also statutorily immunized from liability for publishing content created by others as well as for removing that content'.[513] As Tushnet underlined, Section 230 'allows Internet intermediaries to have their free speech and everyone else's too'.[514]

The immunity granted by these laws leads online platforms to freely choose which values they want to protect and promote, no matter if democratic or anti-democratic and authoritarian. As observed by Roberts, 'videos and other material have only one type of value to the platform, measured by their ability to either attract users and direct them to advertisers or to repel them and deny advertisers their connection to the user'.[515] Since online platforms are private businesses, they would likely focus on minimising economic risks rather than ensuring a fair balance between fundamental rights in the digital environment. In other words, the system of online intermediaries' liability indirectly entrusts online platforms with the role of moderating content based on a standard of protection of free speech influenced by business purposes.

The scope of online platforms' power can be better understood by focusing on how these actors, firstly, set and, then, enforce their internal rules of moderation after balancing conflicting interests. As noted by Belli, Francisco and Zingales,[516] online platforms usually rely on procedures for moderating content they have autonomously established in their Terms of Services ('ToS') or according to internal guidelines. As a result, in the lack of any regulation, users' agreements and internal guidelines set discretionary the standard of protection of free speech playing the role of the law in these digital spaces on a global scale.[517]

A further characteristic concerns the enforcement of ToS rules without any intermediation of public actors. For instance, the removal of content does not require any public order, but online platforms can autonomously perform this activity by virtue of the control over their digital spaces. Indeed,

[510] 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') [2000] OJ L 178/1. See Art 14.
[511] Ibid, Recital 42.
[512] Frank Pasquale, 'Platform Neutrality: Enhancing Freedom of Expression in Spheres of Private Power' (2016) 17 Theoretical Inquiries in Law 487, 497-503.

[513] Danielle Keats Citron and Helen L. Norton, 'Intermediaries and Hate Speech: Fostering Digital Citizenship for our Information Age' (2011) 91 Boston University Law Review 1436, 1439.
[514] Rebecca Tushnet, 'Power Without Responsibility: Intermediaries and the First Amendment' (2008) 76 The George Washington Law Review 986, 1002.
[515] Sarah T. Roberts, 'Digital detritus: "Error" and the logic of opacity in social media content moderation, (2018) 23(3) First Monday <https://firstmonday.org/ojs/index.php/fm/rt/printerFriendly/8283/6649> accessed 28 July 2019.
[516] Luca Belli, Pedro A Francisco and Nicolo Zingales, 'Law of the Land or Law of the Platform? Beware of the Privatisation of Regulation and Police' in Luca Belli and Nicolo Zingales (eds), *How Platforms are Regulated and How They Regulate Us* (FGV Direito Rio 2017).
[517] Luca Belli and Jamila Venturini, 'Private ordering and the rise of terms of service as cyber-regulation' (2016) 5(4) Internet Policy Review <https://policyreview.info/node/441/pdf> accessed 10 March 2019; Edoardo Celeste, 'Terms of service and bills of rights: new mechanisms of constitutionalisation in the social media environment?' (2018) International Review of Law, Computers & Technology <https://www.tandfonline.com/doi/full/10.1080/13600869.2018.1475898?scroll=top&needAccess=true> accessed on 5 April 2019.

online platforms can enforce the rules established by their ToS directly without the need to rely on a public mechanism such as a judicial order. Here, it would be possible to argue that the code plays the role of the law, allowing platforms to directly enforce their rules.[518]

Besides, when moderating content, online platforms assess whether content is unlawful by striking a balance between fundamental rights,[519] a function mirroring that of the judiciary. Recently, the Facebook's proposal to create an independent governance and oversight committee to make decisions about the kinds of content users could post on the site has led to label this new body as a 'Supreme Court'.[520]

Because of these activities, users are subject to the exercise of legitimate authority which, in the digital environment, seems to be exercised by online platforms through a mix of private law and automated technologies (*i.e.* the law of the platforms).[521]

Within this framework, the lack of any users' rights or remedy leads online platforms to exercise the same discretion of an absolute power over its community. Social media usually provide ToS and community guidelines where they explain users the acceptable conducts and content, creating 'a complex interplay between users and platforms, humans and algorithms, and the social norms and regulatory structures of social media'.[522] However, these community rules do not necessarily represent the reality of content moderation. Facebook, for example, relies on internal guidelines which users cannot access and whose drafting process is unknown.[523] According to Klonick, Facebook's content moderation is 'largely developed by American lawyers trained and acculturated in American free-speech norms, and it seems that this cultural background has affected their thinking'.[524]

Moreover, from a technical perspective, the opacity of content moderation derives also from the implementation of machine learning techniques subject to the 'black box' effect.[525] On the one hand, algorithms can be considered as technical instruments facilitating a platform's various functionalities, such as the organisation of online content. Nevertheless, on the other hand, such technologies can constitute opaque self-executing rules, obviating any human control with troubling consequences for democratic values such as transparency and accountability.

Furthermore, it is worth observing that this situation is not only the result of the complexity of content moderation systems but also of a 'logic of opacity'. [526] Platforms are interested in pursuing their 'depoliticization' to escape from their social responsibilities coming from their key social functions. As argued by Roberts, 'yet the process is obscured by a social media landscape that tacitly, if not explicitly, trades on notions of free circulation of self-expression, on the one hand, and a purported neutrality, on the other, that deny the inherent gatekeeping baked in at the platform level by both its function as an advertising marketplace and the systems of review and deletion that have, until recently, been invisible to or otherwise largely unnoticed by most users'.[527]

---

[518] Lessig (n 36).
[519] Marco Bassini, 'Private Enforcement of Fundamental Rights' (2019) 25(2) European Law Journal 198.
[520] Evelyn Douek, 'Facebook's New 'Supreme Court' Could Revolutionize Online Speech' Lawfare (19 November 2018) <https://www.lawfareblog.com/facebooks-new-supreme-court-could-revolutionize-online-speech> accessed 26 July 2019.
[521] Giovanni De Gregorio, 'From Constitutional Freedoms to Power: Protecting Fundamental Rights in the Algorithmic Society' (2019) 11(2) European Journal of Legal Studies 66.
[522] Kate Crawford and Tarleton Gillespie, 'What is a Flag for? Social Media Reporting Tools and the Vocabulary of Complaint' (2016) 18 New Media & Society 410, 411.
[523] Max Fisher, 'Inside Facebook's Secret Rulebook for Global Political Speech' (*New York Times*, 27 December 2018).
[524] Klonick (n 3), 1622.
[525] Frank Pasquale, The Black Box Society. The Secret Algorithms that Control Money and Information (Harvard University Press 2015).
[526] Roberts (n 67).
[527] Ibid.

This logic is also the result of how platforms manage the relationship with their community. Although, at first glance, online platforms usually rely on a narrative promoting a global and safe community, their approach is authoritarian rather than democratic. As Radin explains, businesses exploit contracts to overrule safeguards protecting the rights of the parties.[528] In the case of ToS, Zuboff describes this self-regulatory agreement as 'form of unilateral declaration that most closely resembles the social relations of a pre-modern absolutist authority'.[529] Similarly, according to MacKinnon, online platforms adopt a 'Hobbesian approach to governance' where users consent to give up fundamental rights in exchange for services.[530] In other words, this 'new social contract' leads users in a status of *subjectionis* vis-à-vis online platforms.

The underlined framework shows the rise of a private order whose characteristics do not mirror democratic values but are closer to absolute power. In particular, this phenomenon cannot be defined as the rise of a 'private constitutional order' since neither the separation of power nor the protection of rights is granted in this system, so that some authors have referred to this phenomenon as a return to feudalism,[531] or to the *Ancien Régime*.[532]

4. Users' Right in Online Content Moderation: The *Status Quo*

Within this troubling framework for democratic values, users are subject to a high degree of opacity in the social media environment and, even more importantly, they cannot generally rely on any legal right concerning the moderation of their content. In other words, as observed by Myers West, 'they are exactly the kinds of users who make up the kind of "town square," "global village," or "community" that these platforms themselves say they seek to cultivate—but current content moderation systems do not give them much opportunity to participate or grow as citizens of these spaces'.[533]

From an international perspective, both the Manila principles on intermediary liability and the IGF Dynamic Coalition on Platform Responsibility propose an approach towards the proceduralisation of content moderation.[534] Similarly, the Santa Clara principles on Transparency and Accountability in Content Moderation try to suggest some due process safeguards regarding how content moderation should be performed and what rights users can rely on in the context of this process.[535] Article 19 has proposed the creation of social media councils based on a self-regulatory and multi-stakeholder system of accountability for content moderation complying with international human rights' standards.[536] Similarly, Facebook has published a 'Draft Charter' of its 40-person oversight board explaining the commitments of the board such as transparency and motivation of the decisions.[537]

---

[528] Margaret J. Radin, *Boilerplate The Fine Print, Vanishing Rights, and the Rule of Law* (Princeton University Press 2013).
[529] Zuboff (n 14), 83.
[530] Rebecca Mackinnon, *Consent of the Networked: The Worldwide Struggle for Internet Freedom* (Basic Books 2013).
[531] James Grimmelmann, 'Virtual World Feudalism', (2009) 118 Yale Law Journal Pocket Part 126.
[532] Luca Belli and Jamila Venturini, 'Private Ordering and the Rise of Terms of Service as Cyber-Regulation' (2016) 5(4) Internet Policy Review <https://policyreview.info/ node/441/pdf> accessed 16 June 2019.
[533] Sarah Myers West, 'Censored, Suspended, Shadowbanned: User Interpretations of Content Moderation on Social Media Platforms' (2018) 20(11) New Media & Society 4380.
[534] Manila Principles on Intermediary Liability (2017) and the DCPR Best Practices on Platforms' Implementation on the Right to Effective Remedy" https://www.intgovforum.org/multilingual/index.php?q=filedepot_download/4905/1550
[535] Santa Clara Principles on Transparency and Accountability in Content Moderation (2018).
[536] Article 19, 'The Social Media Councils: Consultation Paper' (2019) <https://www.article19.org/wp-content/uploads/2019/06/A19-SMC-Consultation-paper-2019-v05.pdf> accessed 8 October 2019.
[537] Facebook, 'Draft Charter: An Oversight Board for Content Decision' (2019) <https://fbnewsroomus.files.wordpress.com/2019/01/draft-charter-oversight-board-for-content-decisions-1.pdf> accessed 27 July 2019. See Evelyn Douek and Kate Klonick, 'Facebook Releases an Update on Its Oversight Board: Many Questions, Few Answers' Lawfare (27 June 2019) <https://www.lawfareblog.com/facebook-releases-update-its-oversight-board-many-questions-few-answers> accessed 28 July 2019; Evelyn Douek, 'Facebook's "Oversight Board:" Move Fast with Stable Infrastructure and Humility' (2019) 21(1) North Carolina Journal of Law and Technology (forthcoming).

However, despite the relevance of this proposal, the lack of any binding force of this system leaves online platforms free to decide whether to participate in this mechanism or formally comply with these standards while maintaining their internal rules of procedures. At the same time, the remarkable interest of the UN Special Rapporteur for Freedom of Expression, David Kaye, underlines the increasing pressure on private actors to comply with international human rights law when moderating online content.[538] According to the Special Rapporteur, since social media exercise regulatory functions in the digital environment, these private actors should refer to the existing international human rights law regime when setting their standard for content moderation.[539] Indeed, the international human rights law would allow platforms to apply a universal reference in their activities of content moderation.

Nevertheless, as already underlined, since online platforms are private actors, they are not obliged to respect human rights since international human rights law vertically binds only State actors with the result that the governance of online platforms is based on fragmented national and regional laws as well as soft-regulatory efforts.[540] The same consideration extends to fundamental rights since constitutional provisions bind just public actors to respect them even if there could be some cases where fundamental rights horizontally apply in the relationship between private actors.[541]

Despite the role of self-regulation and corporate social responsibility in building a shared global framework which could overcome any regulatory vacuum,[542] the remedies voluntarily provided by online platforms are highly fragmented and left to their discretion.[543] Moreover, the differences between (public) community guidelines and (private) internal policy as well as the opacity about the use of automated systems in content moderation create a grey area of cases where removal or maintenance of content are decided outside any democratic control.
While, in the US, the legal framework has not changed in the last twenty years, apart from the recent amendments introduced to Section 230 CDA,[544] the Union has started to pave the way towards a new regulatory season of online content moderation.[545] One of the EU objectives is to ensure that online platforms 'protect core values' and increase 'transparency and fairness for maintaining user trust and safeguarding innovation'.[546] According to the Commission, since online platforms give access to information and content to society, their role implies 'wider responsibility'.[547]

[538] David Kaye, *Speech Police: The Global Struggle to Govern the Internet* (Columbia Global Reports 2019).
[539] Report of the Special Rapporteur to the Human Rights Council on online content regulation, A/HRC/38/35 (2018); See, also, Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, A/73/348 (2018); Guiding Principles on Business and Human Rights (2011).
[540] Jennifer Grygiel and Nina Brown, 'Are Social Media Companies Motivated to Be Good Corporate Citizens? Examination of the Connection Between Corporate Social Responsibility and Social Media Safety' (2019) 43 Telecommunications Policy 445.

[541] Some constitutions around the world (e.g. South Africa) horizontally extends the application of fundamental rights in the relationship between private actors. In other case, horizontal application is not the result of a direct constitutional provision but the result of judicial interpretation.
[542] Rolf H. Weber, 'Corporate Social Responsibility as a Gap-Filling Instrument', in Andrew P. Newell (ed.). *Corporate Social Responsibility: Challenges, Benefits and Impact on Business* (Nova 2014), 87.
[543] IGF Dynamic Coalition, 'Best Practices on Platforms' Implementation of the Right to an Effective Remedy' (2018) <https://www.intgovforum.org/multilingual/content/dcpr-best-practices-on-due-process-safeguards-regarding-online-platforms'-implementation-of> accessed 7 August 2019. See, also, Myers (n 85).
[544] See the Stop Enabling Sex Traffickers Act (SESTA) and the Allow States and Victims to Fight Online Sex Trafficking Act (FOSTA) adopted in 2018.
[545] Oreste Pollicino and Giovanni De Gregorio, 'A Constitutional Change of Heart: ISP Liability and Artificial Intelligence in the Digital Single Market' (2019) 18(1) Global Community Yearbook of International Law and Jurisprudence 2018 237.

[546] Commission, 'Online Platforms and the Digital Single Market Opportunities and Challenges for Europe' COM(2016) 288 final.
[547] Ibid.

Within this framework, it is necessary to mention at least two pieces of legislation: the Directive on Copyright in the Digital Single Market ('Copyright Directive'),[548] and the Regulation on tackling the dissemination of terrorist content online ('Regulation on Terrorist Content') voted by the European Parliament on April 2019.[549] These measures constitute a first turning point in online content moderation, requiring online platforms to establish transparent and accountable mechanisms.

The Copyright Directive is the only legal instrument at the EU level introducing a special regime derogating the system established by the e-Commerce Directive for online platforms' liability.[550] Without focusing on this system of liability applying to online content-sharing service providers when hosting copyright-protected content without the prior authorisation of rightholders,[551] it is interesting to look at the new safeguards introduced by the Copyright Directive.

First, online platforms are required to provide rightholders, at their request with adequate information on the functioning of their practices with regard to the cooperation and, where licensing agreements are concluded between service providers and rightholders, platforms are required to provide information on the use of content covered by the agreements.[552] Secondly, online content-sharing service providers have to put in place an effective and expeditious complaint and redress mechanism which users can access to in the event of disputes over the disabling of access to, or the removal of, works or other subject-matter uploaded by them.[553] The Copyright Directive also requires that online platforms handle these complaints without undue delay, and subject their decisions to remove content to human review.[554] Thirdly, Member States have to ensure the availability of impartial out-of-court redress mechanisms for the settlement of disputes which should not deprive users with legal protection afforded by national law.[555]
Similarly, the proposed Regulation on Terrorist Content, which aims to establish a clear and harmonised legal framework to prevent the misuse of hosting services (online platforms) for the dissemination of terrorist content online, is another interesting example of users' rights in online content moderation.[556] Indeed, the Regulation on Terrorist Content provides not only reacting measures, for instance, by requiring online platforms to remove content within one hour from the notice of the competent authority, but also proactive ones to mitigate the risks of exposure to terrorist contents. Among procedural safeguards, transparency is fostered by requiring online platforms and competent authorities to disclose information about their activities concerning

[548] 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 [2019] OJ L 130/92.
[549] European Parliament legislative resolution of 17 April 2019 on the proposal for a regulation of the European Parliament and of the Council on preventing the dissemination of terrorist content online (COM(2018)0640 – C8-0405/2018 – 2018/0331(COD)).
[550] Copyright Directive (n 100), Art 17.
[551] See Giancarlo Frosio and Sunimal Mendis, 'Monitoring and Filtering: European Reform or Global Trend?' in Giancarlo Frosio (ed.), *The Oxford Handbook of Online Intermediary Liability* (OUP, forthcoming 2019); João Quintais 'The New Copyright in the Digital Single Market Directive: A Critical Look' (2019) European Intellectual Property Review (forthcoming); Martin Husovec, 'How Europe Wants to Redefine Global Online Copyright Enforcement' in Tatiana E. Synodinou (ed.), *Pluralism or Universalism in International Copyright Law* (Kluwer, forthcoming 2019); Aleksandra Kuczerawy, 'EU Proposal for a Directive on Copyright in the Digital Single Market: Compatibility of Article 13 with the EU Intermediary Liability Regime', in Bilyana Petkova and Tuomas Ojanen (eds), *Fundamental Rights Protection Online: The Future Regulation of Intermediaries* (forthcoming 2019).
[552] Ibid, Art 17(8).
[553] Ibid, Art 17(9).
[554] Ibid.
[555] Ibid.
[556] Regulation on Terrorist Content (n 101), Art 1. See Joris van Hoboken, 'The Proposed EU Terrorism Content Regulation: Analysis and Recommendations with Respect to Freedom of Expression Implications' (2019) Transatlantic Working Group on Content Moderation Online and Freedom of Expression <https://www.ivir.nl/publicaties/download/TERREG_FoE-ANALYSIS.pdf> accessed 29 July 2019; Aleksandra Kuczerawy, 'The Proposed Regulation on Preventing the Dissemination of Terrorist Content Online: Safeguards and Risks For Freedom Of Expression' (2018) CITIP paper for the Center for Democracy and Technology, <https://cdt.org/files/2018/12/Regulation-on-preventing-the-dissemination-of-terrorist-content-online-v3.pdf> accessed 29 July 2019; Joan Barata, 'New EU Proposal on the Prevention of Terrorist Content Online' (2018) CIS Stanford Law <https://cyberlaw.stanford.edu/files/ publication/files/2018.10.11.Comment.Terrorism.pdf> accessed 26 July 2019.

terrorist content.[557] For instance, hosting providers should explain in their ToS their policy to prevent the dissemination of terrorist content, 'including, where applicable, a meaningful explanation of the functioning of specific measures' as well as provide publicly available annual transparency reports regarding the measures implemented to tackle the dissemination of terrorist content.[558] Besides, where the use of automated tools is involved in content moderation, online platforms are required to provide human oversight and verifications to ensure 'the right to freedom of expression and freedom to receive and impart information and ideas in an open and democratic society'.[559]

Furthermore, the Regulation on Terrorist Content obliges online platforms to ensure content providers, whose content has been removed or access to it disabled, access to complaint mechanisms requesting reinstatement of the content.[560] In this case, online platforms are required to comply with a specific procedure. Within two weeks of the receipt of the complaint, online platforms should provide an explanation in cases where they decide not to reinstate the content.[561] Moreover, online platforms are also required to provide comprehensive and concise information on the removal or disabling of access to terrorist content, the possibilities to contest the decision and a copy of the removal order issued by the competent authority,[562] except when the competent authority decides based on objective evidence and considering the proportionality and necessity of such decision not to disclose information for reasons of public interest such as security.[563]

These two measures are part of a broader strategy of the Union to foster accountability and transparency in online content moderation. In 2016, the Commission issued the Code of Conduct on Countering Illegal Hate Speech Online and, in 2018, the Code of Practice on Online Disinformation resulting from the Communication on Tackling Online Disinformation and, especially, the Communication on tackling illegal content online,[564] then implemented in the Recommendation on measures to effectively tackle illegal content online ('Recommendation').[565] These soft-law measures nudge online platforms to introduce safeguards in content moderation. More specifically, according to the Recommendation, online platforms are encouraged to publish the criteria for removal or blocking of access to online content in a clear, easily understandable and sufficiently detailed way.[566] Furthermore, the Recommendation invites online platforms to send users' confirmation after receiving users' notice.[567] In the case of removal or block of access to the noticed content, online platforms should, without undue delay, inform users about the decision providing also its reasoning as well as the possibility to contest such decision.[568] In turn, the content provider should have the possibility to contest the decision by submitting a 'counter-notice' within a 'reasonable period of time'.[569] The counter-notice can lead to a revision of the previous decision. Moreover, where online platforms implement automated means to process content, decisions to remove or disable access to content considered to be illegal content should be taken by ensuring safeguards such as 'human oversight and verifications, where appropriate and, in any event, where a detailed assessment of the relevant context is required in order to determine whether or not the content is to be considered illegal content'.[570]

[557] Ibid, Arts 8-8(a).
[558] Ibid, Art 8(1).
[559] Ibid, Art 9(2).
[560] Ibid, Arts 9(a)-10.
[561] Ibid, Art 10(2).
[562] Ibid, Arts 10-11.
[563] Ibid, Art 11(3).
[564] Communication on Tackling Illegal Content Online, Towards an enhanced responsibility of online platforms, COM(2017) 555 final.
[565] Recommendation of 1 March 2018 on measures to effectively tackle illegal content online (C(2018) 1177 final)
[566] Ibid, para. 16.
[567] Ibid, paras 5-8
[568] Ibid, paras 9-10
[569] Ibid, paras 11-13
[570] Ibid, para. 20.

The approach of the Union in this field shows a shift from a liberal approach in online content moderation to transparency and accountability obligations. In particular, rather than just focusing on content regulation, the EU approach tends to introduce procedural safeguards for users to dismantle the logic of opacity. However, it is worth underlining how users' safeguards in online content moderation have not been introduced horizontally to cover all content and situation. Indeed, the Union has maintained a vertical approach based on specific categories of content (e.g. copyright). Despite the crucial steps of the EU in this field, users and online platforms face the challenges raised by legal fragmentation in this field since it is still not possible to rely on a general legal framework of safeguards in online content moderation.

5. Injecting Democratic Values in Online Content Moderation

The lack of transparency and accountability in online content moderation limits how users understand how content is processed in the digital environment. Since users cannot generally rely on horizontal and general rights vis-à-vis online platforms, this situation leaves these actors free to decide how to balance and enforce fundamental rights online without any public guarantee. Since the liberal approach to free speech has shown to lead to some perverse collateral effects when applied to the digital environment, it seems that the mere protection of freedom of expression against interferences from public actors is not enough any longer in the digital environment. Therefore, to avoid that the protection of this fundamental right is frustrated by multinational private transnationally, it would be worth proposing to regulate online content moderation as a new form of private power exercised by online platforms.

At first glance, addressing this issue could lead to a change in the liability system of online platforms to increase their degree of responsibility in online content moderation. Nevertheless, this kind of regulatory approach could undermine the economic freedoms of online platforms, which would be overwhelmed by disproportionate obligations. Moreover, this solution would not solve the issue of transparency and accountability in online content moderation. On the contrary, increasing legal pressure on social networks by introducing monitoring obligations would result in 'overly aggressive, unaccountable self-policing, leading to arbitrary and unnecessary restrictions on online behavior'.[571] This risk, known as collateral censorship, could have strong effects on democracy, thus, requiring regulators to avoid threatening online platforms for failing to correctly police content. [572]

Therefore, at this time, the issue to solve is not just relating to the liability of online intermediaries but to the injection of procedural safeguards.[573] Otherwise, without regulating online content moderation, it is not possible to require online platforms, as private actors, to take into consideration fundamental rights when assessing users' requests. Regulating online content moderation would allow users to rely on a procedure against potential violation of their fundamental rights resulting from discretionary decisions by platforms concerning online content. This consideration shows the relevance of the law in this field. Constitutional provisions have been interpreted as a limit to the coercive power of the State or as a source of positive obligation for public actors to act to protect constitutional rights and liberties. As a result, as already underlined, the scope of constitutional rights allows private actors to claim the respect of their rights only vis-à-vis public actors. In the algorithmic society, instead, an equally important and

[571] Milton Mueller, 'Hyper-Transparency and Social Control: Social Media as Magnets for Regulation' (2016) 39(9) Telecommunications Policy 804, 809.
[572] Jack M. Balkin, 'Free Speech and Hostile Environments' (1999) Columbia Law Review 2295.
[573] Recommendation CM/Rec(2018)2 of the Committee of Ministers to member States on the roles and responsibilities of internet intermediaries [2018]; Aleksandra Kuczerawy, 'Safeguards for Freedom of Expression in the Era of Online Gatekeeping' (2018) 3 Auteurs & Media 292; Kate Crawford and Jason Schultz, 'Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms' (2014) 55 Boston College Law Review 93; Danielle K Citron and Frank Pasquale, 'The Scored Society: Due Process for Automated Predictions' (2014) 89 Washington Law Review 1.

pernicious threat for individuals come from those private actors which develop algorithms according to their ethical, economic and self-regulatory frameworks. While, in the past, the threats for individual rights was linked with State actions, today, democratic States deal with the issue of limiting the exercise of freedoms (or powers) exercised by private actors in the digital environment.

As already underlined, the different degree of protection of free speech across the Atlantic is one of the primary reasons for the opposite approaches undertaken by the EU and US. In Europe, serious threats for fundamental rights can be considered as sparkling triggers of the States' positive obligation to regulate private activities to protect fundamental rights as underlined by the European Court of Human Rights.[574] Therefore, the regulation of content moderation would not just result from the need to take into account other fundamental rights rather than freedom of expression but also to ensure the effective protection of the right to freedom of expression under the Convention and the Charter.[575] As observed by Kuczerawy, 'the duty to protect the right to freedom of expression involves an obligation for governments to promote this right and to provide for an environment where it can be effectively exercised without being unduly curtailed'.[576]

In the European framework, the regulation of content moderation would also derive from the need to ensure users a right to access remedies against the violations of their fundamental rights. According to Article 13 ECHR, 'everyone whose rights and freedoms as set forth in this Convention are violated shall have an effective remedy before a national authority notwithstanding that the violation has been committed by persons acting in an official capacity', along with the requirements of Article 1 on the obligation to respect human rights and Article 46 on the execution of judgments of the European Court of Human Rights. This provision requires Contracting parties not just to protect the rights enshrined in the ECHR but especially avoid that the protection of these rights is not frustrated by lack of domestic remedies. As observed by the Strasbourg Court, 'where an individual has an arguable claim to be the victim of a violation of the rights set forth in the Convention, he should have a remedy before a national authority in order both to have his claim decided and, if appropriate, to obtain redress'.[577] Similarly, Article 47 of the Charter provides even broader protection of this right being recognized by a general principle of EU law.[578]

In order to understand the complexities resulting from the regulation of content moderation, the next subsections aim to provide a guide on how the process of content moderation could be regulated by introducing and harmonizing procedural safeguards to increase the degree of transparency and accountability as well as avoiding discretionary interference with users' fundamental rights. More specifically, the next subsections suggest potential users' rights to foster democratic values in the digital environment. In particular, the process of content moderation has been divided into three parts: notice system, decision-making and redress. First, the notice phase includes the process through which users become aware of the various steps of the content moderation procedure either when the user is a notice provider or content provider. Second, the decision-making phase concerns the reasons and effects of content removal or blocking. Thirdly,

[574] See, for example, Von Hannover v Germany [2005] 40 EHRR 1; Verein gegen Tierfabriken Schweiz (VgT) v Switzerland [2001] 34 EHRR 159. See Lech Garlicki, 'Relations between Private Actors and the European Convention on Human Rights' in Andra Sajóand Renata Uitz (eds), The Constitution in Private Relations: Expanding Constitutionalism (Eleven International Publishing 2005).
[575] Aleksandra Kuczerawy, 'The Power of Positive Thinking. Intermediary Liability and the Effective Enjoyment of the Right to Freedom of Expression' (2017) 3 Journal of Intellectual Property, Information Technology and Electronic Commerce Law 182.
[576] Ibid, 186-7.
[577] See Leander v. Sweden, App. No. 9248/81, judgment of 26 March 1987, paragraph 77.
[578] See Case 222/84 Johnston v Chief Constable of the Royal Ulster Constabulary [1986] ECR 1651; Case 222/86 Union nationale des entraîneurs et cadres techniques professionnels du football (Unectef) v Georges Heylens and others [1987] ECR 4097; Case C-97/91 Oleificio Borelli SpA v Commission of the European Communities [1992] ECR I-6313.

the phase of redress regards the possibility for users to ask online platforms for a review of the first decision subject to specific conditions.

The following analysis of users' rights is based on four general principles: ban of general monitoring obligation; transparency and accountability in content moderation processing; proportionality of obligations applying to online platforms; availability of human intervention. More specifically, according to the first principle, States should not oblige platforms to generally moderate online content like established by the e-Commerce Directive.[579] This ban is crucial to safeguard fundamental rights such as freedom to conduct business, privacy, data protection and, last but not least, freedom of expression.[580] Secondly, content moderation rules should be explained to users *ex ante* in a transparent and user-friendly way. The 'content moderation notice' should include the guidelines and criteria used by online platforms to moderate content and explain the company's internal process to ensure that decisions are as predictable as possible. The third principle aims to strike a fair balance between rights of the users and obligations of platforms. Although the lack of transparent and accountable procedures relegates users in a position of *subjectionis,* however, the enforcement of users' rights should not lead to a disproportionate limitation to the right and freedom of online platforms in performing their business, especially for protecting new or small platforms. The fourth principle is based on introducing the principle of human-in-the-loop in content moderation. The role of humans in content moderation could be an additional safeguard allowing users to rely on a human translation of the procedure subject to specific conditions.

5.1 Notice System

The notice system is the first step of the process. It is possible to divide the notice phase into 'content notice' provided by online platforms and 'user's notice' coming from notice providers. The relevance of users' notice for content moderation has been already underlined as a sort of crowd-sourced censorship where users are an active part of the flagging system without being compensated for this activity.[581] Users are critical pieces of the content moderation puzzle since social media also rely on users to flag or, generally, report content.[582]

Despite its relevance, users' notice primarily concerns the phase of post-moderation. Nevertheless, as already underlined, moderation of content is also autonomously performed ex-ante by automated means, for instance, to tackle extreme content like terrorist videos when uploaded. Therefore, requiring online platforms to provide information in a 'content notice' could play a crucial role to explain to users how their content is processed and according to which conditions.

In other words, the content notice would foster transparency in online content moderation by allowing users to understand not only how content is processed by online platforms once they receive users' complaints but also in the phase of pre-moderation. In this case, at least the minimum content of this notice should be prescribed by law to avoid that online platforms can freely choose which information should be disclosed to users. More specifically, online platforms could be required to publish their content moderation guidelines where they explain how the process is organised and the criteria used to moderate content such as definitions of infringing content and the criteria to moderate each type of content.

[579] E-Commerce Directive (n 62), Art 15.
[580] See, for example, Case C-70/10 *Scarlet Extended SA v Société belge des auteurs, compositeurs et éditeurs SCRL (SABAM)* [2011]; Case C-360/10 *Belgische Vereniging van Auteurs, Componisten en Uitgevers CVBA (SABAM) v Netlog NV* [2012].
[581] Eugene Morozov, *The Net Delusion* (Penguin Press 2012).
[582] Crawford and Gillespie (n 74).

Once online platforms provide users with information to understand how content moderation is performed and the available remedies, users would also be aware of the procedures to submit complaints (or users' notice). Since users' notice generally triggers the responsibility of online platforms to act promptly to remove the online content in question, this step plays a crucial role for both online platforms and users. On the one hand, the former can understand when the obligation to remove specific content arises, on the other hand, the latter can know when the process of post-moderation has been initiated. It is worth underlining that users' notice does not trigger in any case the obligation of online platforms to remove content. In the case of the CDA, online platforms are not obliged to remove content unlike the process of 'notice and takedown' introduced by the DMCA, then, also adopted by the e-Commerce Directive. According to this system, once the notice provider submits its complaint to online platforms, the process of online content review starts since users' notice makes platforms aware of the presence of alleged illegal content. Furthermore, users' notice is not the only way for triggering platforms obligation to remove content since their awareness can also derive from other sources, for example, from the news or other events of public interests.

When addressing users' notice, the first step consists of understanding who is the notice provider. First, it would be possible to distinguish between notice sent by public and private actors. When the notice is submitted following a judicial order or the decision of an independent administrative authority, online platforms would be obliged to remove content without having the possibility to assess whether the content is lawful or unlawful. Instead, the notification from the Government and its dependent authorities should not fall under this category to ensure that these public actors do not exploit this preferential notice system as a free way to overcome any accountability and censor online speech. In this case, judicial and independent administrative authorities could have access to a separate process for notification to speed up content review and recover the time spent to assess the lawfulness of specific online content.

The case is different where notice providers are private actors. In this case, the primary issue is to decide whether all users are on the same position or, instead, some notice providers enjoy a privileged status (ie trusted providers). This category would include special notice providers that can rely on privileged channels to signalling content considered illegal. For example, newspapers and publishers could be trusted flaggers for contents involving defamation or disinformation. The same approach could be adopted for other specialised notice providers such as collecting societies for copyright content. Nevertheless, since this choice would empower some entities in deciding about speech online, it would be necessary that the categories of trusted flaggers are provided and periodically reviewed by law or, at least, by independent competent authorities. In both cases, it should be observed that online platforms maintain discretion in deciding whether to remove or block specific content since the notice does not result from an order issued by the judiciary or independent administrative authorities. Therefore, it would be possible to divide notice providers into three categories: public authorities, trusted providers and users.

The second step consists of understanding according to which conditions the notice could be considered valid to trigger the process of online content review. Indeed, it cannot be excluded that the information provided for by the notice provider could be not adequate to process users' requests. More specifically, the notice could lack the URL to identify the content at stake or do not explain what the issue at stake is.[583] This issue is strictly linked with the form according to which notices are sent to online platforms. According to the current system, a notice can also be sent by mail to online platforms. This fragmentation could be mitigated by requiring online platforms to establish forms with mandatory information. However, since, even in this case, this

[583] Case C-324/09 L'Oréal SA and Others v eBay International AG and Others [2011] ECR I-6011, para 122.

discretion would empower platforms to select which information the users should insert even if not necessary, it would be necessary to rely on criteria provided by law or competent authorities.

The third step focuses on determining the flow of notice between three entities: the notice provider, content provider and online platform. Once the notice provider sends its notice to online platforms, the notice provider could receive at least other two notices before the decision. The first notice would consist of an automatic reply confirming that the request has been received and how it will be processed by the platform. The second notice could occur before the decision is implemented. This second contact would allow the notice provider to decide to add other information or withdraw its complaint.

Within this framework, the notice should also involve the content provider. In order to ensure transparency in this process, it would be appropriate that content providers are informed about the review process which could potentially lead to the removal or block of one of their content. This notice could occur once the platform starts its reviewing process after receiving the notice from the notice provider. In this case, the content provider should have the opportunity to submit its observations and proves to contest the notice. In this way, the possibility for content providers to object complaints on their content would inject in this phase the rights to a fair hearing, adversarial proceedings and equality of arms in the process of content moderation. It cannot be excluded that, in the above-mentioned cases, the notice could be limited to protect other interests such as confidentiality or the need to maintain secrecy in ongoing investigations. These exceptions should set by law to avoid that platforms raise several exceptions undermining, *de facto*, the notice system.
Furthermore, in order to avoid any abuse of the notice system, it would be necessary to set mechanisms of compensation against users' misconducts such as compensation for the damage caused by submitting false notice or information. These mechanisms aim to avoid overwhelming platforms with fraudulent requests.
Once these steps are completed and online platforms adopt their decisions, another notice should be sent both to the notice and content provider to inform them about the result of content moderation.

5.2 Decision-making

Once human or machine moderators process content, online platforms decide whether to maintain or remove it. Since decision-making is the phase firmly affecting fundamental rights, additional safeguards should be implemented. Indeed, users should be in the position to understand the criteria online platforms implement to moderate content. Therefore, in this phase, the primary concern regards the possibility to explain the path followed by online platforms to reach a specific output.

First, as already observed, online content moderation is basically performed by algorithmic systems. This system can be implemented to autonomously decide whether to shut down content or suggest potential infringing content to human moderators. Since automation plays a crucial role in moderating content, one of the primary questions concern how to ensure that automated decisions can be foreseeable and transparent. It is no coincide whether transparency is at the core of the debate about algorithms.[584] The risks for fundamental rights and democracy are strictly linked to the lack of transparency about the functioning of automated decision-making

[584] See, in particular, Daniel Neyland, 'Bearing accountable witness to the ethical algorithmic system', (2016) 41 Science, Technology & Human Values 50; Mariarosaria Taddeo, 'Modelling trust in artificial agents, a first step toward the analysis of e-trust' (2010) 20 Minds and Machines 243; Matteo Turilli and Luciano Floridi, 'The ethics of information transparency' (2009) 11 Ethics and Information Technology 105.

processes.[585] Ensuring transparency could be complicated for reasons relating to the protection of other interests such as trade secrets.[586] The issue can be explained due to the impossibility to predict the result of algorithms and reconstruct the elements which have led to a specific output due to the vast amount of data involved.[587]

This possibility is of particular concern for users when observing some pitfalls in algorithmic decision-making processes. In order to understand these challenges, it would be possible to divide the algorithmic process into three phases: input, process and output. First, algorithmic input is made of data which, then, is processed to obtain an output. Therefore, the quality of data firmly affects the algorithmic output. Although the entire automated process could fit with the purposes of content reviewing, however, the way according to which online platforms have trained algorithms could lead to unforeseeable outputs.

Concerning the process, it is necessary to distinguish between deterministic algorithms and systems based on machine learning. In the first case, since the procedure is based on pre-established steps, the prediction of a specific outcome could be possible. When, instead, machine learning is involved in content moderation, it could become complex to explain the process made to reach a specific output. Some algorithms can be considered 'black boxes' since their internal processes are incomprehensible to humans.[588] The aim of algorithms in content moderation is not to censor but to classify information according to specific clusters where content is considered 'lawful' or 'unlawful'. As a result, online content as input is transformed into predictions of the lawfulness of such information as output. This process is based on the system of trial and error where algorithms are trained based on the accuracy of their decision. This mechanism explains why some algorithms still lack that degree of accuracy to detect infringing content or take into consideration the general background. As a result, notwithstanding the output is the most relevant part for users, it is just a small part of the algorithmic jigsaw.

Moreover, despite the relevance of artificial intelligence in content moderation, the role of human moderators in the phase of decision-making cannot be neglected since moderators around the world usually take the last decision. Usually, moderators can rely on less than a minute to decide whether to remove certain content.[589] This strict time frame could be considered a fundamental clue to argue how human moderators cannot consistently comply with either a legal standard or any internal guidelines. Therefore, the process of content moderation is left in the hand of moderators' will.

It is worth observing that this activity is far from the judicial environment where a court decides whether content is illegal. While a trial could last years to decide whether a statement is defamatory, moderators take this decision in a bunch of seconds. Moderators do not only lack legal skills but also come from very different backgrounds since the activity of content moderation

[585] Jenna Burrell, 'How the Machine "Thinks": Understanding Opacity in Machine Learning Algorithms' (2016) 3 Big Data & Society 1; Christopher Kuner and others, 'Machine Learning with Personal Data: Is Data Protection Law Smart Enough to Meet the Challenge?' (2017) 6 International Data Privacy Law 167; Mireille Hildebrandt, 'The Dawn of a Critical Transparency Right for the Profiling Era', in Jacques Bus and others (eds), *Digital Enlightenment Yearbook* (IOS Press 2012); Meg L. Jones, 'Right to a Human in the Loop: Political Constructions of Computer Automation and Personhood' (2017) 47 Social Studies of Science 216.

[586] Luciano Floridi, *The Fourth Revolution: How the Infosphere is Reshaping Human Reality* (Oxford University Press 2014).

[587] Joshua A. Kroll and others, 'Accountable Algorithms' (2016) 165 University of Pennsylvania Law Review 633; Andreas Matthias, 'The responsibility gap: Ascribing responsibility for the actions of learning automata' (2004) 6(3) Ethics and Information Technology 175.

[588] Frank Pasquale, *The Black Box Society. The Secret Algorithms that Control Money and Information* (Harvard University Press 2015); Maayan Perel and Niva Elkin-Koren, 'Black Box Tinkering: Beyond Disclosure in Algorithmic Enforcement' (2017) 69 Florida Law Review 181.

[589] Olivia Solon, 'To censor or sanction extreme content? Either way, Facebook can't win' (The Guardian 23 May 2017) <https://www.theguardian.com/news/2017/may/24/facebook-struggles-with-mission- impossible-to-stop-online-extremism> accessed 25 July 2019.

is often outsourced to countries such as the Philippines.[590] The same concern can be extended to their working conditions which do not allow to perform the activity of content moderation with due care.[591]

Furthermore, automated and human moderation usually fails to reach a degree of granularity that allows taking into consideration the different nuances between contexts around the world. While automated technologies tend to consistently classify content in different clusters, values and principles are local and influenced by cultural diversities. Indeed, even if automated content moderation can help online platforms to perform this activity, their set of values and principles cannot reflect the multiplicities of communities in the world with the result that some content can be penalised for expressing values different from those on which algorithms have been trained and programmed. Although online platforms found their narrative on their role in establishing and promoting the values of an open and global community, it is worth wondering how it is possible to agree on common rules between communities which, in some cases, are also made up of two billion of people. Similar considerations apply for human moderator dealing with content concerning event far not only geographically but also culturally and socially. Moderators usualy decide in less than a minute which content should be removed, no matter whether a specific content comes from different situations or environments. While the activity of content moderation is easier for some content such as child abuse or terrorism, hate speech and disinformation could challenge both human and machine moderators.

Notwithstanding decision-making processes are often complicated to unbox, they ultimately affect users' fundamental rights since possible decisions are just 'ignore' or 'delete'. Within this framework, the primary question concerns the degree of explanation users should have the right to access. In the field of data, this issue has been discussed within the framework of the GDPR. More specifically, the debate about the right to explanation shows the complexity of this issue. Scholars have recently focused on understanding whether the GDPR provides a legal ground for individuals to defend themselves from potentially harmful consequences of the implementation of algorithms, most notably by creating a 'right to explanation' in respect of automated decision-making processes.[592] Articles 13-15 of the GDPR expressly require controllers to provide data subjects the information about 'the existence of automated decision-making, including profiling, referred to in Article 22(1) and (4) and, at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject'. Some scholars argued that the GDPR fosters a form of qualified transparency over algorithmic decision-making.[593] Instead, others support or deny the existence of such a right,[594] or doubt that the GDPR provisions provide a concrete remedy to algorithmic decision-making processes.[595]

More specifically, Article 22 provides a general rule according to which the data subject has the right not to be subject to a decision 'based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her'.[596]

---

[590] Adrian Chen 'The Laborers Who Keep Dick Pics and Beheadings Out of Your Facebook Feed (*Wired*, 23 October 2014) <https://www.wired.com/2014/10/content-moderation/> accessed 22 July 2019.
[591] Roberts (n 3).
[592] Bryce Goodman and Set Flaxman, 'European Union Regulations on Algorithmic Decision-Making and a "Right to Explanation"' (2017) 38 AI Magazine 50; Andrew D. Selbst and Julia Powles, 'Meaningful information and the right to explanation' (2017) 7 International Data Privacy Law 233.
[593] Margot E. Kaminski, 'The Right to Explanation, Explained' (2019) 34(1) Berkeley Technology Law Journal.
[594] Gianclaudio Malgieri and Giovanni Comandè, 'Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation' (2017) 7 International Data Privacy Law 234; Sandra Wachter, Brent Mittelstadt and Luciano Floridi, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' (2017) 7 International Data Privacy Law 76;
[595] Lilian Edwards and Michael Veale, 'Slave To The Algorithm? Why a 'Right to an Explanation' is Probably not the Remedy You are Looking for' (2017) 16 Duke Law & Technology Review 18.
[596] However, some exceptions to this general prohibition are laid down by para. 2. According to Art 22(2): 'Paragraph 1 shall not apply if the decision (a) is necessary for entering into, or performance of, a contract between the data subject

Even more importantly, article 22(3) codifies the interpretation of Recital 71 establishing the obligation for data controller to implement suitable measures to ensure that data subjects' rights among which there is 'at least', the possibility to obtain the human intervention, express his or her point of view and to contest decision. Indeed, Recital 71 specifies that the processing should be subject to suitable safeguards including 'specific information to the data subject and the right to obtain human intervention, to express his or her point of view, to obtain an explanation of the decision reached after such assessment and to challenge the decision'.

The same challenges can be extended to the field of online content since decisions affecting users' rights could be completely automated. First, in the content notice, users should be able to access ex-ante explanations about the logic used by online platforms to moderate content. Second, although it would be burdensome for online platforms to provide a full human motivation for any decision, users should receive at least information about the decision such as the result of the review, information about redress mechanisms, the location, timing and identification number of the moderator who has reviewed a specific content. More importantly, since any restriction of content constitutes an interference with freedom of expression, online platforms should provide a brief explanation of the reason for the removal indicating on which ground the content has been eliminated (prescribed by law), for which purpose (legitimate aim) and the criteria used (proportionality). Therefore, moderators should not limit their activity to a binary decision ('ignore' or 'delete') but insert even brief information about the removal.

Third, human moderation constitutes a crucial safeguard in the decision-making phase to fill the 'black boxes' gap. However, a general rule applying human intervention to all the situation could be a burden for online platforms. In this scenario, the right to rely on human intervention in online content moderation could be applied at users' request and based on the type algorithms used for content moderation. More specifically, in this case, it would be possible to apply a system of 'scale protection' where human intervention is increasingly required as long as algorithms are less deterministic. For example, where machine learning technologies are involved in online content moderation, human intervention could apply by default. Moreover, human intervention could be limited when the decision is the result of a notice coming from public authorities or a trusted notice provider due to their peculiar role.

5.3 Redress

The redress phase is the last and eventual step of content moderation. Once online platforms decide to remove or maintain content, users should be able to ask online platforms to review the previous decision subject to certain conditions. This right aims to provide users with a second chance, especially when decisions are entirely the result of automated processes. In this phase, the provision of human intervention would be mandatory as a minimum standard. In this way, users should have the possibility to rely on humans to review previous decisions. Otherwise, an automated review of the first decision would make ineffective this right, especially when an automated system has been involved in the first decision. This step is particularly important due to the inaccurate assessments that can derive from automated and human moderations.

When addressing redress mechanisms, it is possible to focus on three main steps: modalities of access, reviewing process and remedies. First, it is necessary to focus on whether access to redress mechanism should be opened to content providers and notice providers both when online platforms remove online content and refuse to perform this activity. Recognising the right to

and a data controller; (b) is authorised by Union or Member State law to which the controller is subject and which also lays down suitable measures to safeguard the data subject's rights and freedoms and legitimate interests; or (c) is based on the data subject's explicit consent'.

redress mechanism just in one of the two cases could produce negative effects. On the one hand, when recognising this right only when online platforms refuse to remove or block content, this choice would encourage platforms to censor contents to avoid the burden of redress mechanism with serious risk of collateral censorship. On the other hand, if access to redress mechanism would be possible only in case of removal or block, the gap between the two systems would favour content provider since notice provider could not rely on redress mechanism when their complaint has been rejected. This system would not involve public actors since they usually are those who notify online content. Instead, where public actors are content provider, they could be part of a redress mechanism.

Another condition of access could be based on the use of automated technologies. If, on the one hand, the first decision has been taken solely by automated technologies, the redress mechanism could always be accessible to allow users to rely on a human review. On the other hand, access to redress mechanism could be restricted when the decision has been taken only by humans or by human supported by automated systems. In this case, it is important to inform users in the phase of decision-making whether a human or machine has addressed the case in question. Furthermore, users should be able to rely on the possibility to obtain a second decision from a moderator based in the geographical region as close as possible to the user's location. In this way, moderation of content would be more granular and accurate according to the specific cultural and social context.

A third criterion would be based on the clarity of motivation. It cannot be excluded that platforms would receive similar or almost identical complaints. In this case, if the cases are serial and the motivation of the first decision is provided explaining the reasons for the removal or maintenance of content, the access to redress mechanism could be subject to the discretion of the platform. In other words, a detailed motivation of the first decision could be considered a way to exempt platforms from implementing redress mechanisms and, at the same time, could encourage online platforms to provide more information about the first decision.

The second point involves the review of the first decisions. The primary remedy would consist of dismissing the first decision. Whenever online platforms restrict content, they should ensure the possibility to reinstate content. Indeed, if the user disagrees with the first decision and relies on the redress mechanism established by the platform, it is necessary that the content previously removed is still available in case online platforms review their first decision. Therefore, the reinstatement of content should always be technically possible. Besides, it cannot be excluded a more detailed system where online platforms can review their decisions by restricting the removing or blocking to a geographical area or providing or banning users' profiles in case of repeated infringements. Even in this case, remedies should be provided by law to avoid that online platforms exercise quasi-public roles without any safeguard.

Moreover, it is also worth wondering whether the review procedure should be oral or written, and conducted by employees or independent experts such as an oversight board. Due to the number of potential users' requests, the written procedure would be the most suitable for online content moderation. However, it cannot be excluded that, in some peculiar cases, platforms could provide a sort of alternative dispute resolution mechanism based on different procedures. In this case, without regulating the entire review process but in order to decrease the degree of discretion in the phase of redress, the law could establish at least some conditions, especially concerning the appointments of independent experts and the time frame which should be respected to review the first decision.

Regarding the motivation of the review, providing explanations for every decision could be potentially burdensome since online platforms should invest additional resources. Nevertheless, in this case, since motivation would be an important contribution in setting up a coherent list of

cases based on established precedents to limit further users' appeals, the explanation for removal or reinstatement should be required only to some platforms according to specific thresholds based, for example, on their global turnover. Review decisions should be available to users and published in online platforms' webpages.

6. Conclusions

The current opacity of content moderation constitutes a challenge for democratic societies. If individuals cannot understand the reasons behind decisions involving their rights, especially when automated decision-making systems are involved, the pillars of transparency and accountability on which democracy is based are destined to fall.

While, in the past, the liberal approach to free speech fits with the purpose to safeguard democratic values in the digital environment, today, the emergence of new powers governing the flow of information would require a shift from a negative dimension to a positive approach by regulating content moderation.

The liberal approach adopted at the end of the last century has led online platforms to impose their authoritative regime on content based on a mix of technological and contractual tools. The result of this situation has led users in a status of *subjectionis* where they find themselves forced to comply with standards and safeguards autonomously established by each online platform.

Within this framework, the Union has started to focus on introducing mechanisms of transparency and accountability in online content moderation. For example, the rights to obtain motivation or human intervention in online content moderation are important steps towards a more democratic digital environment. Indeed, these users' rights should not be considered only as instruments to improve transparency and accountability in the digital environment but also to limit the discretion of online platforms operating as private powers outside any constitutional boundary.

Nevertheless, it is necessary to observe that Union efforts are not still enough to ensure a path towards the democratisation of the digital environment. Today, users can rely on certain rights only in the Union and just for specific contents. This choice could lead to an axiological prevalence of some interests in online content moderation since users cannot generally rely on the same rights for all online content. Furthermore, the fragmentation of users' rights affects also the platforms' freedom to conduct business since it requires these actors to set different regimes of content moderation. However, the fragmentation of users' rights is not the only concern at stake. Notwithstanding the Union has introduced new users' rights in content moderation, online platforms still enjoy a broad margin of discretion to decide how to implement them. Regarding the notice system, it is not specified who could be considered trusted notice provider. Besides, the boundaries of motivation in content removal are not entirely clear. The same consideration applies for redress mechanism where the review of the platform's decision is not subject to any due process obligation.

Despite these challenges, the approach of the Union is remarkable and underlines the relevance of EU constitutional law in reacting against new forms of powers raising transnational challenges and undermining democratic values. Like in the field of data protection, the Union has started to pave the way towards the regulation of online platforms activities with an increasing convergence of users' rights in the field of data and content. In other words, the Union approach can be considered a first crucial step towards a new season of content moderation where online platforms are required to operate as responsible actors in light of their gatekeeping role in the digital environment.

Platform Values and Democratic Elections: How can the law regulate digital disinformation?

Professor Chris Marsden, Sussex, Dr Trisha Meyer, Vrije Universiteit Brussel, Dr Ian Brown*[597]

Abstract

This article examines how governments can regulate the values of social media companies that themselves regulate disinformation spread on their own platforms. We use 'disinformation' to refer to motivated faking of news. We examine the effects that disinformation initiatives (many based on automated decision-making systems using Artificial Intelligence [AI] to cope with the scale of content being shared) have on freedom of expression, media pluralism and the exercise of democracy, from the wider lens of tackling illegal content online and concerns to request proactive (automated) measures of online intermediaries. We particularly focus on the responses of the member states and institutions of the European Union. In Section 1, we argue that the apparent significance of the threat has led many governments to legislate despite this lack of evidence, with over 40 national laws to combat disinformation chronicled by March 2019. Which types of regulation are proposed, which actors are targeted, and who is making these regulations? Regulating fake news should not fall solely on national governments or supranational bodies like the European Union. Neither should the companies be responsible for regulating themselves. Instead, we favour co-regulation. Co-regulation means that the companies develop – individually or collectively – mechanisms to regulate their own users, which in turn must be approved by democratically legitimate state regulators or legislatures, who also monitor their effectiveness. In Section 2, we explain the current EU use of Codes of Conduct. In Section 3, we then explain the relatively novel idea that social media content regulation, and specifically disinformation, can be dealt with by deploying AI at massive scale. It is necessary to deal with this technological issue in order to explain the wider content of co-regulatory policy options, which we explain and for which we argue in Section 4. In Section 5 we explain what this means for technology regulation generally, and the socio-economic calculus in this policy field.

Keywords: disinformation, Artificial Intelligence, co-regulation, self-regulation, Internet law, social media regulation, platform regulation, elections, fake news.

[597] Chris Marsden is Professor of Internet Law at the University of Sussex; Trisha Meyer is Assistant Professor and Postdoctoral Researcher at the Vrije Universiteit Brussel; Ian Brown is Senior Research Fellow at Research ICT Africa. This article is based on their expert report, Marsden, C. & Meyer, T. *Regulating Disinformation with Artificial Intelligence. The Effects of Disinformation Initiatives on Freedom of Expression and Media Pluralism*, (Brussels: European Parliament, 2019) doi: 10.2861/003689. See also Meyer, T., Marsden, C. & Brown, I. "Regulating disinformation with technology: analysis of policy initiatives relevant to illegal content and disinformation online in the European Union" in E. Kużelewska, G. Terzis, D. Trottier & D. Kloza (eds.) *Disinformation and digital media as a challenge for democracy*, European Integration and Democracy Series, Vol. 6. (Cambridge: Intersentia 2020, in print). A summary open access article was earlier published, see Marsden, C. and T. Meyer (2019) "How can the law regulate removal of fake news?" Computers and Law, https://www.scl.org/articles/10425-how-can-the-law-regulate-removal-of-fake-news

1. ‘Fake News’ and Disinformation on Social Media Platforms

The digitization of disinformation via social media platforms has been blamed for skewing the results of elections and referenda and amplifying hate speech in many other nations.[598] The evidence of harm, and legislative and judicial responses to deliberate disinformation, are growing but nascent. [599] The apparent significance of the threat has led many governments to legislate despite this lack of evidence, with over forty national laws to combat disinformation chronicled by March 2019.[600] ‘Fake news’ – more properly termed disinformation – has recently become endemic to social networking on the Internet.

We use ‘disinformation’ to refer to motivated faking of news, in line with the European Union (EU)’s institutions[601] and High Level Expert Group on disinformation[602], and the regional and global United Nations (UN) rapporteurs on freedom of information’s use of the term.[603] It is a problem at least as old as written media, but has become more controversial as evidence of state-sponsored domestic and foreign influence peddling online, and micro-targeted political influence marketing via social media, has become ubiquitous. Particularly, the problem of state-sponsored social media inaccuracy was first identified in the Ukraine in 2011, when Russia was accused of deliberately faking news of political corruption.[604]

This article examines how governments can regulate the values of companies (Facebook, YouTube, Twitter in particular) that themselves regulate disinformation spread on their own platforms. We examine the effects that disinformation initiatives (many based on automated decision-making systems using AI to cope with the sheer scale of content being shared) have on freedom of expression, media pluralism and the exercise of democracy, from the wider lens of tackling illegal content online and concerns to request proactive (automated) measures of online intermediaries.[605] We particularly focus on the national and supranational responses of the member states and institutions of the European Union, which in the European Parliament elections of May 2019 held the largest global exercise in democracy behind Indian general elections, with a highly networked electorate, and serious concerns about foreign interference.

In international human rights law, such as Article 10 of the European Convention on Human Rights and Fundamental Freedoms 1950[606], restrictions to freedom of expression must be provided by law, legitimate and proven necessary and the least restrictive means to pursue the aim. In this article, we argue that governments should not push this difficult judgement exercise in disinformation onto online intermediaries, who are inexpert in and not incentivized to judge

---

[598] Euronews *How Can Europe Tackle Fake News in the Digital Age?*, (9 Jan 2019) https://www.euronews.com/2019/01/09/how-can-europe-tackle-fake-news-in-the-digital-age

[599] de Cock Buning, M. (10 Sept 2018) ‘We Must Empower Citizens In The Battle Of Disinformation’, *International Institute for Communications,* http://www.iicom.org/themes/governance/item/we-must-empower-citizens-in-the-battle-of-disinformation

[600] Funke, Daniel (2019) *A guide to anti-misinformation actions around the world*, Poynter Institute, dynamically updated, at https://www.poynter.org/ifcn/anti-misinformation-actions/

[601] Bentzen, Naja, *Understanding Propaganda and Disinformation*, (European Parliament Research Service 2015), http://www.europarl.europa.eu/RegData/etudes/ATAG/2015/571332/EPRS_ATA(2015)571332_EN.pdf

[602] High Level Expert Group on Fake News and Online Disinformation *Report to the European Commission on A Multi-Dimensional Approach to Disinformation*, (2018) https://ec.europa.eu/digital-single-market/en/news/final-report-high-level-expert-group-fake-news-and-online-disinformation

[603] U.N. Special Rapporteur on Freedom of Opinion and Expression et. al., *Joint Declaration on Freedom of Expression and “Fake News,” Disinformation and Propaganda*, U.N. Doc. FOM.GAL/3/17 (Mar. 3, 2017), https://www.osce.org/fom/302796?download=true

[604] *See* Sanovich, Sergey *Computational Propaganda in Russia: The Origins of Digital Misinformation* (Oxford Computational Propaganda Research Project, Working Paper No. 2017.3, 2017), http://comprop.oii.ox.ac.uk/wp-content/uploads/sites/89/2017/06/Comprop-Russia.pdf

[605] ACR techniques became newsworthy in 2016 with the development of eGLYPH for removal of terrorist content: see The Verge (2016) *Automated Systems Fight ISIS Propaganda, But At What Cost?,* https://www.theverge.com/2016/9/6/12811680/isis-propaganda-algorithm-facebook-twitter-google

[606] Convention for the Protection of Human Rights and Fundamental Freedoms as amended by Protocols No. 11 and No. 14, European Treaty series No.5, Signed Rome, 4 November 1950, at https://www.coe.int/en/web/conventions/full-list/-/conventions/rms/0900001680063765

fundamental rights, and not bound by States' international human rights commitments.[607] The UN Special Rapporteur on Freedom of Opinion and Expression recently called for assessments of the impact of technology-based solutions on human rights in general[608], and freedom of expression and media pluralism in particular.[609]

What can be done, by whom, to whom, to address these problems? Which types of regulation are proposed, which actors are targeted, and who is making these regulations? Who should regulate fake news shared online? We argue that regulating fake news should not fall solely on national governments or supranational bodies like the European Union. Neither should the companies be responsible for regulating themselves and ourselves[610]. Instead, we favour co-regulation. Co-regulation means that the companies develop – individually or collectively – mechanisms to regulate their own users, which in turn must be approved by democratically legitimate state regulators or legislatures, who also monitor their effectiveness.[611] The article proceeds as follows. In the following Section 2, we explain the current use of Codes of Conduct. In Section 3, we then explain the relatively novel idea that social media content regulation, and specifically disinformation, can be dealt with by deploying AI at massive scale. It is necessary to deal with this technological issue in order to explain the wider content of policy options, for which we argue in Section 4 that co-regulation is the most likely and appropriate outcome. In Section 5 we explain what this means for technology regulation generally, and the socio-economic calculus in this policy field.

2. Current State of Play for European Disinformation Policy

Within Europe, online disinformation is currently tackled by regulators from a variety of regulatory angles. It can be limited through stipulations and actions against defamation, incitement to hatred and violence, or the ban on certain misleading advertising techniques. Within the context of electoral campaigns, the problem can be tackled by regulating spending and transparency of political campaigns, enforcing data protection rules and bolstering against cyberattacks. The best known example at national legislation is Germany's Network Enforcement Law 2017 ('NetzDG')[612]. This article however focuses on European level responses. Different aspects of the disinformation problem merit different types of regulation. More broadly, institutional support can be provided to safeguard media pluralism, encourage fact-checking and enhance media literacy. We note that all proposed policy solutions explained in section 4 stress the importance of literacy and cybersecurity. Holistic approaches point to challenges within the changing media ecosystem and stress the need to address media pluralism as well.

The most important European policy document dealing with disinformation was the result of policy formation in 2018. The 2018 EU-orchestrated 2018 self-regulatory *Code of Practice on Online Disinformation* followed from the EU High Level Group report, and examined technology-based solutions to disinformation, focused on the actions of online intermediaries (social media

[607] Brown, I. and Korff, K. *Digital Freedoms in International Law: Practical Steps to Protect Human Rights Online* (Global Network Initiative, 2012) https://globalnetworkinitiative.org/wp-content/uploads/2016/10/GNI-Digital-Freedoms-in-International-Law.pdf
[608] UNHRC, *Report of the UN Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression* (29 August 2018) UN Doc A/73/348
[609] UN Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression *Report to the United Nations Human Rights Council on A Human Rights Approach to Platform Content Regulation,* A/HRC/38/35, (6 April 2018) https://undocs.org/A/HRC/38/35
[610] Belli, L., Francisco, Pedro Augusto P.; Zingales, N.eds. *Platform regulations: how platforms are regulated and how they regulate us* (FGV, Rio de Janeiro 2017)
[611] Marsden, C. 'Internet Co-Regulation and Constitutionalism: Towards European Judicial Review', *International Review of Law, Computers and Technology 26(2)* (2012) 212-228
[612] Netzwerkdurchsetzungsgesetz (German Network Enforcement Act) 2017, see EU Code of Practice on Disinformation. Annex II Current Best Practices from Signatories of the Code of Practice (2018) https://ec.europa.eu/digital-single-market/en/news/code-practice-disinformation

platforms, search engines and online advertisers) to curb disinformation online[613]. Though criticized by its own Sounding Board for not stipulating any measurable outcomes,[614] Nielsen argued the Code of Practice produced "three potentially major accomplishments":[615]

- Signatories commit to bot detection and identification by promising to "establish clear marking systems and rules for bots to ensure their activities cannot be confused with human interactions".

- Signatories must submit their efforts to counter disinformation to external scrutiny by an independent third party: "an annual account of their work to counter Disinformation in the form of a publicly available report reviewable by a third party".

- A joint, collaborative effort based on shared commitments from relevant stakeholders including researchers, where signatories promise not to "prohibit or discourage good faith research into Disinformation and political advertising on their platforms".[616]

Other EU initiatives also call for pro-active measures by intermediaries through use of AI to aid removal of illegal content. The proposed EU Regulation on the Prevention of Dissemination of Terrorist Content Online[617] targets rapid removal terrorist content by online intermediaries. Article 17 of the recently passed Copyright in the Digital Single Market Directive 2019 suggests changing intermediary liability protections with a requirement to use filtering technologies[618]. The European Commission has used the overarching phrase "a fair deal for consumers"[619]. These policy developments fit in a context where social media platforms and search engines are increasingly scrutinized on competition grounds[620] and requested to take more responsibility for content removal.

If the socio-technical balance is trending towards greater disinformation, a lack of policy intervention is not neutral, but erodes protection for fundamental rights to information and expression. It is notable that after previous democratic crises involving media pluralism and new technologies (radio, television, cable and satellite), parliaments passed legislation to increase media pluralism by, for instance, funding new sources of trusted local information (notably public service broadcasters), authorizing new licensees to provide broader perspectives, abolishing mandatory licensing of newspapers or even granting postage tax relief for registered publishers,

---

[613] European Commission, *Code of Practice on Disinformation* (2018) https://ec.europa.eu/digital-single-market/en/news/code-practice-disinformation

[614] European Commission *Code of Practice on Disinformation,* Press Release, (26 September 2018) https://ec.europa.eu/digital-single-market/en/news/code-practice-disinformation

[615] Nielsen, R.K. 'Misinformation: Public Perceptions and Practical Responses', *Misinfocon London*, hosted by the Mozilla Foundation and Hacks/Hackers, (24 Oct 2018) https://www.slideshare.net/RasmusKleisNielsen/misinformation-public-perceptions-and-practical-responses/1

[616] Nielsen, R.K. *Disinformation Twitter Thread*, (26 Sept 2018) https://twitter.com/rasmus_kleis/status/1045027450567217153

[617] Proposed EU Regulation on Prevention of Dissemination of Terrorist Content Online (COM(2018) 640 final - 2018/0331 (COD)) https://ec.europa.eu/commission/sites/beta-political/files/soteu2018-preventing-terrorist-content-online-regulation-640_en.pdf

[618] 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, OJ L 130, 17.5.2019, p. 92–125

[619] Vestager, M. 'Competition and A Fair Deal for Consumers Online', *Netherlands Authority for Consumers and Markets Fifth Anniversary Conference,* (26 April 2018, The Hague), https://ec.europa.eu/commission/commissioners/2014-2019/vestager/announcements/competition-and-fair-deal-consumers-online_en

[620] For a scholarly overview and discussion of ongoing platform and search engine competition cases, see Mandrescu, D. 'Applying EU Competition Law to Online Platforms: The Road Ahead – Part I', Competition Law Review 38(8) (2017) 353-365*;* Mandrescu, D. 'Applying EU Competition Law to Online Platforms: The Road Ahead – Part II', Competition Law Review 38(9) (2017) 410-422. For an earlier call to co-regulation, see Marsden, C. (2012) n.15.

and introducing media ownership laws to prevent existing monopolists extending their reach into new media.[621]

Broadcasting is defined in Article 1 of the Audio Visual Media Services (AVMS) Directive as: "editorial responsibility of a media service provider [for the] principal purpose of providing programmes, in order to 'inform, entertain or educate' to general public, conveyed by electronic communications networks"[622]. That is distinguished from Internet communication by its specific audience and that fact that the user chooses the content[623]. Broadcasting law has extensive statute and case law, including that interpreted by the European Convention on Human Rights[624], appealed to the European Court of Human Rights at Strasbourg[625], in regulating election advertising, there was until recently a paucity of case law for the Internet. Baroness Hale has argued that: "In the United Kingdom, and elsewhere in Europe, we do not want our government or its policies to be decided by the highest spenders. …We have to accept that some people have greater resources than others with which to put their views across. But we want to avoid the grosser distortions which unrestricted access to the broadcast media will bring."[626]

Application of broadcast rules to the Internet in the case of video on demand (VOD) services was applied by courts in Belgium nearly two decades ago[627]. Were the same reasoning applied more broadly to the Internet, specific electoral spending law would be rigorously applied by regulators and upheld by courts. The dissenting Strasbourg judges in *Animal Defenders International* argued strongly in 2013 that this should not be so: "Given the comparative potency of newer media such as the Internet, a distinction based on the particular influence of the broadcast media was not relevant. Information obtained through the use of the Internet and social networks is gradually having the same impact, if not more, as broadcasted information."[628] The majority might argue

[621] See e.g. C-288/89 *Stichting Collectieve Antennevoorziening Gouda and others* judgment of 25 July 1991 [1991] ECR I-4007); Directive 89/552/EEC on the Coordination of Certain Provisions Laid Down by Law, Regulation or Administrative Action in Member States concerning the Pursuit of Television Broadcasting Activities, OJ L 298, 17.10.1989, pp.23–30 (particularly Recital 17).

[622] Directive 2010/13/EU of the European Parliament and of the Council of 10 March 2010 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) OJ L 95, 15.4.2010, pp.1–24. Note this Directive is Consolidated, having been adopted in 2007 with transposition date 19 December 2009. Specifically excluded are services including: private correspondence such as email; games of chance, on-line games, search engines (Recital 22); stand-alone text based services (Recital 23); electronic newspapers (Recital 28); services 'where any audiovisual content is merely incidental to the service and not its principal purpose' (Recital 22). See Valcke, P, and Stevens, D. 'Graduated regulation of "regulatable" content and the European Audiovisual Media Services Directive: one small step for the industry and one giant leap for the regulator?' 24 Telematics & Informatics (2007) 285, 295

[623] COM(96) 483, Green Paper on the protection of minors and human dignity in audiovisual and information services, 16.10.97; COM(97) 487, Communication on Illegal and Harmful content on the Internet, 16.10.97; Recommendation 98/560/EC on the development of the competitiveness of the European audiovisual and information services industry by promoting national frameworks aimed at achieving a comparable and effective level of protection of minors and human dignity OJ L 270, 7.10.1998. See further COM(2004)0341 European Parliament legislative resolution on the proposal for a recommendation of the European Parliament and of the Council on the protection of minors and human dignity and the right of reply in relation to the competitiveness of the European audiovisual and information services industry– C6-0029/2004 – 2004/0117(COD).

[624] Convention for the Protection of Human Rights and Fundamental Freedoms, Article 10.

[625] See for instance *Animal Defenders International* (2013) ECHR 362, (2013) 57 EHRR 21. This was a European Court of Human Rights Grand Chamber majority (4-3) judgment in the case, which upheld the Ofcom UK ban on television or radio advertising by the animal-protection organization, on grounds that objectives were "wholly or mainly of a political nature": held that there was no breach of Article 10 in applying Communications Act 2003 Act S.321(2).

[626] See Hale, Brenda (2012) "Argentoratum Locutum: Is Strasbourg or the Supreme Court Supreme?" Human Rights Law Review 12 (1): 65-78 doi: 10.1093/hrlr/ngs001 Hale argued: "This was clearly an interference with freedom of speech, indeed freedom of political speech, which is the most important of the kinds of speech protected by Article 10. It would certainly not be tolerated in the United States".

[627] *Mediakabel BV v. Commissariaat voor de Media*, case C-89/04, 6 November 2002 Belgian Constitutional Court judgment, appealed in Judgment of the European Court of Justice, decision of 2 June 2005, [2005] ECR I-04891.

[628] Dissenting justices Tulkens, Spielmann, Laffranque argued: "The more convincing the general justifications for the general measure, the less importance the Court will attach to its impact" citing *Murphy v. Ireland* (2003) [2003] ECHR 352, (2004) 38 EHRR 13, and *TV Vest AS v. Norway*, no. 21132/05 (2009) 48 EHRR 51 (Chamber, First Section). Further: "Government justified the contested measure by, in particular, need to protect electoral process as part of the democratic order, (*Bowman v. UK* (1998‑I) ECHR 4, Court accepted that a statutory control of the public debate was necessary given the risk posed to the right to free elections…But prohibition in question is not limited to electoral periods, the Bowman judgment and reasoning based on electoral process are of little bearing in this case (TV Vest §66)".

that the onus of regulation needs to be reversed based on the precautionary principle, and that broadcast electoral advertising rules could be extended to the Internet without infringing Article 10.

The extension of broadcast rules to non-broadcast content, whether text-based or in any case at the user's individual choice, would be a significant step that would in all likelihood increase the concentration of online communication in the hands of the largest platforms that can employ economies of scale: deploying proprietary filters to remove harmful content. Examples from the Internet include the attempts to prevent child pornography and terrorist video distribution, as well as copyrighted files. In each case, the use of technologies such as checking hash values in theory permitted removal before publication by the platforms deploying the technology, specifically YouTube and Facebook. In practice, the proliferation of content was restricted but by no means prevented by such technological intervention.[629]

The European Commission has recognized that the platform liability position is becoming very uncertain for platforms, and has agreed to review Notice and Take Down procedures in the E-Commerce Directive (ECD)[630]. It issued a 2013 working paper on its progress[631], 2016 proposals on the Digital Single Market Strategy[632], which may result in legislative proposals in the 2019-24 legislative period.

As platforms await a legislative option, they are also threatened by the possibility of filtering offered by the CJEU case of *Eva Glawischnig-Piesczek v Facebook Ireland*, decided on 3 October 2019[633]. It is useful to briefly explain the decade of prior CJEU case law in legal liability of intermediary platforms, whether as now for disinformation laws or earlier cases involving copyright violations. In *Scarlet Extended*, the CJEU had to balance rights holders against access providers and users' rights[634]. The CJEU recognized that the risk of preventing access to lawful content through over-blocking or over-filtering is a relevant factor to take into account. *Scarlet Extended* is an extension of the earlier CJEU reasoning in *Promusicae*[635]: the Belgian court ordered an access provider to filter all traffic for copyright infringement and 'pay for the privilege' of enforcing copyright on behalf of rights holders. Paragraphs 43–44 in *Scarlet Extended* are critical for general guidance:

> "The protection of the right to intellectual property is indeed enshrined in Article 17(2) of the Charter of Fundamental Rights of the European Union ('the Charter'). There is, however, nothing whatsoever in the wording of that provision or in the Court's case-law to suggest that that right is inviolable and must for that reason be absolutely protected. As paragraphs 62 to 68 of the judgment in *Promusicae* make clear, the protection of the fundamental right to property, which includes the

[629] See for example the difficulties encountered by platforms attempting to restrict sharing of video footage of the Christchurch shootings: Herne, A. & Waterson, J. (2019) 'Social Media Firms Fight to Delete Christchurch Shooting Footage', The Guardian, 19 March, https://www.theguardian.com/world/2019/mar/15/video-of-christchurch-attack-runs-on-social-media-and-news-sites

[630] 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') OJ L 178/1 pp.1–16, 17 July 2000

[631] European Commission "Report on the implementation of the e-commerce action plan" 23/04/2013 SWD(2013) 153 final.

[632] European Commission (2016) Proposal for a Regulation of the European Parliament and of the Council on addressing geo-blocking and other forms of discrimination based on customers' nationality, place of residence or place of establishment within the internal market and amending Regulation (EC) No 2006/2004 and Directive 2009/22/EC, 25 May.

[633] Case C-18/18 *Eva Glawischnig-Piesczek v Facebook Ireland Limited* ECLI:EU:C:2019:458 decided 3 October 2019.

[634] Case C-70/10 *Scarlet Extended SA v Société Belge des auteurs, compositeurs et éditeurs (SABAM)* OJ C 113, 1 May 2010: 20–20. Decided 24 November 2011, OJ C 25/6, 28 January 2012

[635] Case C-275/06 *Productores de Música de España (Promusicae) v Telefónica de España SAU*, judgment of 29 January 2008 [2008] ECR I271 at para 70

rights linked to intellectual property, must be balanced against the protection of other fundamental rights."

The CJEU stated that the Belgian injunction in issue would be a serious infringement of the freedom of the access provider concerned to conduct its business, since it would require it to install a complicated, costly, permanent computer system at its own expense. The Belgian court's order would have emasculated Article 15 of the ECD. This reasoning was reconfirmed and extended from access to social networking platforms by the February 2012 decision in *SABAM v Netlog*.[636] *Scarlet Extended* is a short decision (as is *Netlog*), and the question asked was set at the most extreme end of the scale, an injunction that was: (a) preventative; (b) entirely at the access provider's expense; (c) for an unlimited period; (d) applied to all customers indiscriminately; (e) for all kinds of communications. Useful guidance for national courts in the judgment included that the complexity/cost of the proposed Belgian system weighed against it, that Internet Protocol addresses are personal data, that the Belgian injunction was overbroad and could interfere with lawful as well as unlawful use.[637] Confirmation that IP addresses may be considered personal data arrived in 2016[638]. The remedy of URL blocking in *Scarlet* is indiscriminate, whereas the United Kingdom 'Cleanfeed' system deployed in the *Newzbin2* judgment of the English High Court was already in place and the cost essentially negligible.[639] As a result UK rights holders requested ISPs to block access to the file sharing website Pirate Bay.[640] Other EU countries have also seen successful applications for injunctions against ISPs.[641] The law as it stands awaits the European institutions' decisions in 2020 on how to reform the ECD, which will affect the liability environment for platforms in their disinformation polices, together with other content issues such as copyright.

While many previous media law techniques are inappropriate for online social media platforms, and some of these measures were abused by governments against the spirit of media pluralism, legislators need to consider which regulatory measures may protect freedom of information and expression by providing a bulwark against disinformation.

In Section 3, we will argue that AI is in the short to medium term highly unlikely to replace human judgement, and there is no possibility of restricting disinformation at source such that no-one views it. A key lesson from the peer-to-peer file sharing discussion has been that alternate forms of offline digital distribution are powerful replacements for online sharing: copyrighted material and other forms of content can easily be shared via a cheap ubiquitous technology such as a USB key or other form of portable storage, with much less user education than using a 'dark net' encrypted service such as a Tor relay.

3. Machine Learning and AI as 'Solutions' for Disinformation

Written evidence of disingenuous or 'fake news' is as old as the cuneiform tablets of Hammurabi.[642] Technological solutions risk threatening freedom of speech and media pluralism.

---

[636] Case C-360/10 *SABAM v Netlog*, 2 C.M.L.R. 18. 3 (2012) decided 16 February.

[637] See C-70/10 *Scarlet*, at paragraph 52: 'injunction could potentially undermine freedom of information since that system might not distinguish adequately between unlawful content and lawful content, with the result that its introduction could lead to the blocking of lawful communications'. See further *SABAM v Netlog*, at paragraphs 36–38.

[638] Case 582/14 *Patrick Breyer v Germany*, decided 19 October 2016, ECLI:EU:C:2016:779, 1 WLR 1569 (2017).

[639] *Twentieth Century Fox Film Corporation and Others v British Telecommunications Plc (No 2)* [2011] EWHC 2714 (Ch) 26 October 2011. This was the first order in the UK under Section 97A of the Copyright Designs and Patents Act 1988, which implements Article 8(3) of Directive 2001/29/EC, though it was already possible under UK law to seek injunctions against intermediaries.

[640] *Dramatic Entertainment Limited v British Sky Broadcasting Limited* [2012] EWHC 1152 (Ch) 2 May at: <http://www.bailii.org/ew/cases/EWHC/Ch/2012/1152.html>.

[641] Court of Appeal of Antwerp Case 2010/AR/2541 *VZW Belgian Anti-Piracy Federation v NV Telenet* 26 September 2011.

[642] Discussed in Enriques, L. *Financial Supervisors and RegTech: Four Roles and Four Challenges* (Oxford University, Business Law Blog, 9 Oct 2017) http://disq.us/t/2ucbsud

The problem has become far more visible and arguably acute online; social networks such as Facebook, YouTube, Twitter and WhatsApp allow information, authentic or otherwise, to spread globally and instantly. Hildebrandt explains the scale and scope that can create disinformation problems in social media platforms:

> "Due to their distributed, networked, and data-driven architecture, platforms enable the construction of invasive, over-complete, statistically inferred, profiles of individuals (exposure), the spreading of fake content and fake accounts, the intervention of botfarms and malware as well as persistent A/B testing, targeted advertising, and automated, targeted recycling of fake content (manipulation)."[643]

She warns that we must avoid the machine learning version of the Thomas self-fulfilling prophecy theorem – that "if a machine interprets a situation as real, its consequences becomes real".[644]

Within machine learning techniques that are advancing towards AI, automated content recognition technologies are textual and audio-visual analysis programmes that are trained to identify potential 'bot' accounts and unusual potential disinformation material[645]. In this article, AI refers to the use of automated techniques in the recognition and moderation of content and accounts, to assist human judgement.[646] Moderating content at larger scale requires AI as a supplement to human moderation (editing).[647] The shorthand 'AI' is used in the remainder of the article to refer to both these technologies. Where necessary, we specify which of the two (recognition or moderation) is implied.

Can AI solve the fake news problem? One argument being put forward by the owners of online platforms is that new technologies can solve the very problems they create. Chief among those technologies is machine learning or AI, alongside user reporting of abuse. However the notion that AI is a 'miracle cure', the panacea for fake news, is optimistic at best. Platforms argue that the use of automated content filtering systems, that use algorithmic processes to identify harmful content, provide a means for effective self-regulation by platforms. AI algorithms cannot be the only way to regulate content in future. Automated technologies such as AI are not a silver bullet for identifying illegal or "harmful" content. They are limited in their accuracy, especially for expression where cultural or contextual cues are necessary. The illegality of terrorist or child abuse content is far easier to determine than the boundaries of political speech or originality of derivative (copyrighted) works. The European Commission consultations on online platforms, assessment of the formally self-regulatory Code of Conduct fighting hate speech, and its overall

---

[643] Hildebrandt, M. 'Primitives of Legal Protection in the Era of Data-Driven Platforms', Georgetown Law Technology Review 2(2) (2018) at p. 253 footnote 3.

[644] Merton, R.K. 'The Self-Fulfilling Prophecy', *The Antioch Review 8(2),* (1948) 193-210.

[645] Artificial Intelligence refers to advanced forms of machine learning, generally classified as algorithmic processes powered by advanced computing techniques such as neural networks and including in particular Deep Learning. The technical literature is vast, but of relevance, see Klinger, J., Mateos-Garcia, J.C., and Stathoulopoulos, K. *Deep Learning, Deep Change? Mapping the Development of the Artificial Intelligence General Purpose Technology,* (2018) DOI: http://dx.doi.org/10.2139/ssrn.3233463. See also Zuckerberg, M. "A Blueprint for Content Governance and Enforcement", *Facebook Notes*, (15 Nov 2018) https://www.facebook.com/notes/mark-zuckerberg/a-blueprint-for-content-governance-and-enforcement/10156443129621634/ stating: "Some categories of harmful content are easier for AI to identify, and in others it takes more time to train our systems. For example, visual problems, like identifying nudity, are often easier than nuanced linguistic challenges, like hate speech". See also Chang, J., Boyd-Graber, J., Wang, C., Gerrish, S., and Blei, D. "Reading Tea Leaves: How Humans Interpret Topic Models", in Y. Bengio, D. Schuurmans, J. Lafferty, C. Williams, and A. Culotta (Eds.) *Advances in Neural Information Processing Systems* (Cambridge, MA: MIT Press 2009), pp. 288–96; Monroe, B., Colaresi, M., and Quinn, K. "Fightin' Words: Lexical Feature Selection and Evaluation for Identifying the Content of Political Conflict", *Political Analysis 16(4)* (2008) 372-403; Azevedo, L. "Truth or Lie: Automatically Fact Checking News", in *Companion Proceedings of The Web Conference 2018 (WWW '18),* International World Wide Web Conferences Steering Committee, Geneva, Switzerland, (2018) pp. 807-811, DOI: https://doi.org/10.1145/3184558.3186567

[646] See Epstein, R. & Robertson, R.E. '"The Search Engine Manipulation Effect (SEME) and its Possible Impact on the Outcomes of Elections", *112 Proc Nat'l Acad. Sci.* (2015) E4512

[647] Klonick, K. 'Why The History Of Content Moderation Matters', *Content Moderation at Scale 2018 Essays: Techdirt*, (2018) https://www.techdirt.com/articles/20180129/21074939116/why-history-content-moderation-matters.shtml

Digital Single Market strategy all arguing for greater "responsibility" by online platforms[648]. This means faster private enforcement which can also be seen as censorship given the lack of 'put back' appeal guarantees[649].

One of the problems is that they are responding to a perceived need from politicians to remove more content, rather than addressing fair process and due process. The informal incentive structure may require platforms to demonstrate to politicians how much content they have removed, when a very important factor in accountability for legal content posted may be "examples of successful appeals to put content back online". While the involvement of private parties in EU administrative governance has the clear advantage of delivering policies which are based on the expertise of the regulatees, private-party rule-making raises significant concerns in terms of its legitimacy. The EU response has been to effectively ignore the inconvenience that Internet self- or co-regulation is private censorship of free speech. The UK Parliament Artificial Intelligence (AI) Committee reported on some of these issues in 2017[650]. There are an enormous number of false positives in taking material down, which need human intervention to analyze. Google and Facebook announced in 2018 their intention to employ 50,000 more people as content moderators (subcontracted to so-called Mechanical Turks). 'Mechanical Turks' are people employed—subcontracted, typically—to carry out these activities[651], in parts of the world where their own cultural understanding of the content they are dealing with may not be ideal[652]. Subcontracting to people on very low wages in locations other than Europe is a great deal cheaper than employing a lawyer to work out whether there should be an appeal to put content back online.

Technical research into disinformation has followed several tracks:

- identifying and removing billions of bot as distinct from human accounts[653];
- identifying the real world effects of Internet communication on social networks[654],

[648] COM (2015) 192 A Digital Single Market Strategy for Europe, final, 6 May 2015, para. 3.3. EC (2015) Public Consultation on the Regulatory Environment for Platforms, Online Intermediaries, Data and Cloud Computing and the Collaborative Economy available at https://ec.europa.eu/digital-single-market/news/public-consultation-regulatory-environment-platforms-online-intermediaries-data-and-cloud; EC (2016) Fighting Illegal Online Hate Speech: First Assessment of the New Code of Conduct, available at http://ec.europa.eu/newsroom/just/item-detail.cfm?item_id=50840 COM (2016) 288 Online Platforms and the Digital Single Market: Opportunities and Challenges for Europe, 25 May 2016, p. 9

[649] See Frosio, Giancarlo F. (2017) 'From horizontal to vertical: an intermediary liability earthquake in Europe' 12 Journal of Intellectual Property Law and Practice 565, 575. See European Commission (2016) 'Full Report on the Results of the Public Consultation on the Regulatory Environment for Platforms, Online Intermediaries and the Collaborative Economy' 25 May.

[650] House of Lords (2017) AI Select Committee: AI Report Published https://www.parliament.uk/business/committees/committees-a-z/lords-select/ai-committee/news-parliament-2017/ai-report-published/ (note the report is published in non-standard URL accessed from this link)

[651] Hara, Kotaro; Adams, Abi; Milland, Kristy; Savage, Saiph; Callison-Burch, Chris; Bigham, Jeffrey (2017) "A Data-Driven Analysis of Workers' Earnings on Amazon Mechanical Turk" eprint arXiv:1712.05796 Conditionally accepted for inclusion in the 2018 ACM Conference on Human Factors in Computing Systems (CHI'18) Papers program

[652] See Ross, J., Irani, L., Silberman, M., Zaldivar, A., & Tomlinson, B. (2010). Who are the crowdworkers?: shifting demographics in mechanical turk. In CHI'10 extended abstracts on Human factors in computing systems (pp. 2863-2872). Association of Computing Machinery. See effects in Youtube Transparency Report (2018) https://transparencyreport.google.com/youtube-policy/overview

[653] Gilani, Z., Farahbakhsh, R., Tyson, G., Wang, L., and Crowcroft. J. (2017) 'Of Bots and Humans (on Twitter)', in *ASONAM '17 Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining,* pp. 349-354; Perez, B., Musolesi, M., and Stringhini, G. (2018) 'You are Your Metadata: Identification and Obfuscation of Social Media Users using Metadata Information', *ICWSM*.

[654] Including the 'Dunbar number' of friends that can be maintained, which has not measurably increased with the Internet: Dunbar, R. I. M. (2016) 'Do Online Social Media Cut Through the Constraints that Limit the Size of Offline Social Networks?', *Royal Society Open Science 2016(3)*, DOI: 10.1098/rsos.150292. Quercia, D., Lambiotte, R., Stillwell, D. Kosinski, M., and Crowcroft, J. (2012) 'The Personality of Popular Facebook Users', in *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work (CSCW '12),* pp. 955-964, https://doi.org/10.1145/2145204.2145346

- assessing the impact of disinformation via media consumption and electoral outcomes[655];

- researching security threats from disinformation;

- researching discrimination and bias in the algorithms used to both propagate and increasingly to identify and/or disable disinformation[656].

Online disinformation consumption includes that of video news and newspapers, whose readerships have largely migrated online,[657] but also images and amateur montages of video ('deep fakes') that are far harder to detect as disinformation. Textual analysis of Twitter or news sites can only explore the tip of the iceberg of disinformation, as video and images are much more difficult to examine comprehensively. Partial evidence of AI effectiveness is supplied by corporates. Facebook stated in 2018 that its automated systems detect 99% of the terrorism-related content it removes, as well as 96% of nude images and 52% of hate speech.[658] It has also been reported to automatically detect "nearly 100 percent of spam… 98.5 percent of fake accounts… and 86 percent of graphic violence-related removals".[659]

This evidence of AI removals is only unaudited company claims. Note Facebook's AI claims to detect "just 38 percent of the hate speech-related posts it ultimately removes, and at the moment it doesn't have enough training data for the AI to be very effective outside of English and Portuguese".[660] Researchers have claimed that trained algorithmic detection of fact verification may never be as effective as human intervention, with serious caveats (each has accuracy of only 76%): "future work might want to explore how hybrid decision models consisting of both fact verification and data-driven machine learning judgments can be integrated".[661] This is a sensible approach where resources allow for such a wide spectrum of solutions.

AI therefore cannot be the only way to regulate content in future.[662] Subcontracting to people on very low wages in locations other than Europe is a great deal cheaper than employing a lawyer to work out whether there should be an appeal to put content back online. The current incentive structure is for platforms to demonstrate how much content they have removed, when a very important factor may be examples of successful appeals to 'put back' legitimate content online.[663] Content moderation at scale still needs human intervention to interpret AI-flagged content.

A satisfactory solution to algorithmic transparency might be the ability to replicate the result that has been achieved by the company producing the algorithm. Transparency and explanation is necessary, but it is a small first step towards better regulation.[664] Veale, Binns and Van Kleek

---

[655] Zannettou, S. et al. (2018) *Disinformation Warfare: Understanding State-Sponsored Trolls on Twitter and Their Influence on the Web*, arXiv:1801.09288v1

[656] Alexander J., and Smith, J. (2011) 'Disinformation: A Taxonomy', *IEEE Security & Privacy 9(1),* 58-63, doi: 10.1109/MSP.2010.141; Michael, K. (2017) 'Bots Trending Now: Disinformation and Calculated Manipulation of the Masses [Editorial]', *IEEE Technology and Society Magazine 36(2),* 6-11, doi: 10.1109/MTS.2017.2697067

[657] Nielsen, R.K. and Ganter, S. (2017) 'Dealing with Digital Intermediaries: A Case Study of the Relations Between Publishers and Platforms', *New Media & Society 20(4),* 1600-1617, doi: 10.1177/1461444817701318

[658] Zuckerberg, M. (2018) 'A Blueprint for Content Governance and Enforcement', 15 November, https://www.facebook.com/notes/mark-zuckerberg/a-blueprint-for-content-governance-and-enforcement/10156443129621634/?hc_location=ufi

[659] Koebler, J., and Cox, J. (23 Aug 2018) 'The Impossible Job: Inside Facebook's Struggle to Moderate Two Billion People', *Motherboard*, https://motherboard.vice.com/en_us/article/xwk9zd/how-facebook-content-moderation-works

[660] Koebler and Cox (2018)

[661] Perez-Rosas, V., Kleinberg, B. Lefevre, A. and Mihalcea, R. (2018) *Automatic Detection of Fake News*, http://web.eecs.umich.edu/~mihalcea/papers/perezrosas.coling18.pdf

[662] Schaake, M. (2018) 'Algorithms Have Become So Powerful We Need a Robust, Europe-Wide Response', *The Guardian* https://www.theguardian.com/commentisfree/2018/apr/04/algorithms-powerful-europe-response-social-media

[663] Google (2018) *YouTube Transparency Report*, https://transparencyreport.google.com/youtube-policy/overview

[664] Edwards, L. and Veale, M. (2017) *Slave to the Algorithm? Why a 'Right to Explanation' is Probably Not the Remedy You are Looking for*, https://ssrn.com/abstract=2972855. Erdos, D. (2016) 'European Data Protection Regulation and Online New Media: Mind the Enforcement Gap', *Journal of Law and Society 43(4)* 534-564, http://dx.doi.org/10.1111/jols.12002

explain how to move beyond transparency and explicability to replicability: to be able to run the result and produce the answer that matches the answer they have.[665] Replicability would be the ability to look at the algorithm in use at the time and, as an audit function, run it back through the data to produce the same result. It is used in medical trials as a basic principle of scientific inquiry. It would help to create more trust in what is otherwise a black box that users and regulators simply have accept.

Hildebrandt explains that "data-driven systems parasite on the expertise of domain experts to engage in what is essentially an imitation game. There is nothing wrong with that, unless we wrongly assume that the system can do without the acuity of human judgment, mistaking the imitation for what is imitated".[666] Some of the claims that AI can 'solve' the problem of disinformation do just that. Over time, AI solutions to detect and remove illegal/undesirable content are becoming more effective, but they raise questions about who is the 'judge' in determining what is legal/illegal, and undesirable in society. Underlying AI use is a difficult choice between different elements of law and technology, public and private solutions, with trade-offs between judicial decision-making, scalability, and impact on users' freedom of expression.

In Section 4 we explore the options for dealing with this tool in analyzing disinformation, where an imitation game is insufficient to identify truth and falsehood, and human intervention on a large scale is required.

4. Options for regulating AI in disinformation introduced

In this Section, we explore the policy options that are available in some depth, and which would form the basis of EU legal policy towards disinformation[667]. Policy options have moved beyond the 'Goldilocks' theory of a three card trick (one too hot, one too cold, one just right) to encompass a range of self- and co-regulatory options[668].

[665] Veale, M., Binns, R., and Van Kleek, M. (2018) 'The General Data Protection Regulation: An Opportunity for the CHI Community? (CHI-GDPR 2018)', *Workshop at ACM CHI'18,* 22 April 2018, Montreal, arXiv:1803.06174

[666] The imitation game is often known as the Turing test, after Turing, A.M. (1950) *'Computing Machinery and Intelligence'*, *Mind 49*, 433-460.

[667] European Union, Inter-institutional agreement on better law-making, OJ L 123, 12.5.2016, p. 1, at https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016Q0512(01)&from=EN. See also Council of the European Union, Impact Assessment - Indicative guidance for Working Party Chairs, Brussels, 9 June 2016, 9790/16, at http://data.consilium.europa.eu/doc/document/ST-9790-2016-INIT/en/pdf

[668] Dunlop, C., & Radaelli, C. "Impact Assessment in the European Union: Lessons from a Research Project" European Journal of Risk Regulation, 6(1) (2015). 27-34. doi:10.1017/S1867299X00004256; Damonte, A., Dunlop, C., & Radaelli, C. "Regulatory Reform: Research Agendas, Policy Instruments and Causation" European Journal of Risk Regulation, 8(1), (2017) 72-76. doi:10.1017/err.2016.12

Figure: Reeve model of regulatory pyramid[669]

4.1 Co-regulation as a Regulatory Technique explained

Internet self- and co-regulatory arrangements have a legal foundation, and specific legal constraints or conditions to be respected. The Internet developed self-regulation based on the Codes of Conduct (CoC) and Terms of Use (ToU) that early Internet users employed, in the scientific institutions that first developed the protocols and social standards[670]. The use of such ethical standards is more corporate social responsibility than law[671]. Given the rapid growth, complex inter-relationships and dynamic changes that have taken place in the current century, governments have broadly accepted that a more flexible and innovation-friendly model of regulation is required[672]. Cafaggi stated in regard to sanctioned regulation: "An intermediate hypothesis between delegated private regulation and ex post recognized private regulation is that in which private regulation, produced by the private or self-regulator, has to be approved by a public authority to become effective."[673] It is a pragmatic acceptance that the models used for regulation should be as flexible as possible, to permit significantly greater user innovation and freedom than with other types of communications (notably telecoms and broadcasting)[674]. This includes using both hard and 'soft law' forms of regulation[675]. The Internet self-regulatory

[669] Reeve, B'. "The Regulatory Pyramid Meets the Food Pyramid: Can Regulatory Theory Improve Controls on Television Food Advertising to Australian Children?" Journal of Law and Medicine 19(1) (2011) 128-46.
[670] Werbach, Kevin *Digital Tornado: The Internet and Telecommunications Policy*, Office of Plans and Policies Working Paper 29. (Washington: Federal Communications Commission 1997)
[671] Abbott K, Snidal D. "The Governance Triangle: Regulatory Standards Institutions and the Shadow of the State" Chapter 2 in Mattli W, Woods N (eds) *The Politics of Global Regulation*, (Princeton University Press 2009), pp.44-88. Abbott K, Snidal D. (2004) "Hard and soft law in international governance", International Organization 54, pp.421-422; Helin, S., & Sandström, J. "An inquiry into the study of corporate codes of ethics", Journal of Business Ethics 75 (2007), pp.253–271. Higgs-Kleyn, N., & Kapelianis, D. "The role of professional codes in regulating ethical conduct", Journal of Business Ethics, 19 (1999), 363–374. Vrielink, Mirjan Oude, Cor van Montfort, Meike Bokhorst Codes as hybrid regulation, ECPR Standing Group on Regulatory Governance, (June 17-19 2010), Dublin.
[672] Generally on the role of smart regulation, see Gunningham N., Rees J. "Industry Self-regulation: An Institutional Perspective", Law & Policy 19(4) (1997); Gunningham, N. and Grabosky, P. *Smart Regulation: Designing Environmental Policy*, (Oxford University Press 1998); Gaines, Sanford E. and Cliona Kimber "Redirecting Self-Regulation", Env. Law 13 (2001), p.157. More generally, see Black, J. "Constitutionalising Self-Regulation", Modern Law Review, Vol. 59, No. 1, (1996) pp. 24-55 at p.59; Black, J. *Managing the Financial Crisis – The Constitutional Dimension*, LSE Legal Studies Working Paper No. 12/2010 (2010).
[673] See Cafaggi, F. (2006) Rethinking private regulation in the European regulatory space, *EUI Working Paper LAW No. 2006/13*, at http://cadmus.eui.eu/bitstream/handle/1814/4369/LAW2006.13.PDF?sequence=1 at p.24.
[674] Goldsmith, J. and Wu, T. (2006) Who Controls the Internet? Oxford: Oxford University Press.
[675] On the role of 'soft law' more generally, see Senden, L. (2005) Soft Law, Self-Regulation and Co-Regulation In European Law: Where Do They Meet? Electronic Journal of Comparative Law, vol. 9.1 at http://www.ejcl.org/91/abs91-3.html. Cosma, H. & Whish, R. (2003) Soft Law in the Field of EU Competition Policy, European Business Law Review, Vol. 14,. Pt. 1 pp.25-56. Hodson, Dermot and Imelda Maher (2004) Soft law and sanctions: economic policy co-

paradigm has been increasingly challenged by the growth and evolution of the Internet and associated technologies including cloud computing, blockchains, smart contracts and Artificial Intelligence[676].

Formal co-regulation comprises a regulatory system in which the regulator is independent from government, making regulation subject to prior approval of codes of conduct, systems for funding and independent appeal[677]. In Germany, this is known as regulated self-regulation.[678] This is a hybrid system subject to statutory control. Examples from the internet regulatory ecosystem are:

- Nominet, largest European Domain Name System Registry operator, which operates the .uk domain since 1996, under ultimate control by government via Digital Economy Act 2010;[679]

- EURID which regulates and operates registries under the .eu domain since 2003.[680]

European co-regulation in wider consumer protection legislation was detailed in 2002, and became official policy in December 2003, with the Inter-Institutional Agreement on Better Law-Making, which defined co-regulation[681]. Self-regulation is viewed as making standards and practices across industry that the Commission, or a Member State, views agnostically in pre-legislative or legislative terms. Government then analyze the extent to which self-regulation approaches the standards of 'representativeness' which co-regulation is meant to demonstrate as a best practice. The Inter-Institutional Agreement confirmed in 2003 that forms of regulation short of state regulation: "will not be applicable where fundamental rights or important political options are at stake or in situations where the rules must be applied in a uniform fashion in all Member States." The European Commission in 2005 went on to analyze co-regulation in terms of 'better regulation' (COM/2005/97). This was immediately made part of internal EC practice in the Impact Assessment Guidelines (SEC/2005/791) which the Commission must follow before bringing forward a new legislative or policy proposal, updated in 2015[682].

---

ordination and reform of the Stability and Growth Pact, Journal of European Public Policy, Volume 11 Issue 5 pp.798–813

[676] Werbach, Kevin (2018) The Blockchain and the New Architecture of Trust, MIT Press; Cohen, Julie E. (2016) The Regulatory State in the Information Age, 17 Theoretical Inq. L. 369-414; Finck, Michèle "Digital Co-Regulation: Designing a Supranational Legal Framework for the Platform Economy", European Law Review (2018): https://ssrn.com/abstract=2990043; Hildebrandt, Mireille "Law as Information in the Era of Data-Driven Agency", Modern Law Review, 79: (2016), 1–30, at doi:10.1111/1468-2230.12165; Werbach, Kevin "Contracts Ex Machina", 67 Duke LJ (2017) 101

[677] Marsden, C. T. (2011) Internet Co-Regulation: European Law, Regulatory Governance and Legitimacy in Cyberspace, Cambridge University Press, explores 25 such self- and co-regulatory schemes.

[678]See Hoffmann-Riem, W. (2001) *Modernisierung in Recht und Kultur,* Frankfurt: Suhrkamp; Huyse, L., and Parmentier, S. (1990) 'Decoding Codes: The Dialogue between Consumers and Suppliers through Codes of Conduct in the European Community', *Journal of Consumer Policy 13(3),* 253–272, at 260; Joerges, C., Meny, Y. and Weiler, J.H.H. (Eds., 2001) *Responses to the European Commission's White Paper on Governance,* European University Institute; Kleinstuber, H. (2004) 'The Internet between Regulation and Governance', in *Organisation for Security and Co-operation in Europe, The Media Freedom Internet Cookbook*, pp61-100; Latzer, M., Just, N., Saurwein, F., and Slominski, P. (2003) 'Regulation Remixed: Institutional Change through Self- and Co-Regulation in the Mediamatics Sector', *Communications and Strategies, 50(2),* 127-157.

[679]See Marsden, C. (2011), N. 60 at p61. By 2018, there were 12 million UK domains registered, see Nominet (2018), *UK Domains*, https://www.nominet.uk/uk-domains/

[680]Regulation (EC) No 874/2004 Laying Down Public Policy Rules concerning the Implementation and Functions of the .eu Top Level Domain and the Principles governing Registration

[681] Inter-Institutional Agreement on Better Law-Making, Official Journal of the European Union December 2003, 2003/C 321/01 at http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2003:321:0001:0005:EN:PDF See variously COM 2002/704 Towards a reinforced culture of consultation and dialogue - General Principles and minimum standards for consultation of interested parties by the Commission, 11 December
COM(2002) 275 European Governance: Better Lawmaking, 5 June; COM/2002/0278 Action plan: Simplifying and improving the regulatory environment. Later plans contained more detail: see COM(2009) 504 Report From The Commission On Subsidiarity And Proportionality (16th report on Better Lawmaking) at http://ec.europa.eu/governance/better_regulation/documents/com_2009_0504_en.pdf and COM (2005) 97 Better Regulation for Growth and Jobs in the EU.

[682] European Commission (2015) Better regulation for better results - An EU agenda, 19 May, SWD(2015) 110 final

This European regulatory activity in defining co- and self-regulation was matched by its continued research into the impact of the Internet and its own legislative and policy initiatives since 1998. The European Commission thus commissioned substantial independent research from 2001 onwards in assessing Internet regulation and the enforcement thereof by private actors[683]. Price and Verhulst examined private Internet enforcement via internal self-organisation: they identified increasing realism in recognising competition problems, emerging monopolies and dominance beginning to emerge in the early 2000s[684]. A 2004 report for the European Commission based on a three year study of private law enforcement concluded: "There is a danger that some aspects of internet self-regulation fail to conform to accepted standards. We recommend co-regulatory audit as the best balance of fundamental rights and responsive regulation" [685]. Latzer et al provided excellent analysis of the types of co-regulation beginning to develop, and their institutional path dependency[686]. Self Regulatory Organisations (SROs) generally form as single issue bodies, often crisis-driven, but then develop according to their institutional environment, for instance broadcast content self-regulation bodies can develop into video games, video film, or Internet content self-regulation. They note that there are different economic as well as political incentives for self-regulation, and analysis is needed with attention to the loss of constitutional guarantees[687]. Transparency and explanation by the SRO is necessary, but small first steps towards greater co-regulation[688]. Digital information policy is critically concerned with relationships between existing government-industry actors and 'prosumer' groups, whose role in production, distribution and consumption is growing rapidly, and whose motivations and activism are often non-monetary, requiring a more sophisticated interdisciplinary method for assessing contributions, motivations and sustainability of the 'prosumption economy', the growth of the virtual polity and social communities online, and a new prosumer law and policy to govern the regulation of the digital information ecology. This calls for a new form of consumer and citizen protection, which Brown and Marsden termed 'prosumer law'[689]. Helberger et al have called this 'networked consumer' law[690]. The European Commission has used the overarching objective of "a fair deal for consumers" online[691]. Commissioner Vestager has explained that Internet social networks are essentially addiction platforms[692].

---

[683] Tambini, D., Leonardi, D., and Marsden, Christopher T. (2008) Codifying Cyberspace: Communications Self-Regulation in the Age of Internet Convergence, Routledge: London, as well as Brown and Marsden (2013) supra. European Community of Practice Agora on Better Self and Coregulation (2013-17) : https://ec.europa.eu/digital-single-market/en/newsroom-agenda/event/cop-better-self-and-coregulation

[684] Price, M. and Verhulst, S. (2004) Self Regulation And The Internet, Kluwer Law International.

[685] Directorate-General for Communications Networks, Content and Technology (European Commission , Programme in Comparative Law and Policy (2004) *Self-Regulation of Digital Media Converging on the Internet: Industry Codes of Conduct in Sectoral Analysis, Final Report of IAPCODE Project for European Commission DG Information Society Safer Internet Action Plan*, 30 April, Section 12.7 at https://publications.europa.eu/en/publication-detail/-/publication/b7c998d9-75d6-464d-9d91-d59aa90a543c/language-en

[686] Latzer, Michael, Price, Monroe E., Saurwein, Florian, Verhulst, Stefaan G. *Comparative Analysis of International Co- and Self-Regulation in Communications Markets*, Research report commissioned by Ofcom, September, Vienna: ITA (2007) at www.mediachange.ch/media//pdf/publications/latzer_et_al_2007_comparative_analysis.pdf

[687] See regulators' analyses: Ofcom, *Criteria for promoting effective co and self-regulation: Statement on the criteria to be applied by Ofcom for promoting effective co and self-regulation and establishing coregulatory bodies* (2004) www.ofcom.org.uk/consult/condocs/coreg/promoting_effective_coregulation/co_self_reg.pdf. Office of Regulation Review *A Guide to Regulation, Second Edition,* December 1998 at www.pc.gov.au/orr/reguide2/reguide2.pdf; .

[688] Edwards, Lilian and Veale, Michael, Slave to the Algorithm? Why a 'Right to Explanation' is Probably Not the Remedy You are Looking for: Edwards, Lilian and Veale, Michael, "Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For" 16 Duke Law & Technology Review 18 (2017) http://dx.doi.org/10.2139/ssrn.2972855. Erdos, David (2016) European Data Protection Regulation and Online New Media: Mind the Enforcement Gap Journal of Law and Society, Vol. 43, Issue 4, pp. 534-564, http://dx.doi.org/10.1111/jols.12002

[689] Brown, I. and Marsden, C. *Regulating code: good governance and better regulation in the information age* (Cambridge, MA: MIT Press, 2013) at Chapter 8. See also Jasmontaite, L. "The European Data Protection Supervisor (EDPS) Opinion 4/2015 towards new digital ethics". European Data Protection Law Review (EDPL) 2(1), (2016) 93-96 at 95.

[690] Helberger, Natali, Borgesius, Frederik J. and Reyna, Agustin "The Perfect Match? A Closer Look at the Relationship between EU Consumer Law and Data Protection Law" Common Market Law Review, Vol. 54, No. 5 (2017)

[691] Vestager, M. (2018) "Competition and a fair deal for consumers online", Netherlands Authority for Consumers and Markets Fifth Anniversary Conference, The Hague, 26 April https://ec.europa.eu/commission/commissioners/2014-2019/vestager/announcements/competition-and-fair-deal-consumers-online_en

[692] See https://www.b.dk/globalt/eu-commissioner-margrethe-vestager-facebook-is-designed-to-create-addiction-like

Can social media platforms, addictive or not, be left to privately enforce the online regime, or will determined co-regulatory intervention make this happen? Two judgments of the European Court of Human Rights shed light on this. The first was an Estonian reference to the Grand Chamber, *Delfi*[693], in which a news website was made liable for the comments that were underneath the news article. It was fined for the comments, which led news websites across Europe to think that they would have to either pre-moderate, which would require a great deal of investment, or alternatively remove comments altogether. That case has since been followed by *MTE v. Hungary*, which restored some kind of balance, and came to a different conclusion on the facts, deciding that pre-moderation of comments was not required. As a lower chamber decision, it could not overturn *Delfi*[694]. If the law requires prior approval of comments, whether it be on Twitter, a news website or wherever else, that requires a great deal more investment, and websites may well choose to exclude all comments[695]. Revisiting the protections from liability for hosting third party content, as suggested by the new European Commission in 2019 inspired in part by disinformation threats[696], may cause the entire co-regulatory structure to partially unravel.

In the following section, the options are laid out in more detail.

4.2 Options for regulating AI in disinformation explained

Six options are provided for technical means to moderate and remove disinformation, ranging from Option 0 (no new regulation but, further research and analysis into current self- and state regulation) to Option 5 (specific legislative instruments):

- **Option 0**: Status quo, noting that this would entail permitting both 'natural' technical experiments in moderation, research into creating evidence-based policy as outlined above, and the legislative responses that already exist.

- **Option 1**: Non-audited self-regulation, with increasing industry-government coordination, but no sanction on those companies choosing not to cooperate in standards.[697]

- **Option 2**: Audited self-regulation, under which for instance the code of practice on disinformation would be subjected to formal published audit by a commonly agreed self-regulator.[698]

- **Option 3:** A formal self-regulator, recognised by the European institutions and ideally with funding separate from the industry.

- **Option 4**: Formal co-regulation, in which the regulator is independent from government yet subject to prior approval of codes of conduct, systems for funding and arbitration.

[693] Delfi AS v Estonia [GC], 64569/09 ECHR [2015]
[694] MTE v Hungary ECHR 22947/13. For case comment, see Bjarnadóttir, María Rún (2017) Case Law, Strasbourg: Einarsson v Iceland, Defamation on social media and Article 8, Inforrm Blog, 14 November, at https://inforrm.org/2017/11/14/case-law-strasbourg-einarsson-v-iceland-defamation-on-social-media-and-article-8-maria-run-bjarnadottir/
[695] Kerr, A. & Musiani, F. & Pohle, J. "Communication and internet policy: a critical rights-based history and future" Internet Policy Review, 8(1) (2019) DOI: 10.14763/2019.1.1395
[696] von der Leyen, Ursula *A Union that strives for more: My agenda for Europe*, p.13, at https://ec.europa.eu/commission/sites/beta-political/files/political-guidelines-next-commission_en.pdf
[697] Marsden, C. Internet Co-regulation: European Law, Regulatory Governance and Legitimacy in Cyberspace Cambridge University Press (2011), pp.107-113.
[698] Such as UK Safer Internet Centre (2018) for reporting and removing child sex abuse images online, https://www.saferinternet.org.uk/

- **Option 5**: Statutory regulation, in which a regulator is tasked to combat disinformation directly by licensing of content providers and their systems for content moderation. Current electoral and broadcast regulators already perform this function for offline media.

Note that the options are interdependent – where regulation is proposed, it sits atop a pyramid of activities including co-regulation, self-regulation, technical standards and individual company initiatives. There is no single option to solve the problem of disinformation. Given what AI use and abuse reveals about disinformation practices, potential actions are summarised in the table below.

Table: Typology of regulation and implications

Option and form of regulation	**Typology of regulation**	**Implications/Notes**
0 Status quo	Corporate social responsibility, single-company initiatives	Note that enforcement of the General Data Protection Regulation, AVMS Directive, and the proposed revised ePrivacy Regulation, would all continue and likely expand
1 Non-audited self-regulation	Industry code of practice, transparency reports, self-reporting	Corporate agreement on principles for common technical solutions
2 Audited self-regulation	European Code of Practice of September 2018; Global Network Initiative published audit reports	Open interoperable publicly available standard e.g. commonly engineered/designed standard for content removal to which platforms could certify compliance
3 Formal self-regulator	Powers to expel non-performing members, dispute resolution ruling/arbitration on cases	Commonly engineered standard for content filtering or algorithmic moderation. Requirement for members of self-regulatory body to conform to standard or prove equivalence. Particular focus on content 'put back' metrics and efficiency/effectiveness of appeal process
4 Co-regulation	Industry code approved by Parliament(s) or regulator(s) with statutory powers to supplant	Government-approved technical standard – for filtering or other forms of moderation. Examples from broadcast and advertising regulation
5 Statutory regulation	Formal regulation – tribunal with judicial review	National regulatory agencies – although note many overlapping powers between agencies on e.g. freedom of expression, electoral advertising and privacy

Option 0: Status quo

This option would entail permitting both 'natural' technical experiments in moderation, and the legislative responses that already exist, such as that of Germany's Network Enforcement Law (NetzDG)[699]. However, it would also rely on individual corporate efforts to enforce, rather than an industry self-regulation scheme or democratically legitimate institutional oversight. Individual users would continue to rely on companies' terms of service enforcement for their own and others' freedom of expression (with widely varying content standards, definitions of abusive/harmful content etc.).

Individual companies would continue to pursue disparate aims according to their own judgement of brand interest (e.g. Google decided not to accept political advertising during the 2018 referendum on the Thirty-sixth Amendment of the Constitution Act in Ireland, whereas Facebook only banned foreign actors' adverts). The idea that a multinational public social media company acts as its own government with its own 'supreme court' was promulgated by Mark Zuckerberg in April 2018,[700] but is clearly a case of corporate social responsibility over-reach.[701] However, much can be achieved using non-traditional regulatory tools to control AI use. A highly influential Shorenstein Center report for the Council of Europe, outlines responses by platforms, news providers and governments identified separately.[702]

The benefits of no regulation are the classic United States common law of the libertarian 'marketplace of ideas' to combat disinformation. However, the costs are that only research and evaluation could be carried out by government, with no carrot-and-stick threat to regulate. Sustainability would be jeopardised by any political calculation that disinformation has overwhelmed the media ecosystem's own established defences, and this article concludes that the 2016-17 electoral/referendum evidence shows substantial failures in the regulatory ecosystem for the media, notably with regard to bot accounts and unregulated online political advertising. An unregulated online free-for-all is unappealing to European policy makers.[703]

Much detailed internet regulation is self-regulation despite such profound constitutional issues of fundamental rights. This is because US companies have implemented in terms of service the 'negative liberty' framework of the US First Amendment which stops Congress intervening in the liberty of the press. By contrast, European law has positive obligations including Article 10 Paragraph 2 of the European Convention on Human Rights, permitting states to intervene to protect rights[704]. Despite US claims of the exceptionalism of free speech, Option Zero is not a realistic option for European legislators.

---

[699] Netzwerkdurchsetzungsgesetz (German Network Enforcement Act) 2017, see EU Code of Practice on Disinformation. Annex II Current Best Practices from Signatories of the Code of Practice (2018) https://ec.europa.eu/digital-single-market/en/news/code-practice-disinformation

[700]Kozlowska, H. (3 April 2018) 'Mark Zuckerberg Floated a 'Supreme Court' for Facebook. What Does That Mean?', *Quartz*, https://qz.com/1243203/mark-zuckerberg-floated-a-supreme-court-for-facebook-what-does-that-mean/

[701]On the role of multinationals in regulation generally, see Ruggie, J. (2018) 'Multinationals as Global Institution: Power, Authority and Relative Autonomy, *Regulation & Governance (2018)12*, 317–333, https://onlinelibrary.wiley.com/doi/pdf/10.1111/rego.12154

[702]Wardle, C. and Derakhstan, H. (2017) *Information Disorder: Toward an Interdisciplinary Framework for Research and Policy Making* (DGI(2017)09), Shorenstein Center on Media, Politics and Public Policy at Harvard Kennedy School for the Council of Europe, https://shorensteincenter.org/information-disorder-framework-for-research-and-policymaking

[703]Described by French former culture minister Jack Lang as 'the freedom of the fox in the barnyard': see Muravchik, J. (1998) *The Future Of The United Nations: Understanding The Past To Chart A Way Forward*, American Enterprise Institute for Public Policy Research, Washington, D.C., https://epdf.tips/the-future-of-the-united-nations-understanding-the-past-to-chart-a-way-forward.html, at p.85.

[704] Most recently, see Keller, Daphne, *Dolphins in the Net: Internet Content Filters and the Advocate General's Glawischnig-Piesczek v. Facebook Ireland Opinion*, Stanford Center for Internet and Society, September 4, 2019

This article therefore argues that the initiatives identified by Wardle et al. should be targeted and encouraged by European institutions, in the interests of a better approach to tackling disinformation[705]. It also proposes additional regulatory measures explained in Options 1-5 below. Option Zero is only effective if the disinformation problem is held to be capable of self-healing by market actors and individuals without the need for more formal coordination, investment or even direct regulation.

Option 1: Non-audited self-regulation

This option would increase platform activity compared with Option zero in terms of preventing immediate regulatory intervention, with increasing industry-government coordination, but no sanction on those companies choosing not to cooperate. Many examples can be found in the Shorenstein table above. Government and private industry research funding could be increased to encourage machine learning-based and other forms of content moderation.[706] The EU code of practice on disinformation proposed by companies under the aegis of the European Commission would continue to be developed. However, the lack of formalised transparency processes (other than reporting) makes this option ineffective and potentially damaging to the European policy process, and thus it is an unsatisfactory hybrid option as compared to Option Zero or Option 2.

The Santa Clara Principles for Content Moderation are a step towards Option 1. European Union funding for the World Wide Web Consortium is an example of technical sponsorship to help internet self-regulated standards.[707] In the AI space, standards for ethical algorithms are being developed by for instance the IEEE P7000 scheme,[708] but critics have pointed out that these ethical norms are the predecessor to legal standards.[709] Therefore, any ethical code that becomes an industry standard for certification, especially in an area affecting fundamental rights like algorithmically determined content recognition, is likely to lead to a call for legislative standards and enforcement.

Option 2: Audited self-regulation

Under audited self-regulation, the self-regulatory scheme is subject to regular (even annual) independent audit to ascertain the degree to which members are cohering to the criteria. For instance, the code of practice would be subjected to formal published audit by a commonly agreed self-regulator; an example is INHOPE, the pan-European hotline associated co-funded originally under the safer internet action plan.[710] Members of the EU High Level Expert Group on Disinformation argue that: "[f]act-checking technology has an important role to play, provided it is independent and free from any political influence. Platforms can provide client-based interfaces for control and guidance on selecting, for example, priorities in news searches and news feeds,

[705] Wardle et al (2017) n. 104.
[706] See for instance publications of the European Union funded ENCASE Social Computing project: https://encase.socialcomputing.eu/publications Zinonos, S., Tsirtsis, A., and Tsapatsoulis, N. (2018) 'Twitter Influencers or Cheated Buyers?', *IEEE Cyber Science and Technology Congress*; Mariconti, E. et al. (2018) "You Know What to Do": Proactive Detection of YouTube Videos Targeted by Coordinated Hate Attacks', *ArXiv*; Zannettou, S. et al. (2018) 'On the Origins of Memes by Means of Fringe Web Communities', *ACM Internet Measurement Conference (IMC)*; Zannettou, S. et al. (2018) 'The Web of False Information: Rumors, Fake News, Hoaxes, Clickbait, and Various Other Shenanigans', *ArXiv*; Founta, A-M. et al. (2018) 'A Unified Deep Learning Architecture for Abuse Detection', *ArXiv*; Founta, A-M. et al. (2018) 'Large Scale Crowdsourcing and Characterization of Twitter Abusive Behavior', *International AAAI Conference on Web and Social Media (ICWSM)*; Zannettou, S. et al. (2018) 'The Good, the Bad and the Bait: Detecting and Characterizing Clickbait on YouTube', *1st Deep Learning and Security Workshop, co-located with the 39th IEEE Symposium on Security and Privacy.*
[707] Marsden, C. (2011) n.60, pp. 107-113.
[708] IEEE (2018) *Global Initiative on Ethics of Autonomous and Intelligent Systems,* https://standards.ieee.org/industry-connections/ec/autonomous-systems.html
[709] @rcalo: 'Now that I'm on my high horse, let me *specifically disavow* @IEEEorg's efforts to create an ethical certification program. IEEE is an important organisation we should look to for thought leadership. But offering an ethical certification is as dangerous as it is premature.' (23 October 2018) https://twitter.com/rcalo/status/1054834789570633729
[710] UK Safer Internet Centre (2018).

diversity of opinions on consumer time lines and the re-posting of fact-checked information. Platforms need to be transparent about their algorithms".[711]

In the AI disinformation scheme, this audit could be undertaken by the industry body, or by a self-regulator from an associated industry, for instance broadcasting or games classification (see Option 3). The HLEG members argue that: "Google, Facebook and Twitter have now taken a public commitment to work with researchers who can independently assess the spread and impact of disinformation. The [EC disinformation] report specifically calls on major technology companies to provide data that would allow the independent assessment of efforts like Google's fact-check tags, Facebook's use of fact-checks as Related Articles or the downgrading of disinformation in the News Feed". Jiménez Cruz et al. argue for: "[t]he creation of a network of Research Centers focused on studying disinformation across the EU, [as the] current knowledge base is almost entirely focused on the United States data".[712] This is a vital area for further funded research by the European institutions.

The cost-benefit of audited self-regulation depends on the level of independence and rigour of the auditor function. It allows for flexible regulation, though efficiency depends on industry actors' commitment to the independence and rigour of the auditor in the absence of any penalty for lack of compliance, often a fatal failing.[713] Lower costs and more responsive regulation are possible, free riders are very likely to exist, though the scale of the larger platforms and the existing code of practice commitments may ensure greater scrutiny. In essence, Jiménez Cruz et al. argue that Option 2 is best suited to the current evidence, for a "structured process ahead that will document progress made and expose anyone not taking their responsibilities seriously".[714]

Feasibility and effectiveness depend on the implementation of audit. Sustainability of audited self-regulation is very low, given the possibilities for non-compliance identified above. Human rights challenges will exist even with an independent multi-stakeholder board, so that self-audit is inevitably judged inadequate and may be supplanted by more formal regulatory bodies. Risks and future uncertainties are thus very high, and there is no satisfactory example of audited self-regulation on the internet without the backstop of formal regulation. Take for example the time-limited Google Advisory Council on the Right to be Forgotten[715], a legal right which is subsequently subject to regulatory and court enforcement and was thus not an example of audited self-regulation. The Global Network Initiative claims such an audit function, but annual reports do not give detail such that it would satisfy these criteria.[716]

Option 3: Formal self-regulator

This regulator would be recognised by the European institutions and ideally with funding separated from the industry. Recognition does not signal statutory power to intervene or to direct the regulator, but does indicate that the institutions wish to guide the choice of self-regulatory scheme employed, short of intervention via legislation.

---

[711] Jiménez Cruz, C., Mantzarlis, A., Nielsen, R.K., and Wardle, C. (12 March 2018), 'Six Points from the EU Commission's New Report on Disinformation, *Medium,* https://medium.com/@hlegresponse/six-key-points-from-the-eu-commissions-new-report-on-disinformation-1a4ccc98cb1c

[712] Ibidem. Note a network of Centres on Internet and Society already exists, and is currently studying this area, with circa 35 European centres, chaired over time by Politecnico de Torino (NEXA Centre) and Humboldt University: see https://networkofcenters.net/centers

[713] In the expert interview, Monique Goyens (Director-General at European Consumer Organisation – BEUC, 31 August 2018) expressed it in the following way: 'I have been in the job of consumer activism for more than thirty years. I have seen a lot of self-regulation. I have not seen much that has worked.'

[714] Jiménez Cruz, C., Mantzarlis, A., Nielsen, R.K., and Wardle, C. (12 March 2018), n.113.

[715] Google (2015) *Google Advisory Council on the Right to be Forgotten*, https://archive.google.com/advisorycouncil/

[716] Global Network Initiative (2018) *Annual Report 2017: Reinforcing a Global Standard,* https://globalnetworkinitiative.org/global-network-initiative-annual-report-2017-reinforcing-a-global-standard/

An example is the Pan European Game Information (PEGI) scheme, under which 30,000 computer game products have been labelled and classified to indicate violence, sexual content, and other types of content that may give human dignity/child protection concerns, using the graphical warnings of the Netherlands *Kijkwijzer* scheme implemented by the Netherlands Institute for the Classification of Audio-visual Media, and the UK Video Standards Council.[717] App store games are regulated using the International Age Rating Coalition system.[718] PEGI is not formally regulated, but claims: "PEGI is used and recognised throughout Europe and has the enthusiastic support of the European Commission. It is considered as a model of European harmonisation in the field of the protection of children".[719]

Applied to AI and disinformation, this schematic would suggest a multistakeholder or at least EU institutions-industry dialogue establishing general principles applying to an AI regulator, while the self-regulator would set out details of the scheme design. Such principles may include, for instance, the principle that no account can be suspended without human intervention to correct for false positive identification of a bot account, and the potential for account holder appeal against such a deletion. As noted, the UN Special Rapporteur on Freedom of Opinion and Expression has recommended such a body to deal with online content moderation. Human-regulated AI is more likely to be guaranteed with robust co-regulation than self-regulatory schemes (see following section).

The cost-benefit of self-regulation is held in general to allow for very flexible regulation, though efficiency depends on industry actors confirming to the rating scheme. Lower costs and more responsive regulation are possible, though free riders who fail to conform fully may exist. Feasibility and effectiveness depend on the initial design, as well as the implementation of that design by the self-regulator. A problem can be that the lack of sanctions for inappropriate labelling or failure to conform to standards may not be subject to a robust system of audit and correction.

Sustainability of self-regulation is always an issue. Internet regulation is often implemented directly by legislatures due to particularly profound constitutional and human rights challenges including freedom of expression and prevention of harm, so that self-regulation is judged inadequate and supplanted by state regulatory bodies. Risks and future uncertainties are thus closely tied to the regulatory commitment to making self-regulation an end state (subject to satisfactory independent audit of procedures) rather than an interim measure.

Coherence with EU objectives are easier to assess with co-regulation than with self-regulation because the national statutory criteria establishing the co-regulator must conform to European law principles, and ex-post comparative evaluation across Member States can more easily be undertaken given these common criteria. The divergence of regulatory means used for areas such as child protection and video on demand over the two decades of European consumer internet law show that a level of co-existence of different regulatory schemes is possible with national differences.

Potential ethical, social and regulatory impacts revolve around the media pluralism dilemma. The fundamental rights issues with co-regulation are similar to those for less direct regulatory interventions – freedom of expression as a fundamental right may be held inappropriate for anything but state regulation, a constant issue in internet regulation.

[717] *Kijkwijzer* (2018) *Netherlands Institute for the Classification of Audio-visual Media,* http://www.kijkwijzer.nl/nicam and Pan European Game Information (2018) *How We Rate Games*, https://pegi.info/page/how-we-rate-games
[718] International Age Rating Coalition (2018) *How IARD Works,* http://www.globalratings.com/how-iarc-works.aspx
[719] Marsden, C. (2011) n.60, p187 and PEGI (2018) *PEGI Age Ratings*, https://pegi.info/page/pegi-age-ratings

Option 4: Formal co-regulation

Note that this body would censor citizens directly, so the right to appeal to an independent adjudicator must be built in. The regulator could be associated with and certified/approved by state regulatory bodies, such as the EU Fundamental Rights Agency or European Data Protection Board.

Co-regulation offers the statutory underpinning and legitimacy of parliamentary approval for regulatory systems, together with general principles of good regulation, such as independence from regulatees, appeal processes, audit and governance principles. It also devolves the responsibility for these practices to an independent body, which theoretically gives agility and flexibility to the regulator within these general principles. As the Regulation establishing the .EU domain explains:

> 'Internet management has generally been based on the principles of non-interference, self-management and self-regulation…implementation of the .eu TLD may take into consideration best practices in this regard and could be supported by voluntary guidelines or codes of conduct where appropriate'[720].

Co-regulation is therefore a good example of the pyramid of regulation, with a statutory tip of regulatory principles and authorisation for the regulator, a co-regulator layer that sets out regulatory design, and industry-shaped rules and codes to provide the detailed implementation.

Applied to AI and disinformation, this schematic would suggest a statute laying out the general principles applying to an AI regulator, while the regulator would set out details of the scheme design. Such principles may include, for instance, the principle that no account can be suspended without human intervention to correct for false positive identification of a bot account or egregious content, and the potential for account holder appeal against such a deletion. This would be a minimum requirement to maintain freedom of expression for social media users, to ensure accounts are not deleted without due process. A civil society stakeholder argued that: "Any measure to tackle the complex topic of online disinformation must not be blindly reliant on automated means, AI or similar emerging technologies without ensuring that the design, development and deployment of such technologies are individual-centric and respect human rights"'[721].

This human-regulated AI is more likely to be guaranteed with robust co-regulation than self-regulatory schemes. The parallels with domain names are instructive, as accounts cannot be removed from owners without a formal process (even if the owner is deceased). The cost-benefit of such co-regulation is held in general to allow for more efficient and flexible regulation. That theoretically can provide both lower costs and more responsive regulation, though in practical terms exceptions may exist. Feasibility and effectiveness depend on the initial statutory design as well as the implementation of that design by the co-regulator. There are many examples of successful internet co-regulation, though disinformation is a particularly rapidly moving target. Experience with another open internet issue, network neutrality, shows that such feasibility challenges can be overcome with appropriate multistakeholder engagement[722].

Sustainability of co-regulation is an issue. While it is more robust than less interventionist regulatory designs, internet co-regulation is often chosen due to the particularly profound constitutional and human rights challenges, so that self-regulation is judged inadequate. Thus, a

[720] Regulation (EC) No 733/2002 on the Implementation of the .eu Top Level Domain, at Recital 9.
[721] EDRi (19 October 2018) *Civil Society Calls for Evidence-Based Solutions to Disinformation*, https://edri.org/civil-society-calls-for-evidence-based-solutions-to-disinformation/, quoting Statement of Hidvégi, Fanny, European Policy Manager with Access Now.
[722] See Marsden C. *Network neutrality: From Policy to Law to Regulation* (Manchester University Press, 2017).

frequent failing of co-regulation is that it is eventually supplanted by state regulatory bodies, as for instance with video on demand under the Audiovisual Media Services Directive[723]. Though the direction of travel from self-regulation to state regulation is not inevitable, it can be made due to pressure from both government and from regulates seeking regulatory certainty. In such situations, the costs of co-regulation can escalate as the scheme attempts to shadow state regulation. Risks and future uncertainties are thus closely tied to the regulatory commitment to making co-regulation an end state rather than an interim measure. As explained for Option 3, coherence with EU objectives are easier to assess with co-regulation than with self-regulation.

Potential ethical, social and regulatory impacts revolve around the media pluralism dilemma, that increasing pluralism and diversity with regulation risks regulatory capture and the danger that the regulated diversity does not satisfy the users' needs in a free society. The fundamental rights issues with co-regulation are similar to those for less direct regulatory interventions – freedom of expression as a fundamental right may be held inappropriate for anything but state regulation, a constant issue in internet regulation.

Option 5: Statutory regulation

In Option 5, a regulator would be tasked to combat disinformation directly by licensing of content providers and their systems for content moderation. Current electoral and broadcast regulators already perform this function for offline media. The UK Parliament states that '[i]n this rapidly changing digital world, our existing legal framework is no longer fit for purpose'[724] and has suggested this option. Hearings are ongoing on the role of the UK Information Commissioner and communications regulator Ofcom in such a scheme.[725] Each national context will differ, but in general a regulator would encompass: reformed, strengthened powers for the: electoral commission, data protection authority, advertising regulator, and communications regulator (broadcast, newspaper); police enforcement of criminal law regarding fraud (bot accounts) and other malicious (illegal) communications. It is unclear what such a regulator could achieve without invoking direct censorship of non-conforming organisations[726].

AI systems may be forced to conform to a mandatory national or regional standard, which could lead to dominant standards being enforced anti-competitively. While this was overcome in, for instance, the 3G standard for mobile telephony, there is no convincing example of content moderation subject to technical standards being successfully mandated. The UK government's example of mandatory age rating that it is introducing in 2018 is not a promising approach.[727]

A merger of many regulators is not necessary to combine the functions via coordinated federated networks of those regulators. The UK Information Commissioner report makes this clear as the most effective and sustainable method in the short- to medium-term: “The Government should conduct a review of the regulatory gaps in relation to the content, provenance and jurisdictional scope of political advertising online”. Best practice from the various Member States should be collated, analysed and disseminated, ideally by the European Parliament with assistance from

[723] See European Commission, Proposal for a Council Directive 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 in view of changing market realities, COM(2016) 287 final

[724]UK House of Commons (2018) *Interim Report on Disinformation and 'Fake News'*, Select Committee on Media, Culture and Sport, https://publications.parliament.uk/pa/cm201719/cmselect/cmcumeds/363/36302.htm

[725]UK Information Commissioner's Office (2018) *Democracy Disrupted? Personal Influence and Political Influence*, https://ico.org.uk/media/action-weve-taken/2259369/democracy-disrupted-110718.pdf, Recommendation 10 at p. 46

[726]Marsden, C. (2018) “Prosumer Law and Network Platform Regulation: The Long View Towards Creating Offdata”, 2 Georgetown Tech. L.R. 2, pp.376-398 at 387.

[727]UK Department for Digital, Culture, Media and Sport (2018) *Explanatory Memorandum To The Online Pornography (Commercial Basis) Regulations 2018,* http://www.legislation.gov.uk/ukdsi/2018/9780111173183/pdfs/ukdsiem_9780111173183_en.pdf For criticism, see Hill, R. ('UK.gov To Press Ahead with Online Smut Checks (but expects £10m in Legals in Year 1)', *The Register* (17 October 2018) https://www.theregister.co.uk/2018/10/17/age_verification_legislation_bbfc/

the EU Fundamental Rights Agency.[728] The Digital Rights Clearinghouse set up by the EU Data Protection Supervisor with data protection, consumer protection and competition authorities is another example[729].

Given the speed and flexibility of response demanded by the political priority to combat disinformation, it may be that the reform of existing legislation is a more effective and sustainable form of regulation. For instance, electoral advertising rules can be brought within the ambit of the existing regulator without necessarily reforming primary legislation. The removal of bot accounts is ongoing, and appeal processes could be built into the removal of disinformation, ideally within Option 3. A raft of incremental improvements will be more compatible with the mission to control disinformation and the uses of AI therein, than a more disruptive change at this stage.

4.3 Focus on freedom of expression and media pluralism

The impacts of policies in this area are universally high, and Option 1 remains the least favourable option throughout. The costs of uncertainty are much higher for the less regulatory options, and regulatory sustainability and protection of fundamental rights (including freedom of expression/media pluralism) is more strongly supported for the more regulatory Options 4/5.

Noting that the objective of free and fair parliamentary elections are the highest political priority, regulatory Option 5 would specifically ensure electoral online advertising is regulated online, as it currently is offline. However, that is not a proposal for any kind of super-regulator. Overall, we argue legislation may be premature and potentially hazardous for freedom of expression: co-regulation between different stakeholder groups with public scrutiny is preferable, where effectiveness can be independently demonstrated via audit. Furthermore, noting that Option Zero means a lack of protection of fundamental rights, including appeal against account suspension, as well as exposure to unregulated disinformation, we argue that options to ensure independent appeal and audit of platforms' regulation of their users be introduced as soon as feasible. When technical intermediaries need to moderate content and accounts, detailed and transparent policies, notice and appeal procedures, and regular reports are crucial. It is believed this is also valid for automated removals.

We advise against regulatory action that would encourage increased use of AI for content moderation purposes, without strong human review and appeal processes. There is scope for standardising (the basics of) notice and appeal procedures and reporting, and creating a self-regulatory multi-stakeholder body, such as the UN Special Rapporteur's suggested social media council.[730] As recommended by the Special Rapporteur, this multi-stakeholder body could, on the one hand, have competence to deal with industry-wide appeals and, on the other hand, work towards a better understanding and minimisation of the effects of AI on freedom of expression and media pluralism.

This article emphasises that disinformation is best tackled through media pluralism and literacy initiatives, as these allow diversity of expression and choice. Source transparency indicators are preferable over (de)prioritisation of disinformation, and users need to be given the opportunity to understand how their search results or social media feeds are built, and edit their search results/feeds where desirable. Finally, noting the lack of independent evidence or even detailed

[728]For FRA activities in this area, see European Union Agency for Fundamental Rights *Enabling Human Rights and Democratic Space in Europe*, (2018) http://fra.europa.eu/en/event/2018/enabling-human-rights-and-democratic-space-europe

[729] See EDPS (2019) *Big Data & Digital Clearinghouse*, at https://edps.europa.eu/data-protection/our-work/subjects/big-data-digital-clearinghouse_en On why regulatory intervention is needed to ensure legitimacy in this area, see European Data Protection Supervisor Opinion on online manipulation and personal data 3/2018 p.20; Article 29 Working Party Opinion: Guidelines on Consent under Regulation 2016/679 p.19.

[730] UN Special Rapporteur (2018) n.13, paragraphs 58, 59, 63, 72

research in this policy area, the risk of harm remains far too high for any degree of regulatory certainty. We reiterate that far greater transparency must be introduced into the variety of AI and disinformation reduction techniques used by online platforms and content providers[731].

4.4 Summarizing the Cost-Benefit of Disinformation Regulation

Fighting disinformation does have a cost. Unless European citizens are engaged to work independently on behalf of platform companies - this will be unpopular because this is expensive - policy cannot solve this problem in Europe. What European institutions, whether as a bloc or among its constituent national governments, need to do is to make sure that what companies do is engage European fact-checkers to work with their AI programmes to properly resource their own attempts to stop 'fake news'. They also need European lawyers to work on appeals. Executives in California, ex-politicians such as Nick Clegg, or thousands of badly-paid contractors hired off the internet, from the Philippines or India, cannot regulate European fake news: it has to be Europeans. They must have training in journalism and European human rights law to make judgements on journalistic opinion and freedom of expression. That such a proposal appears highly optimistic is a sign of little regard platforms have thus far been required to show for European human rights standards.

While it would appear to be in the platform owners' best interests to reduce the dissemination of disinformation, the means of doing so could prove to be a sticking point. As ever, it comes down to a question of money. The platforms claim results from AI, not least because it is much cheaper than employing enough humans to solve the problem. The accurate way to deal with fake news is to have a hybrid model of trained humans working on problems that AI has identified. Humans have to make the value judgements. That is expensive for Facebook and YouTube, but absolutely essential to accuracy. They will only make those investments in qualified European values, fact-checkers and 'fake news' spotters if co-regulation is introduced, if they are forced to do so by governments.

Does the evidence support any further legal intervention to control disinformation? First, note that the evidence base is growing rapidly in 2019, and there is strong recent evidence that electoral outcomes have been affected by online disinformation. Second, the UK's Information Commissioner is engaged in auditing the activities of the Brexit campaigners in the 2016 UK referendum, having issued £120,000 fines on 1 February 2019 for three separate illegal uses of personal data.[732] Third, there remain significant questions about Online Behavioural Advertising (OBA), in electoral periods, as a campaign tool more widely, and as an effective and appropriate use of personal information more broadly under the General Data Protection Regulation (GDPR)[733]. We have just begun the investigation of regulating online disinformation and its uses in our democracies.

5. Conclusion: Whom to Regulate, Why and How?

We conceptualised the value dimensions as the protection of representative and electoral democracy. We acknowledged that the size of the economic actors involved means that economic

[731] See further Marsden, C. and R. Nicholls "Interoperability: A solution to regulating AI and social media platforms" Computers and Law (2019), at https://www.scl.org/articles/10662-interoperability-a-solution-to-regulating-ai-and-social-media-platforms

[732] ICO *ICO to Audit Data Protection Practices at Leave.EU and Eldon Insurance after Fining Both Companies for Unlawful Marketing Messages*, (1 Feb 2019) https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2019/02/ico-to-audit-data-protection-practices-at-leaveeu-and-eldon-insurance-after-fining-both-companies-for-unlawful-marketing-messages

[733] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) OJ L 119, 4.5.2016, p.1–88 ELI: http://data.europa.eu/eli/reg/2016/679/2016-05-04

value creation is affected by their regulation, though issues concerning democratic and social values are paramount. Public choice theory theorizes that politicians will mundanely pursue their self-interested course, and the conduct of elections is their primary concern. Given that distinction between electoral regulation and all other forms of public policy, it is unsurprising that electoral reform is central to political concerns. We caution that elections are conducted in a multimedia environment that varies by nation, and that is converging on digital media, but that existing forms of media still predominate. Thus Internet browsing on smartphones even on social media platforms involves largely consumption of content created by existing media organizations that predate the Internet, whether that be television or radio news clips or online versions of newspaper articles.

When the political settlement of these older media was made in the period prior to the 1990s, the political concern with forms of representative democracy resulted in a regulatory settlement that placed political concerns alongside economic concerns. For instance, legislation introduced bans on political advertising on the dominant European forms of social media, broadcasting. The exception to these bans was the United States, where political ownership of local television and radio stations led to a very different policy outcome. The emergence of US-dominated social media platforms has led to a largely unquestioned adherence to the United States model of permitting political advertising as a form of free expression. This has only been effectively challenged in Canada (2019) and Ireland (2018), where the requirements for transparency and a ban on overseas donations to political campaigns led social media platforms to ban all political advertising on these media.

A second set of important questions concerns what kind of institutions and regulatory tools can identify, protect and uphold the policy values in electoral regulation Processes and mechanisms to restore democratic values and social justice and infuse them into digital platforms included transparency, media literacy, and the introduction of forms of human-centred co-regulation. The rents captured in the Internet advertising economic value chain were acknowledged, and regulation for the elimination of those rents has been proposed, by for instance requiring social media companies to redistribute revenue to media organisations, and to donate substantially to fact-checking and other forms of disinformation awareness campaigning. To “follow the value” in this case was the clearly preferred option, in this case to focus on the social media platforms themselves. To regulate access to that value in a way that aligns the incentives of economic operators with those of society was the explicit goal. To refrain from interfering with the economic value process was far less considered given the primary importance to politicians of preventing interference with democratic processes. The alternative for welfare improvement across the value chain would be the wholesale importation of United States ‘richest takes all’ political campaigning, a policy most vociferously opposed by one of the ‘fathers of Internet regulation’, Lawrence Lessig[734]. To think more clearly about what constitutes value creation and value extraction in this policy arena, in order not to recommend regulation that creates equal or greater economic rents, is to argue for the reform, abolition or illegality of ‘recommender systems’ (targeted advertising using personal data online)[735], or even for the abandonment of the capitalist model of digital information creation[736]. While that ambition lies outside the scope of this article, it is notable that several prominent experts now suggest that is the direction in which future policy should be oriented.

[734] Lessig, L. Republic, *Lost: How Money Corrupts Congress—and a Plan to Stop It* (Twelve Publishing, 2011) ISBN 978-0-446-57643-7. This is the first of several book-length arguments about financial corruption of political processes on which Lessig has focussed since his research focus shifted from Internet policy.
[735] Briant, Emma L ‘LeaveEU: Dark Money, Dark Ads and Data Crimes’ (2019) in Paul Baines; Nancy Snow & Nicholas O’Shaughnessy (Eds) *Sage Handbook of Propaganda*, Sage: London; Cobbe, Jennifer and Singh, Jatinder, *Regulating Recommending: Motivations, Considerations, and Principles* (April 15, 2019). Available at http://dx.doi.org/10.2139/ssrn.3371830
[736] Moglen, Eben, *The dotCommunist Manifesto* (January 2003) at http://moglen.law.columbia.edu/publications/dcm.html

Finally, to what entities do we apply rules based on specific values? Who are the recipients of the regulation aimed at fostering the value(s) we chose and protecting the value we create? The answer is once again individuals, but also the social media platforms, and the electoral system itself. It is the uses and abuses of existing rules for elections and social media which have combined to produce a toxic disinformation environment online. In this respect, we note that social media regulation is an ongoing process that has built on earlier instances of Internet regulation, and call for more study of the history of Internet law. Phenomena such as Distributed Ledger Technology (using so-called 'blockchain'[737]), AI and disinformation can be regulated using many of the co-regulatory lessons learnt from Internet regulatory history, and such history should be researched, broadcast and applied[738].

We also urge historical context: disinformation is as old as the written word, as explained in Section 2. It cannot be "solved", but its worst effects can be somewhat ameliorated using those policy options outlined in section 4 and summarized in Section 5. As with so many technological regulatory problems, from railways to nuclear power to the Internet to AI, the lessons of regulatory history are important to adapting existing, and deploying new, regulation for new technology[739]. The complex socio-economic deployment of innovations is what creates regulatory issues, not the technology itself[740]. Elections have a long history, and fake news has played a role in outcomes. The regulation we apply to social media disinformation is a further layer to place over the existing layers of media and election regulation, a further bandage over a gaping wound in imperfect democratic processes. Recognizing the added complexity of local content on digital media, and more internationally sourced disinformation, adds a new and disparate element to the regulation of representative democracy.

[737] Guadamuz, A. and Marsden, C. 'Blockchains and Bitcoin: regulatory responses to cryptocurrencies' First Monday, 20(12) (2015) ISSN 1396-0466
[738] See K. Werbach ed. *After the Digital Tornado: Networks, Algorithms, Humanity*, (Cambridge University Press 2020, in press).
[739] Marsden, C. (2018) *"*Prosumer Law and Network Platform Regulation: The Long View Towards Creating Offdata*"* 2 Georgetown Tech. L.R. 2, pp.376-398 at p380.
[740] Guadamuz and Marsden (2015) supra n.139; Marsden C. (2017) *Net neutrality*, n.125 at Chapter 8.

The Progressive Policy Shift in the Debate on the International Tax Challenges of the Digital Economy: A "Pretext" for Overhaul of the International Tax Regime?

** Dr. Alessandro Turina, Senior Research Associate, International Bureau of Fiscal Documentation (Amsterdam).* *This contribution is current through September 2019. The Author acknowledges the Editors and the anonymous reviewers for their constructive comments. Any errors are the author's.*

Abstract

This contribution presents a critical overview of the policy and legal debate (primarily from a tax treaty law perspective) surrounding the challenges raised by the digitalisation of the economy for the international tax regime. The article addresses some key policy challenges inherent to the proposals for reform under current consideration, focusing in particular on the difficulties associated with fitting the concept of "value creation" within the pre-existing framework based on "source" and "residence", the perceivable drift from a primarily "supply" approach to a "supply and demand" approach in the understanding of "source" as well as, more broadly, a perceivable shift in the policy debate on the tax implications of the digitalised economy, from a targeted analysis aimed at incrementally reforming the existing regime in order to address some specific frictions to a full-blown reconsideration of certain fundamental concepts on which said regime has rested in the last century and, even more notably, to an attempt to renegotiate the current distributive rules of cross-border income, which perhaps herald a partial revenue shift towards "market" countries.

1 Introductory Remarks

1.1 Outline of this contribution

This contribution is meant to provide a critical overview of the current state of play in the area of the possible reform of the international tax regime with the view of addressing the challenges raised by the digitalization of the economy.

From a tax policy angle, development of the digital economy has resulted in non-resident companies operating in market jurisdictions in fundamentally different manners today than at the time international tax rules were designed[741]. Whether they concern business to business (B2B) or business to consumer (B2C) transactions, new business models have emerged (for example online retail, social media, subscription, collaborative platforms models, etc.)[742]. These models challenge in particular the assumption that a non-resident enterprise may only significantly operate in a market jurisdiction through a physical presence[743], typically a fixed place of business

[741] OECD, *Addressing the Tax Challenges of the Digital Economy, Action 1 - 2015 Final Report*, OECD/G20 Base Erosion and Profit Shifting Project, OECD Publishing, Paris, 2015, p. 98, N. 246 (hereafter: OECD, Action 1 - 2015 Final Report).
[742] European Commission, *A Fair and Efficient Tax System in the European Union for the Digital Single Market*, COM(2017) 547 final (21 Sept. 2017), available at: https://ec.europa.eu/taxation_customs/sites/taxation/files/1_en_act_part1_v10_en.pdf., p. 4 (hereafter European Commission, COM(2017)).
[743] OECD, Action 1 - 2015 Final Report, p. 98 N 246.

in the form of a permanent establishment[744] (PE) or through a dependent agent, so called "agency PE"[745].

The French Google case[746] is in this respect particularly emblematic. The case dealt with a structure where Google Ireland Limited did not have any place of business in France. On the other hand, another group company, Google France, provided administrative and marketing support to Google Ireland for which it charged a service fee. In fact, Google France did not accept orders for advertising for display in France from French customers, which had to be approved by Google Ireland[747]. From an international tax law perspective, the question underlying such a fact pattern could have been reformulated as follows: when a non-resident company sells advertising services to local customers and another group company in the market jurisdiction is only in charge of marketing and administrative functions without formally concluding contracts with customers, can the latter be considered a permanent establishment (most notably, an agency permanent establishment) of the former? The Administrative Tribunal of Paris provided an answer to this question in July 2017[748]: the answer was, in substance, negative. In particular, the French Court determined that Google France could not in fact be considered to have the authority to conclude contracts[749]. Quite interestingly, opposite conclusions – leading to the detection of an agency PE - were reached, with regard to the interpretation of virtually equivalent provisions and analogous factual circumstances, by Courts of other countries.[750]

This example shows how the current prevailing modes of conducting business in the digital era can easily allow to conduct significant economic activities in a country without thereby establishing a PE and, thus, without being subject to income taxation thereupon and how the traditional tax treaty rules may not provide adequate answers to such challenges. Indeed, BEPS Action 7, discussed in further detail in Section 3.1.1, by lowering the threshold for the detection of an agency PE has raised the issue and tried to address it; however, as it will also be shown in Section 3.1.1, lowering the P.E. threshold may not yield the desired results without a matching revision of the rules governing the attribution of profits to PEs, something that has not been done and that raises several questions. The underlying policy question to all this is, how much income should the market jurisdiction receive: this question can be considered possibly the fundamental one within the whole debate surrounding international taxation and the digitalization of the economy and different answers are possible.

In light of the foregoing, it may not come as a surprise that the digitalization of the economy has raised an unprecedentedly animated debate concerning the very foundations of the rules governing the cross-border taxation of business profits. It should also not come as a surprise that

[744] Art. 5 OECD Model Convention (MC) 2017.
[745] Art. 5(5) OECD MC 2017.
[746] Google Ireland Limited v. Administration générale des finances publiques, Case 1505113/1-1, Tribunal administratif de Paris (12 July 2017) (hereafter : Google France Case).
[747] For a more in depth outline of the factual background of the case, see J. Schwarz, *Permanent Establishment : La lutte continue*, Kluwer International Tax Blog, 24 July 2017 (available at: http://kluwertaxblog.com/2017/07/24/permanent-establishment-la-lutte-continue/)
[748] *See* generally Google France Case.
[749] See, Google France Case, Para. 16, where the Court, characterising the facts, determined that, while Google France could not be considered an independent agent – which would have ruled out the possibility of detecting an agency PE, it could not be said to have the required characteristic (to give rise to an agency PE as per Art. 5(5) in its pre-2017 version) of "acting on behalf of an enterprise" and having and habitually exercising, in a Contracting State « an authority to conclude contracts in the name of the enterprise. ». In fact, in the reconstruction of the Court, Google France was considered to be only in charge of marketing and administrative functions (e.g., keeping contacts with clients, providing after-sale services, see Para. 11 of the Decision) without formally concluding contracts with customers, considering that, as remarked by the Court in Para. 13 of the Decision, this latter prerogative remained exclusively with Google Ireland.
[750] See in particular, the 2014 Spanish Dell Case delivered by the Audiencia Nacional (AN, 8 June 2015, No. 182/2012) in the wake of similar decisions (e.g.,the Roche Vitaminas case rendered by the Spanish Supreme Court, TS, 12 Jan. 2012, No. 1626/2008). For an overview of this alternative stream of case law, which could be labeled as "substance oriented" rather than "form oriented" and which appears to have become the leading case law in this area of international tax law in Spain, see A. Martin Jiménez, *The Spanish Position on the Concept of a Permanent Establishment: Anticipating BEPS, beyond BEPS or Simply a Wrong Interpretation of Article 5 of the OECD Model?*, 7 Bull. Int'l. Taxn. 458 (2016).

this is an extremely dynamic area : while having delivered an in-depth illustrative report[751] on the matter within the framework of its Base Erosion and Profit Shifting Project in 2015, the OECD did not proceed to set forth specific recommendations as it had done in other areas of the BEPS Project. This occurred despite the circumstance that the tax digital agenda was perceived as one of the most topical items on the agenda by a variety of jurisdictions. It was thus not a surprise that, building on the descriptive work laid down in the BEPS Action 1 Report, some countries decided to implement or planned to implement unilateral measures, typically building upon or revisiting the policy options contemplated by BEPS Action 1. Against such a background, in July 2017, the G20 mandated in the OECD to provide an interim report on the tax issues arising from digitalization by Spring 2018, and a final report by 2020.[752] In this context, the OECD invited public comments on these issues in September 2017[753] and presented the received inputs during a conference held in California in November 2017[754].

Similarly, the European Commission took up the matter and, in September 2017, released a Communication devoted to a « A Fair and Efficient Tax System in the European Union for the Digital Single Market »[755]. The Communication outlined some possible avenues for action, basically revisiting the options laid down in BEPS Action 1 and contextualizing them in the framework of the current EU debate (e.g., aiming at incorporating possible measures centered upon the development of the concept of a "digital permanent establishment" (digital PE) within the framework of its pending work on the Common Consolidated Corporate Tax Base). A set of proposals by the European Commission, in the form of a Directive Proposal was released in March 2018. This timeframe roughly coincided with the Interim Report that the OECD Task Force on the Digital Economy was asked to submit in view of the G20 Finance Minister and Central Bank Governors Meetings to take place on March 19 and April 20 2018.

The two Directive Proposals dealt respectively with an "interim solution", centered upon a "digital service tax" (Directive Proposal COM(2018) 148 final) and a "long-term solution" based on "significant digital presence" (Directive Proposal COM(2018) 147 final). As the European Commission indicated, the concept of significant digital presence is intended to establish a nexus in a jurisdiction that leads to the creation of a permanent establishment. Therefore, the proposal was to be considered an addition to the permanent establishment concept. The use of different thresholds (number of users, number of contracts or amount of revenues) would assure that the proposal applies to a broad scope of business models, irrespective of their size. As it can thus be seen, in light of the Proposal, the significant economic presence would not create a new set of rules to address the taxation of digital economy, but only adds a new criterion to the definition of permanent establishment, such as the case of the construction sites.

The two Directive Proposals however encountered some lukewarm response at the Council level and basically entered into a deadlock. It may be argued, in a somewhat ironic heterogony of ends, that the only impact of the Proposals and, most notably, of the one concerned with the "interim solution", has been to further foster the proliferation of unilateral measures in the form of "digital services taxes". As the equalization levy and the Italian digital service tax may attest, the fundamental template for this type of approach pre-dated the directive proposals but, at least on the European Continent, it cannot be denied that the (for the time being aborted) Directive

[751] OECD, Action 1 - 2015 Final Report, p. 11.
[752] OECD (2017), OECD Secretary - General Report to G20 Leaders, http://www.oecd.org/ctp/oecd-secretary-general-tax-report-g20-leaders-july-2017.pdf, p. 14.
[753] OECD, *OECD invites public input on the tax challenges of digitalization*, http://www.oecd.org/tax/beps/oecd-invites-public-input-on-the-tax-challenges-of-digitalisation.htm (accessed 22 Feb. 2018).
[754] OECD, *Tax Challenges of the Digitalisation: Comments Received on the Request for Input – Part I,* 1 (2017) and OECD, *Tax Challenges of the Digitalisation: Comments Received on the Request for Input – Part II*, 279 (2017), paras 7-13 and 20-42. For a summary of these inputs, *see* S. de Jong,W. Neuvel, A. Uceda*, 'Dealing with data in a Digital Economy'*, (2018) 25 International Transfer Pricing Journal 6.
[755] *See* generally European Commission, COM(2017).

provided a crucial inspiration[756]. At the same time, this shift towards (temporary?) unilateralism[757] may perhaps be considered in a more positive light (provided the concerned measures do not actually enter into force) if regarded as a stimulus to achieve multilateral consensus solutions more swiftly than originally planned, as the acceleration in the output by the OECD/Inclusive Framework in the course of 2019 would seem to attest.

The latest development of the ongoing debate, is the release of a policy note by the OECD[758] followed by a public consultation document released on 13 February 2019. In the context of the policy note and of the public consultation, the OECD appears to having left outside of the scope of the debate possible solutions based on "digital service taxes" or source taxation approaches in favour of, either:

- an approach based on the "user participation" paradigm as developed originally by the British Treasury; or
- an approach based on addressing the challenges of the digitalised economy in light of an approach based on transfer pricing and, in particular, of attributing greater weight to the use of marketing intangibles as a criterion for the attribution of taxing prerogatives; or
- an approach based on the concept of "significant economic presence", somewhat broadening the "significant digital presence" concept mentioned above[759].

The above documents were further consolidated in a "Programme of Work to Develop a Consensus Solution to the Tax Challenges Arising from the Digitalisation of the Economy"[760], released in the course of the summer. This document is very much concerned with the so-called "Pillar II" of the digitalization project, which falls outside the scope of this contribution. Nonetheless, the "Programme of Work" anticipates some interesting developments that should be consolidated in a public consultation document to be released in the last quarter in 2019 and, in particular, the objective to reach a "unified approach" based on a common denominator of the three approaches outlined hereabove[761]. The "Programme of Work" signals that certain inputs from business constituencies received within the framework of the latest public consultation have been taken into great consideration and have added new articulations to the terms of the debate[762].

[756] Even though this profile of analysis transcends the scope of this paper, besides the issues of compatibility with tax treaties addressed further in section 4 of this paper, the compatibility of measures akin to digital service taxes with primary EU Law, both in the area of fundamental freedoms as well as in the area of state aid. For further consideration of these profiles, see, inter alia, A. Turina, *Which 'Source Taxation' for the Digital Economy?*, 46 (6/7) *Intertax*, 495, in particular at 508 – 512 (2018); R. Mason, L. Parada, 92(12) *Digital Battlefront in the Tax War*, 92 Tax Notes International 12, 1183 (2018); J. Nogueira, *The compatibility of the EU digital services tax with EU and WTO law: requiem aeternam donate nascenti tributo*, 1(1) Intl. Tax Stud. (2019).

[757] Which has been heavily criticised also on a policy plane, see, inter alia, J. Becker, J. Englisch, *EU Digital Services Tax: A Populist and Flawed Proposal* (Kluwer International Tax Blog 2018), retrievable at the following link: http://kluwertaxblog.com/2018/03/16/eu-digital-services-tax-populist-flawed-proposal/

[758] OECD Inclusive Framework. Addressing the Tax Challenges of the Digitalisation of the Economy – Policy Note. 23 January 2019.

[759] The Discussion Draft also contemplates a "Pillar 2" based on a "Global Anti Base Erosion" (Globe) Proposal. No comments are provided in this contribution on the proposal for the 2nd Pillar included in the public consultation document, since this proposal does not address the analysis of the nexus and profit allocation rules, but the implementation of specific measures to avoid BEPS, regardless of the jurisdiction that taxes digitalized businesses.

[760] See OECD/G20, *Programme of Work to Develop a Consensus Solution to the Tax Challenges Arising from the Digitalization of the Economy* (OECD/G20 Base Erosion and Profit Shifting Project, 2019).

[761] See in particular p. 11 of the *Programme of Work*, *supra* n. 17.

[762] This seems to be in particular the case of the proposal set forth by Johnson and Johnson based on a distributor-based approach and relying on safe harbours. See Johnson & Johnson, Comments on Public Consultation Document: Addressing the Tax Challenges of the Digitalization of the Economy (Mar. 3, 2019), retrievable at https://www.dropbox.com/s/hou6dvuckmahoft/OECD-Comments-Received-Digital-March-2019.zip?dl=0&file_subpath=%2FJohnson%26Johnson.pdf. For the incorporation of these inputs in the current "Programme of Work", see OECD/G20, Programme of Work, supra n. 20 at 15.

In light of the very dynamic context described above[763], this contribution outlines the background to the discussion by placing the digital taxation debate in the broader BEPS framework, highlighting its interconnections with the pre-existing policy discussion on source and residence and displaying how the latest contributions to the debate may shift the debate towards a different angle.

1.2 Base Erosion, Profit Shifting and Source Vs Residence

The OECD/G20 BEPS Initiative aims at "fixing" the international tax regime on the basis of coherence, substance and transparency[764].

The underlying policy objective is to operate, to the largest possible extent, a shift from unilateralism to multilateralism. From a tax treaty perspective, 7 June 2017 witnessed a historic example of this policy with the signing ceremony of the Multilateral Instrument (MLI) to Implement Tax Treaty Related BEPS Measures[765].

The impact of the BEPS initiative in addressing structural issues in the international tax regime cannot be overestimated and this applies in particular to the affirmation of a new core principle which should inspire international tax rules, namely, that income should be taxed where value is created: this is explicitly affirmed in relation to transfer pricing outcomes[766] but appears more broadly as a principle inspiring the whole BEPS Action[767].

In order to contextualise this contribution in the broader debate on platform values, It should primarily be made clear that the understanding of value that is currently being held as the focus of the tax policy debate currently ongoing is primarily concerned with the concept of economic

---

[763] Which is paired by an exponential growth of the literature dealing with these matters. See, ex multis and in additions to the contributions already cited, R. Avi-Yonah, *International Taxation of Electronic Commerce,* 52 Tax Law Rev. 507 (1997*)*; D. Pinto, *E-Commerce and Source-Based Income Taxation*, Vol. 6 (IBFD Doctoral Series, 2003); P. Hongler & P. Pistone, *Blueprints for a New PE Nexus to Tax Business Income in the Era of the Digital Economy,* IBFD White Paper (2015); M. Olbert & C. Spengel, *International Taxation In The Digital Economy : Challenge Accepted?,* 9 (1) World Tax J. , 3 (2017); W. Schön, *Ten Questions about Why and How to Tax the Digitalized Economy*, 72 Bull. Intl. Taxn. 4/5, p. 278 (2018); Y. Brauner, P. Pistone, *Adapting Current International Taxation to New Business Models: Two Proposals for the European Union*, 71(12) Bull. Intl. Taxn. (2017); R. Danon, *Can Tax Treaty Policy Save Us? The Case of the Digital Economy*, in *Tax Treaties after the BEPS Project*: A Tribute to Jacques Sasseville (B.J. Arnold, Ed; Canadian Tax Foundation, 2018); L. Spinosa & V. Chand, *A long-term Solution For Taxing Digitalized Business Models: Should the Permanent Establishment Definition Be Modified to Resolve the Issue or Should The Focus Be on a Shared Taxing Rights Mechanism?*, 46 (6/7) Intertax, 476; I. Grinberg, *User Participation in Value Creation*, 4 BTR, 407 (2018); M. Devereux & J. Vella, *Taxing the Digitalised Economy: Targeted or System-Wide Reform?*, 4 BTR, 387 (2018); I. Grinberg*, International Taxation in the era of Digital Disruption*: *Analyzing the Current Debate,* 20-22 (28 October 2018), retrievable at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3275737 ; P. Oosterhuis & A. Parsons, *Destination Based Income Taxation: Neither Principled Nor Practical?,* 71 Tax Law Rev., 515 (2018*)*; R. Avi-Yonah, *Designing a 21st century Taxing Threshold: Some International implications of South Dakota vs. Wayfair,* Public law and Legal Theory research paper series, Paper 611 (2018*)*, retrievable at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3201418 ; J. Becker & J. Englisch, *Taxing Where Value Is Created: What's 'User Involvement' Got to Do with It?*, 47 Intertax 2, 161 (2019); P. Pistone, J. Nogueira & B. Andrade, *The 2019 OECD Proposals for Addressing the Tax Challenges of the Digitalization of the Economy: an Assessment*, Intl. Tax Stud. 2 (2019); A. Baez & Y. Brauner, *Taxing the Digital Economy post BEPS…Seriously,* (March 1, 2019); W. Schön, *One Answer to Why and How to Tax the Digitalized Economy*, Max Planck Institute for Tax Law and Public Finance, Working paper 2010-10, 3-12 (2019); W. Haslehner, K. Pantazatou, A. Rust, *Tax and the Digital Economy. Challenges and Proposals for Reform* (Kluwer Law International, 2019); P.Pistone, D. Weber (Eds.), *Taxing the Digital Economy* (IBFD Publications, 2019).

[764] *See* generally OECD, Action Plan on Base Erosion and Profit Shifting, OECD Publishing, 2013 (hereafter OECD Action Plan 2013).

[765] For a general overview of the functioning of the MLI, see in particular R. Danon, H. Salomé, *The BEPS multilateral instrument : General overview and focus on treaty abuse*,3 IFF Forum für Steuerrecht (2017), 197.

[766] BEPS Action 8-10 devoted to transfer pricing is significantly titled " Aligning Transfer Pricing Outcomes with Value Creation". (OECD (2015), *Aligning Transfer Pricing Outcomes with Value Creation, Actions 8-10 – 2015 Final Reports*, OECD/G20 Base Erosion and Profit Shifting Project, OECD Publishing, Paris. *http://dx.doi.org710.1787/9789264241244-en* (hereafter: OECD Actions 8-10 – 2015 Final Report)).

[767] OECD Action Plan 2013, p. 10.

value creation and, most notably, with its extraction by States in the form of income taxes or other types of analogous levies.

What does not appear completely clear is how such a novel emphasis on the place of value creation as a key jurisdictional link for exercising taxation on business profits should fit in or interact with the long-debated controversial relation between the residence versus source balance and the BEPS initiative. On the one hand, the OECD Action plan clearly states that *: "(…) While actions to address BEPS will restore both source and residence taxation in a number of cases where cross-border income would otherwise go untaxed or would be taxed at very low rates, these actions are not directly aimed at changing the existing international standards on the allocation of taxing rights on cross-border income"* [768]*.*

Prima facie, therefore, it seems that the BEPS initiative would only be concerned by cases of profit shifting or instances in which profits are taxed nowhere ("stateless income ") or at an unintended very low rate[769]. The idea of reuniting income with value creation or substantial activities, which represents the core principle of several BEPS action items (whether or not consisting in minimum standards), could however also be seen as a way to revisit residence versus source through the back door[770]. European corporate tax policy, the most important example of regional implementation of the BEPS initiative, is also driven by similar considerations. For example, in its recent communication on a fair and efficient tax system in the European Union for the Digital Single Market, the Commission reiterates that: *" Since the start of its mandate, this Commission has taken action to ensure the principle that all businesses operating in the EU should pay their taxes where profits and value are generated. This principle is essential for a fair and effective taxation in the Single Market, and it can only be enforced through common and coordinated measures "*[771].

The fundamental ambiguity that can be appreciated in this area may have some unwarranted effects. In this regard, it has recently been observed that: *"this new principle (of value creation) is simply different from the existing principles inherited from the 1920s. As the basic structure is being kept in place and the new principle is being overlaid on top it, the post-BEPS international tax system is likely to be more incoherent, with taxing rights being aligned with economic substance in some cases but not in others. There does not appear to be any principle for distinguishing between the two set of cases; at best, reliance will be placed on vague and arbitrary tests such as "artificial" and "excessive""*[772]*.*

It thus appears clear that the notion of "value creation" as undertaken within the framework of the debate on the digitalization of the economy is not a traditional concept belonging to international tax law[773]. Nonetheless, as recent inter-disciplinary scholarship has observed, the concept of value creation can have implications for the determination of sufficient taxable nexus[774]: while the existing criteria underpinning international taxation assign value creation exclusively to the "supply side" when it comes to determining a nexus for establishing the presence of taxing rights, a "more economic" understanding of value creation would implicitly require that the role of the demand side also be taken into consideration, as according to an understanding of value creation

[768] OECD Action Plan 2013, p.12.
[769] See R. Danon, *Can Tax Treaty Policy Save Us?,* supra n*. 21*, 188. About the notion of "stateless income" please refer to E.D. Kleinbard, *Stateless Income*, 11 Fla. Tax Rev., 699 (2011).
[770] Ibidem.
[771] European Commission, COM(2017), p. 11.
[772] M. P. Devereux, J. Vella, *Implications of digitalization for international corporate tax reform*, in : WP 17/07, 8.
[773] J. Becker, J. Englisch, Taxing Where Value Is Created: What's User Involvement Got to Do with It?, (2019) 47 Intertax 161.
[774] J. Becker., J. Englisch, *op. cit*., 164.

based on the mainstream "subjective" theory of value[775], supply cannot create value completely independent of demand[776].

So far reference has been made to the expression "market jurisdiction", which is currently oft reoccurring in the policy debate on the tax challenges arising from the digitalization of the economy facilitated by the OECD/Inclusive Framework. The expression is in itself rather intuitive, in that it aims to vividly depict the "demand side" vis-à-vis a "supply side" in cross-border transactions. In policy jargon, this dichotomy is often referred to interchangeably as the relationship between "source" and "residence". As a matter of fact, this may risk to appear as an oversimplification but, rather interestingly, it is illustrative of a tension that characterizes the debate surrounding the cross-border taxation of business profits. Namely, while with regard to other items of cross-border income, such an approximation would not be too inaccurate, in the case of cross-border corporate taxation the situation is much more complex and may be argued that this is where the crux of the matter currently lies. In other words, would the agenda concerned with the taxation of the digital economy possibly constitute a testing ground for reconsidering the terms of the debate, instead, through an "origin versus destination"[777] perspective in light of a reconsideration of the concept of source?[778].

Through legal lenses, the notion of source could, prima facie, appear rather tautological per se, merely constituting a proxy[779] to postulate the existence of a "genuine link"[780] that would allow a certain state to exert its tax jurisdiction (a peculiar and composite form of jurisdiction, but a jurisdiction nonetheless). Also in this case, however, such an understanding would seem to be called into question, as recent streams of scholarship in public international law have identified what could be called a "territoriality bias" in the international legal discourse.[781]

Indeed, through economic lenses, the most updated contributions in the literature would seem to suggest that source (implying reference to a "source of income") is a problematic concept, as it has been observed that income lacks geographical attributes.[782] However, if this the case, all source rules would automatically have to be considered not only artificial, but also arbitrary.

A less ambiguous approach to the topic could lie in distinguishing, within the scope of source, "source as origin" and "source as destination".[783] This appears in line with the main dichotomy developed in this area by economic literature, namely the supply-based approach and the supply-demand based approach.[784]

---

[775] See, ex multis, the definition of value of goods as "arising from from their relationship to our needs and is not inherent in the goods themselves", see C. Menger, *Principles of Economics*, 120, as reported by J. Becker, J. Englisch, *op. cit*., 163.
[776] J. Becker, J. Englisch, *op. cit*., 164.
[777] M. de Wilde, *Tax Jurisdiction in a Digitalizing Economy: Why "Online Profits" Are So Hard to Pin Down?*, 43 *Intertax* 12, at 796, 797 (2015). On the introduction of a destination-based cash flow tax and, more broadly, on the concept of taxation at destination, *see* A. Auerbach, M. Devereux, M. Keen & J. Vella, *Destination-Based Cash Flow Taxation*, Saïd Business School Working Paper 2017/09 (2017).
[778] For further considerations on this point, see Turina, *Which Source Taxation, supra* n. 16, at 495.
[779] In this regard, it has been observed that "source is no 'a priori' concept. See K. Vogel, *Worldwide vs. Source Taxation of Income – A review and Re-Evaluation of Arguments* (Part I), 16 (8/9) Intertax, 216, at 217.
[780] On the international public law underpinnings of the jurisdiction to tax, *see* J. Martha, *The Jurisdiction to Tax in International Law: Theory and Practice of Legislative Fiscal Jurisdiction* (Kluwer 1989). In more recent literature, see further on this, J. Kokott, *The 'Genuine Link' Requirement for Source Taxation in Public International Law*, in Haslehner et. Al (Eds), *Tax and the Digital Economy*, supra n. 21.
[781] P.D. Szigeti, *The Illusion of Territorial Jurisdiction*, 56 Tex. Intl. L.J., at 369 (2017).
[782] For this effective formulation, *see* M. de Wilde, *Sharing the Pie: Taxing Multinationals in a Global Market* (IBFD 2017). The original submission of such a proposition, although in different terms, would seem to be attributable to H.J. Ault & D.P P. Bradford, *Taxing International Income: An Analysis of the US. System and Its Economic Premises*, in *Taxation in the Global Economy*, at 30 (A. Razin & J. Slemrod eds., U. Chicago Press 1990). *See also* R. Avi-Yonah, *International Tax as International Law: An Analysis of the International Tax Regime*, at 38 (Cambridge U. Press 2007). For an economic perspective, *see* M.P. Devereux, *Taxation of Outbound Direct Investment: Economic Principles and Tax Policy Considerations*, 24 Oxford Rev. Econ. Policy 4, at 698 et seq. (2008).
[783] De Wilde, *Tax Jurisdiction*, *supra* n. 25, at 797-800.
[784] R.A. Musgrave & P.B. Musgrave, *supra* n. 23, at 83.

Under the supply-based approach, profits originate from where the factors that produce the profits operate, and the source of the "normal" return of equity capital should therefore be identified as "the location in which the actual operation of the capital occurs".[785] Citing more recent scholarship, the OECD's Technical Advisory Group on Monitoring the Application of Existing Treaty Norms for Taxing Business Profits,[786] in its 2003 Final Report,[787] observed that, pursuant to the supply-based approach, "[t]he mere consumer market does not represent a factor contributing to the added value of the company",[788] provided that economic profits should be related to the situs of the locational rents that generate these profits.

On the other hand, the alternative conceptual platform, i.e. the supply-demand approach, implies that the interaction of supply and demand is what creates business profits, so that it would be necessary to take account of the fact that the demand for the products arises from the consumer market.[789]

Quite remarkably, a large majority of the members of the Technical Advisory Group rejected the supply-demand approach. More specifically, in their opinion:

> [t]he mere fact that the realization of business transactions requires an interaction between the supply of goods or services by an enterprise and the demand in a market state has not historically been considered[790] by countries to provide a sufficient link for considering that the profits of the enterprise arising from these transactions should, for purposes of income taxation, be sourced in the market state.[791]

Fifteen years later, the theoretical conceptualization (supply approach vs. supply-demand approach) should remain the same, but a fundamental option aimed at reconsidering the favour afforded the "supply approach" in favour of a "supply and demand" one would seem to have emerged. It should be noted that this shift may have unforeseen consequences in the sense of treating source and sales as one and the same[792], by contrast, as it has been observed, sales are ultimately a measure of trade while source should more correctly be approached and ascertained as income production function[793].

In light of such a shift, while this is might not always be the case under existing rules, for what concerns the debate surrounding the digital economy, the approximation between the notion of source as traditionally understood and that of destination would seem to be progressing slowly, but perhaps inexorably, even though a distinction may have to be made. Namely, the application of source-based taxation as commonly implemented in its conventional perception, i.e. through a form of withholding tax levied "in the source State"[794] would indeed attribute tax jurisdiction over the concerned income to the country where the goods or services are supplied. This would, in fact, fulfil the supply-demand criterion in the allocation of business profits.

---

[785] R.A. Musgrave & P.B. Musgrave, *supra* n. 23.
[786] The Technical Advisory Group (TAG) on Monitoring the Application of Existing Treaty Norms for Taxing Business Profits was set up by the OECD Committee on Fiscal Affairs in January 1999 with the general mandate to "examine how the current treaty rules for the taxation of business profits apply in the context of electronic commerce and examine proposals for alternative rules".
[787] OECD, TAG, *Are The Current Treaty Rules For Taxing Business Profits Appropriate For E-Commerce ?*, para. 40 (OECD Publishing 2003).
[788] OECD, TAG, *supra* n. 44, *citing* A. Schäfer & C. Spengel, *ICT and International Taxation: Tax Attributes and Scope of Taxation*, Discussion Paper 02-81, Centre for European Economic Research, at 11 (2002).
[789] *Ibidem.*
[790] F. Vanistendael, *Digital Disruption in International Taxation*, Tax Notes Intl., at 177 (8 Jan. 2018).
[791] OECD, TAG, *supra* n. 44, para. 41.
[792] See R. Tavares, *Multinational Firm Theory and International Tax Law, Seeking Coherence*, 8(2) World Tax J., 243, at 275 (2016).
[793] Ibid. See further on analogous earlier reconstructions W. Schön, *International Tax Coordination for a Second Best World* (Part III), 2 (3) World Tax J., 227 (2010), 227.
[794] It is well known in international tax scholarship that this expression is not a technical one. *See* in this regard the analysis carried out in K. Vogel, *"State of Residence" May as Well Be "State of Source": There is No Contradiction*, 59 Bull. Intl. Taxn. 10, at 420 (2005).

At the same time, if an approach based on "taxation at source" in the traditional sense were to be affirmed, according to which tax would typically be levied in the form of a withholding tax to be applied by a payer resident in the source state, the destination-based logic would surely be fully fulfilled only with regard to business-to-business situations. By contrast, if the payer were expected to act as a "tax intermediary" in relation to an aggregate of business-to-consumer transactions, it would be fairly easy to delink the concerned payer from the underlying consumer market by channelling transactions through another country. Under this scenario, taxation would still nominally take place at source (i.e. in the country where the payer is the resident) but, indeed, it could not be considered to ultimately take place at destination.

1.3 A Testing Ground: The International Tax Challenges of the Digital Economy

As highlighted in the BEPS Action 1 Report[795], the 'digital economy' is not a sector of the economy per se but concerns the entire economy. This can be explained by the fact that '*the digital economy is the result of a transformative process brought by information and technology (ICT)…*'[796]. Therefore, this phenomenon should be apprehended as a spectrum, rather than a specific business field, where traditional 'brick-and-mortar' businesses are at one end, and highly digitalized models are at the other. Thus, the key features characterizing the so-called digital economy are actually true for all businesses, should they concern business-to-business (B2B) transactions, business-to-consumer (B2C) transactions or consumer-to-consumer (C2C) transactions, and traditional businesses using digitalized means to generate profits or new business models that operate in the digital space such as online retailers, online advertisers, social media platforms, cloud computing providers, online platform intermediaries, etc. For this reason, ring-fencing the digital economy for the purpose of taxes would not be desirable[797] because it would create a gap between the reality of businesses and a fictitious reality arising from the international tax framework.

Thus, it is necessary to understand the new features of the digital economy and whether it has modified the way enterprises generate profits in order to adapt the current tax system to this new 'era'. To this end, the BEPS Action 1 Report highlighted six features that are key to the digital economy. They are: i) the mobility of intangibles, users and business functions; (ii) reliance on data; (iii) network effects; (iv) the use of multi-sided business models; (v) tendency towards monopoly or oligopoly; and (vi) volatility[798]. While some of these characteristics mainly characterize the digitalization of business models from a macroeconomic perspective (for instance, this phenomenon leads towards monopolistic situations), other directly challenge international tax standards. In particular, mobility and the use of intangibles allow enterprises to generate profits in a market jurisdiction without physical or tangible nexus[799] therein, which questions the concept of PE[800], whether constituted by a fixed place of business or through an agent[801], as highlighted in Section 1. This results from the fact that the manner in which digitalized businesses operate differ from the ways of making business at the time of the design of international tax rules. As highlighted by the 'Task Force on the Digital Economy' (TFDE) in BEPS Action 1, this may '*create opportunities for achieving double non-taxation, for example due to the lack of nexus in the market country under current rules coupled with lack of taxation in the*

[795] OECD, Action 1 - 2015 Final Report, p. 11.
[796] OECD, Action 1 - 2015 Final Report, p. 11.
[797] *See* OECD, Action 1 - 2015 Final Report, p. 11; Committee of Experts on International Cooperation in Tax Matters, "*Tax challenges in the digitalized economy: Selected issues for possible consideration*" E/C.18/2017/CRP.22 (17-20 October 2017), p. 4-5 N 11 (hereafter: UN Report).
[798] *See* OECD, Action 1 - 2015 Final Report, p. 64-65 N 151.
[799] OECD, Action 1 - 2015 Final Report, p. 98 N 246.
[800] Art. 5 OECD MC 2017.
[801] On the subject of the agency PE, *see* the Google case discussed in Section 1.

jurisdiction of the income recipient and of that of the ultimate parent company'[802]. In other words, it can be argued that the digital economy disrupts the traditional residence versus source balance when it comes to the taxation of non-resident enterprises (NRE) in the market jurisdiction. However, it seems that the digital economy raises more fundamental concerns, such as raising '*questions regarding the paradigm used to determine where economic activities are carried out and value is created for tax purposes, which is based on an analysis of the functions performed, assets used and risks assumed*'.[803]

2. The Work on BEPS Action 1

With the 'Base Erosion and Profit Shifting' (BEPS) Project, the OECD and the G20 intends to '*"fixing" the international tax system on the basis of coherence, substance and transparency*'[804] with the policy objective of operating a shift from unilateralism to multilateralism.[805] Against this backdrop, Action 1 '*Addressing the Tax Challenges of the Digital Economy*' aims at highlighting the tax issues raised by the digital economy and proposing solutions to address them. In particular, the options advanced by the OECD to tackle these issues rare: 1) the amendment of the existing '*permanent establishment*' definition[806], 2) the introduction of a '*Significant Economic Presence*' test (SEP);[807] 3) the application of a withholding tax on goods or services;[808] 4) the adoption of an '*equalization levy*';[809] and 5) a VAT solution.[810] It is however worth mentioning that in both the 2014 Interim Report[811] and the Final Report of Action 1, none of these options were recommended[812]. This can be justified by the fact that it is expected that some of the tax challenges raised by the digital economy will be mitigated by other actions of the BEPS package which indirectly tackle these challenges, such as Action 7 'Preventing the artificial Avoidance of Permanent Establishment Status'[813]. However, in the meantime, actual tax rules are unable to cover all of the business transactions arising from businesses operating in the digital space[814], which subsequently brings tax uncertainty to the table[815].

3. Policy Options under consideration since the launch of the BEPS Project

3.1 Treaty-Based Solutions

3.1.1 Introducing a "Digital PE" Concept

[802] OECD, Action 1 - 2015 Final Report, p. 99 N 249.
[803] OECD, Action 1 - 2015 Final Report, p. 99 N 249.
[804] R. Danon, *Can Tax Treaty Policy Save Us?*, supra *n.* 21, p. 188; on the BEPS project, see OECD Action Plan 2013.
[805] Ibidem.
[806] *See* OECD, Action 1 - 2015 Final Report, p. 86-87 N 204-217; *see also* OECD (2015), *Preventing the Artificial Avoidance of Permanent Establishment Status*, *Action 7 – 2015 Final Report*, OECD/G20 Base Erosion and Profit Shifting Project, OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264241220-en (hereafter: OECD, Action 7 - 2015 Final Report).
[807] *See See* OECD, Action 1 - 2015 Final Report, p. 107 N 277. It could result in a new form of nexus or take the form of a SEP PE. For more information on the proposal, *see* OECD, Action 1 - 2015 Final Report, p. 107 N 277-294.
[808] *See* OECD, Action 1 - 2015 Final Report, p. 113 N 292. For more information on the proposal, *See* OECD, Action 1 - 2015 Final Report, p. 113-115 N 292-301.
[809] *See* OECD, Action 1 - 2015 Final Report, p. 115-16, N 302. For more information, *see* OECD, OECD, Action 1 - 2015 Final Report, p. 115-117, N 302-308.
[810] For more information on that proposal, *See* OECD, Action 1 - 2015 Final Report, p. 120-129, N 309-339.
[811] OECD (2014), Addressing the Tax Challenges of the Digital Economy, OECD/G20 Base Erosion and Profit Shifting Project, OECD Publishing. http://dx.doi.org/10.1787/9789264218789-en
[812] OECD, Action 1 - 2015 Final Report, p. 13
[813] OECD (2015), *Preventing the Artificial Avoidance of Permanent Establishment Status, Action 7 – 2015 Final Report*, OECD/G20 Base Erosion and Profit Shifting Project, OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264241220-en (hereafter: OECD, Action 7 – 2015 Final Report).
[814] IMF & OECD (2017), Tax Certainty: IMF/OECD Report for the G20 Leaders. http://www.oecd.org/tax/tax-policy/tax-certainty-report-oecd-imf-report-g20-finance-ministers-march-2017.pdf, p. 22 (hereafter IMF/OECD Report)
[815] IMF/OECD Report, p. 22-23.

The profits of an enterprise are taxable in the State of residence of the enterprise[816], unless the latter carries on its activities in a Contracting State through a Permanent Establishment (PE) as defined by Article 5 of the OECD Model Tax Convention (MTC)[817]. Thus, when defining whether a non-resident enterprise has to pay taxes in a market jurisdiction, the PE concept of utmost importance.

In the 2014 version of the OECD Model, there exists three types of PE, that is a fixed place or physical PE as addressed by paragraph 1, a PE as a consequence of construction-related activities as addressed by paragraph 3 (which will not be discussed in this contribution since it does not directly relate to the digital economy) and an agency PE as addressed by paragraphs 5 and 6. Similarly, there are also exceptions to the PE definition as evidenced in paragraph 4.

The issue *vis-à-vis* the digital economy is that, as highlighted by the United Nations' Committee of Experts on International Cooperation in Tax Matters, '*it is clear that all the variables contributing to the existence of a PE depend on physical presence of either personnel or fixed place of business*'.[818] As illustrated in section 1, one of the key feature of digital businesses is that they challenge the need for a physical or tangible nexus in the market jurisdiction since they can generate profits in these jurisdictions without such a presence therein. Thus, businesses operating in the digital space can easily circumvent the threshold set by paragraphs 1 as well as 5 and 6, either by relying on the exemptions listed in paragraph 4[819] or through the use of commissionnaire or similar arrangements as set in paragraph 5 instead of the use of subsidiaries as evidenced in paragraph 6.[820]

Since the PE concept is the historical option[821] to tax the profits generated by NREs in the market State, amending the PE definition could be argued to be a natural solution. Therefore, Action 7 of the BEPS package tackles this option by lowering the PE threshold, notably through the amendments of paragraphs 4, 5 and 6 of Article 5 of the 2014 OECD MTC. Although these are not minimum standards, they are now incorporated in the 2017 update of the OECD model. In particular, in a pre-BEPS world, a company would have been able to circumvent the constitution of a PE in the market jurisdiction by ensuring that the functions performed in this market jurisdiction were limited to 'preparatory or auxiliary'[822] ones, as per paragraph 4. Following the

---

[816] See Art. 7(1) of the OECD Model (OECD (2015), *Model Tax Convention on Income and on Capital 2014 (Full Version)*, OECD Publishing. http://dx.doi.org/10.1787/9789264239081-en (hereafter 2014 OECD Model)). This holds true for the new version of the OECD Model (OECD (2017), *Model Tax Convention on Income and on Capital: Condensed Version 2017*, OECD Publishing. http://dx.doi.org/10.1787/mtc_cond-2017-en (hereafter: OECD Model 2017)). However, since the latter entails amendments brought by, among others, BEPS Action 7, references to both models will be made depending on the discussed topic.

[817] See 2017 OECD Model, Art. 5.

[818] UN Report, p. 16.

[819] OECD, Action 7 - 2015 Final Report, p. 28 ss. N 10-15; *see also* OECD, *Commentary on Article 5 Concerning the Definition of Permanent Establishment* (available at: http://www.oecd-ilibrary.org/taxation/model-tax-convention-on-income-and-on-capital-condensed-version_20745419), p. 132 ss. N. 58-78, (hereafter: OECD Commentary 2017 on Article 5).

[820] OECD, Action 7 - 2015 Final Report, p. 15-16, N 5-9; *see also* OECD Commentary 2017 on Article 5, p. 141 ss. N 82-101 and p. 146 ss. N 102-114. Although this transcends the scope of this paper, it is worthy to mention that many of the dysfunctionalities surrounding commissionaire-based business models are ultimately rooted in irreconcilable differences between the common and civil approaches to these categories. As it has been observed: "[t]he common law concept of agency and the civil law concept of commissionnaire are not equal and their interpretation in light of the article 5 of the OECD model treaty must be carried out according to the understanding of their different legal nature. As the commissionaire concept is unknown in common law countries, this is treated as a simple undisclosed agent. Consequently, the commissionnaire is deemed to bind the principal, and thus, a commissionnaire arrangement will almost always be deemed to give rise to a PE in common law countries. Nevertheless, the risk for a commissionnaire of having a taxable presence in civil law countries only would arise when domestic law treats the commissionnaire as binding the principal.", see L. Parada, *Agents vs. Commissionnaires: A Comparison in Light of the OECD Model Convention*, 72 *Tax Notes International* 1, 59 (2013). It is thus not surprising that virtually all the case law dealing with the matter of abuse of commissionaire structures has emerged in Continental European countries, with France and Norway (inter alia) typically adhering to a more formalistic view and Spain taking a more substance-oriented path; for a critical analysis of the Spanish case law trends on the matter, see A. Martin Jimenez, *The Spanish Position*, *supra* n. 10, 458.

[821] R. Petruzzi, R. Holzinger, *Profit Attribution to Dependent Agent Permanent Establishments in a Post-BEPS Era*, 9(2) World Tax Journal, 267(2017).

[822] OECD, Action 7 - 2015 Final Report, p. 10.

amendments proposed in Action 7, '*Article 5(4) is modified to ensure that each of the exceptions included therein is restricted to activities that are otherwise of a "preparatory or auxiliary" character*'.[823] Additionally, amendments to this paragraph also ensure that multinational enterprises cannot circumvent the PE status though the fragmentation of activities[824]. With regards to the agency PE concept, in a pre-BEPS era, the use of commissionaire would not have created a PE in the market jurisdiction because of the wording of Article 5(5) of the 2014 OECD MTC. As a matter of fact, although such contracts were concluded *on behalf of* the NRE, they were not concluded *in the name of* the NRE.[825] Moreover, when contracts were finalized or authorized abroad, or when the person usually exercising the authority to conclude contracts was considered to be an independent agent as per paragraph 6, a PE in the market jurisdiction did not arise[826]. BEPS Action 7 thus modified the agency PE notion to ensure that '*as a matter of policy, where the activities that an intermediary exercises in a country are intended to result in the regular conclusion of contracts to be performed by a foreign enterprise, that enterprise should be considered to have a taxable presence in that country unless the intermediary is performing these activities in the course of an independent business*'.[827] In light of the foregoing, it may be argued whether a situation such as one that gave rise to the French Google case would be decided differently in a post-BEPS world. The answer appears positive, as the new rules may support the detection of an agency PE in analogous fact patterns; at the same time, the more thorny issue of how profits should be attributed to such a PE remains unaddressed.

Although the modification of the PE definition seems to be a promising option to address the new manners businesses use to generate profits in the market jurisdiction, this solution brings along further challenges. Firstly, for the amendments to apply, both parties to a covered tax agreement must have ratified the corresponding article in the multilateral instrument. In particular, it is necessary that they have selected the same option within the proposal (for instance, to amend paragraph 4 of the OECD MTC, two options are proposed)[828]. Thus, it can be assumed that these changes will not be uniformly implemented. Additionally, even though a PE would be constituted in the market jurisdiction following these amendments, profit attribution issues arise. As a matter of fact, when it comes to paragraph 4 of Article 5, the taxable profit in the market State would be restricted to the functions performed in that State. In other words, no real increase in the taxable profits by the market jurisdiction would take place. Similarly, with regards to the agency PE, if the intermediary in the market State is remunerated at arm's length, it can be assumed that no further profit should be attributed to the PE.[829]

For this reason, tax policy makers and academics are proposing 'enhanced' versions of these amendments. For instance, BEPS Action 1 itself proposes a new nexus threshold, a *significant economic presence*, which would allow market jurisdictions to be granted taxing rights. However, as outlined by the OECD, this option would allow a meaningful allocation of income to the new nexus if, and only if, actual attribution rules are revisited[830]. This holds true for the new PE nexus proposed by Hongler/Pistone[831] as well, which aims at taxing business income in the era of the digital economy. Thus, it can be argued that if modifying the PE definition is the recommended option, modifications to attribution rules will have to follow.

[823] OECD, Action 7 – 2015 Final Report, p. 10.
[824] 2017 OECD Model, Art. 5, para. 4.1.
[825] OECD, Action 7 – 2015 Final Report, p. 15, N. 5.
[826] OECD, Action 7 – 2015 Final Report, p. 10.
[827] OECD, Action 7 – 2015 Final Report, p. 10.
[828] OECD, Action 7 – 2015 Final Report, p. 28-29 and 38.
[829] University of Lausanne, Tax Policy Center, R. Danon/V. Chand, *Comments on the Discussion Draft* (2017), Example 2, Paras. 28-35.
[830] OECD, Action 1 – 2015 Final Report, p. 112 N 285.
[831] See Hongler, Pistone, *supra* n. 21.

3.2 Source-Based Approaches: Withholding Taxes and Equalization Levies

In addition to a revisiting of the permanent establishment by foreseeing a "significant economic presence", BEPS Action 1 Report mentions that a withholding tax[832] could, in theory, be imposed alternatively[833]as:

- a standalone gross-basis final withholding tax on certain payments made to non-resident providers of goods and services ordered online; or
- as a primary collection mechanism and enforcement tool to support the application of the nexus option based on significant economic presence.

The first configuration of the concerned withholding tax could be applied to transactions for goods or services ordered online (i.e. digital sales transactions) or to all sales operations concluded remotely with non-residents. Under the second configuration, the withholding tax would be non-final and would be used as a tool to support net-basis taxation. In this scenario, a broad scope of application covering all remote supplies could be foreseen, the tax so withheld could be claimed against any outstanding tax liability resulting from the detection of SEP or, should no SEP be detected, be claimed back by the affected taxpayer.

The BEPS Action 1 Report would seem to express a preference for the use of the withholding tax approach as a back-up mechanism to enforce net-basis taxation (i.e., in connection with the implementation of a digital nexus approach)[834]. Such an approach would imply the need to foresee a credit system enabling taxpayers to pay any tax due on net income in addition to the tax withheld, or for taxpayers that are in a loss position on a net basis at the end of the fiscal year to claim a tax refund[835].

As earlier mentioned, the third option foreseen by the BEPS Action 1 Report would be the introduction of an "equalisation levy"[836]. While the withholding tax approach is defined, first and foremost, by its concrete administrative configuration; at least in the context of BEPS Action 1 Report, the equalisation levy would seem to be defined more by its rationale than by its concrete implementation[837]. At the same time, it may be derived that, from a more technical viewpoint, an equalisation levy would constitute a form of excise tax[838]. From an implementation viewpoint, it appears that the levy would be imposed on the gross value of the goods or services provided to in-country customers and users, paid by in-country customers and users, and collected by the foreign enterprise via a simplified registration regime, or collected by a local intermediary[839].

3.3 A Survey of Unilateral Measures Centered Upon Source-Based Approaches

India was the first country to adopt the option of an equalisation levy on digital transactions within the scope of its 2016 Finance Bill. Although possibly too skeptical as a conclusion, it may be

[832] A further implementation model of withholding-based approach that would not be limited only to specified "digital transactions" may be found in Y. Brauner, A. Báez Moreno, *Withholding Taxes in the Service of BEPS Action 1: Address the Tax Challenges of the Digital Economy* (February 2, 2015). WU International Taxation Research Paper Series No. 2015 – 14 (available at SSRN: https://ssrn.com/abstract=2591830).
[833] *See* OECD, Action 1 - 2015 Final Report, p. 113-115 N 292-301.
[834] *See* OECD, Action 1 - 2015 Final Report, p. 115 N 301.
[835] Ibidem.
[836] *See* OECD, Action 1 - 2015 Final Report, p. 115-117 N 302-308.
[837] In this regard, the BEPS Action 1 Report observes that "(equalisation levies) are intended to address a disparity in tax treatment between domestic corporations and foreign corporations" and that [an] equalisation levy could be structured in a variety of ways depending on its ultimate policy objective. In general, an equalisation levy would be intended to serve as a way to tax a non-resident enterprise's significant economic presence in a country." *See* OECD, Action 1 - 2015 Final Report, p. 115-16 N 302.
[838] Ibidem.
[839] See OECD, Action 1 - 2015 Final Report, p. 116 N 304.

speculated that the introduction of an equalisation tax by India may be considered a less subtle attempt than the Israeli one to engage in "treaty dodging"[840] by delinking the taxation of digital transactions from tax treaties introducing a new levy not covered therein.

Even though India features an extensive array of aggressive sourcing rules within its domestic legislation[841], these provisions have been more successful in increasing the number of tax disputes, than in subjecting digital transactions to tax, provided that the Indian Courts have consistently rejected positions taken by Tax Authorities by relying on the favourable treaty provisions[842]. As a result, there was a perceived need, especially in the perspective of the local Tax Authorities, to introduce a new legislation that would grant mutual exclusivity to Indian tax law in subjecting digital transactions to tax. According to the Indian Tax Authorities, such new legislation, as it would be carved out of income taxation, would enable India to keep the levy outside of the scope of application of tax treaties[843]. As a result, the Indian "equalization levy" constitutes a tax which cannot be credited on the tax paid by the for foreign company in its residence country.

With regard to the actual design and implementation of the levy, the Indian equalisation levy would be imposed on the gross value of the goods or services provided to in-country customers and users, paid by in-country customers and users, and collected by the foreign enterprise via a simplified registration regime or by a local intermediary. Such a levy is described as an equalisation levy as the word "equalisation" represents the objective of ensuring tax neutrality between different businesses using differing business models or residing within or outside the taxing jurisdiction[844]. In addition to the equalisation levy, India introduced a surtax of 6% to be levied on payments to foreign companies for online advertising services and to be withheld by the resident payors.

The conceptual forerunner of the Indian "equalisation levy" may possibly be traced to the British "diverted profits tax". In the same vein, Australia adopted "multinational anti-avoidance law" (anti-abuse legislation)[845]. These measures share with the Indian "equalisation levy" the characteristic of being unilateral and outside of the scope of application of tax treaties, thus being susceptible to generate international double taxation issues that may largely be unsolvable due to the likely non-creditability of such taxes in the country of residence of the companies affected by such levies. While it may be too early to be in a position to gather statistics in this respect, the potential concern about international double taxation arising from these measures is in re ipsa. In fact, assuming that affected taxpayers would already be subject to "ordinary" income tax obligations in

[840] In this case, the dodged interpretation may be the one concerning the "taxes covered" provision to be found in tax treaties.

[841] An analytical overview of these provision would fall outside the scope of this paper. As a recent example, the "secondary source rule" for royalties can be cited. Source rules for investment income such as royalties are typically based on the residence of the payor; by contrast, the Indian rule attracts India's taxing jurisdiction, as income thereby sourced, royalties that are related to the conducting of a business in India.

[842] India is perhaps the Country that has given rise to the greatest volume of tax treaty-related or cross-border tax case in the last year, as the largest international tax treaty database, the IBFD tax treaty database (retrievable at www.ibfd.org) attests. To build up on the previous example, see supra note 72, on the "secondary source rule", ex multis, the *Qualcomm* case rendered by the Delhi Income Tax Appelate tribunal can be cited (150 TTJ 661, 26 June 2009) can be cited. The Tribunal considered that the licensed IP was used by the company in manufacturing products outside India and the sale to India "was without any operations being carried out in India which would amount to business with India and not business in India". For a commentary of the case from an Indian domestic perspective, see R. Nayak, A. Jain, 'Ruling on royalty secondary source rule under Indian tax laws', (2013) International Tax Review (retrievable at www.internationaltaxreview.com)

[843] See A. K. Lahiri, G. Ray, D. P. Sengupta, *Equalisation Levy*, Brookings Institution, Working Paper 02 (Jan. 2017); S. Wagh, *The Taxation of Digital Transactions in India: The New Equalisation Levy*, 70(9) Bulletin for International Taxation 547 (2016). For a brief overview of the recently adopted measure see also M. Agrawal, *India at the Forefront in Implementing BEPS-Related Measures: Equalization Levy in Line with Action 1*, 23(4) Intl. Transfer Pricing J., 323 (2016).

[844] Ibidem

[845] It is interesting to remark that. while the British government explicitly mentions digital businesses as the main target group (see for instance this related HRMC press release: https://www.gov.uk/government/news/government-ramps-up-efforts-to-tackle-digital-multinational-tax-risks), Australia has pointed out in more general terms to some large multinationals suspected of diverting profits (see the following press release published on the website of the Australian Government: http://www.budget.gov.au/2015-16/content/glossy/tax/html/tax-03.htm).

their country of residence, the circumstance of being subject to an "equalization levy" or other type of analogous levy in the country of source which would not be creditable in the country of residence, primarily because domestic relief mechanisms typically only cover income taxes, and, even where a tax treaty is in place, if the equalization levy is considered to fall outside of its scope – as it appears to be most likely – then no double taxation relief as per the treaty would be provided. The design of the Indian equalization levy would offer a way out for double taxation in all those cases the foreign service providers acknowledges the presence of a Permanent Establishment in India but of course this is limited to specific situations as in many instances it is likely that the targeted foreign service providers would not integrate the requirements for the detection of a Permanent Establishment on the Indian territory as foreseen by tax treaty provisions based on Art. 5 of the OECD or UN Model.

Italy appears as the first EU Member State to have concretely adopted a short-term solution in the form of a "digital transaction tax". The new tax has been included in the 2018 Finance Bill approved in December 2017[846] and should apply from the 1st of January 2019. The new levy, which is defined as a tax *("imposta")* but which would be subject to the procedural rules foreseen for VAT[847], has been included in the 2018 Finance Bill approved in December 2017[848] and should apply as from the 1st of January 2019. The legislative history behind the tax displays a very strong debate surrounding its introduction, as evidenced by a substantive revision of some of the key features of the tax compared to the original draft[849] . In its final version, the new tax would require Italian resident service recipients (as well as Italian permanent establishments of non-residents) to withhold a three percent tax on consideration paid for services rendered through digital means[850], broadly defined as services rendered through the internet with minimal human intervention and that could otherwise not be supplied without recourse to information technology[851]. Online retail would not be targeted by the tax. The Italian Ministry of Finance should have provided in a decree to be issued by the end of April 2018 the list of the services included in the scope of application of the tax[852], along with further implementation guidance. In its final version, the applicability of the new tax would not be subject to monetary thresholds but a de minimis rule linked to the conclusion of 3,000 relevant digital transactions per year has been foreseen[853]. Notably, the tax would apply equally to non-residents and residents. In addition, in the case of residents, the tax would not be creditable against income taxes. It appears that for this reason, the original rate was reduced from six to three percent[854]. The tax is meant to be collected by means of withholding and paid by the resident service recipients but a compulsory transfer of the bearing of the tax to the service providers is expressly foreseen[855].

For the sake of completeness, it should be mentioned that in November 2017, the United Kingdom set forth the proposal of a tax measure that could have vast repercussions on the digital tax agenda and that could represent the implementation of a withholding tax-based approach along the lines of the developments that have been unfolding at the UN level[856]. In a nutshell, the proposed British measure would introduce a royalty withholding tax to be triggered when a non-resident entity engaged in the supply of products or services in the United Kingdom pays a royalty to a connected party in a low tax jurisdiction. The withholding tax would however be waived in

[846] Art. 1, Para. 1011 – 1019 of Bill No. 2960-B of 23 December 2017 (hereafter: Finance Bill 2018).
[847] Art. 1, Para. 1016, Finance Bill 2018.
[848] Art. 1, Para. 1011, Finance Bill 2018.
[849] For instance, the original rate was reduced from six to three percent and it was determined would apply to resident and non-resident suppliers alike (see Art. 1, Para. 1013), without foreseeing a specific creditability of the new tax against the income tax.
[850] Art. 1, Para. 1013, Finance Bill 2018.
[851] Art. 1, Para. 1012, Finance Bill 2018.
[852] Ibidem
[853] Art. 1, Para. 1013 Bill, Finance Bill 2018.
[854] Art. 1, Para. 1013, Finance Bill 2018.
[855] Art. 1, Para. 1014, Finance Bill 2018.
[856] Where, for instance, the taxation at source of technical service fees has been foreseen under the new Art. 12A. The British initiative would however not specifically single out fees for technical services but, rather, concentrate on the possible base-eroding impact of royalty flows.

case the non-resident entity already has a Permanent Establishment in the United Kingdom or is subject to the Diverted Profits Tax. Another peculiar feature of the royalty withholding tax would be that the UK-resident related parties to the non-resident supplier would be made jointly and severally liable for the tax[857].

It should be mentioned that, addition to the above jurisdictions, in fact, the most notable digital tax development in 2019 has been the introduction or planning of digital services taxes in several Countries, including many that are OECD members: Malaysia, Singapore, South Korea and, in Europe, notably: France[858] and Austria, where the measures have already been enacted as well as other Countries where the measures are pending, such as Spain, Belgium, Czech Republic and the United Kingdom (with scheduled enactment from April 2020). The British experience would appear particularly notable, not only because it would overlay the already introduce diverted profits tax but also because it would constitute the first concrete application of the "user participation" conceptual framework outlined in the following section. Namely, the tax would be meant to be levied, at a 2% rates on revenues derived from British users' creation of value for digital services businesses. All the measures mentioned so far have been expressly labeled as "interim measures" by the concerned jurisdictions, in the sense that all jurisdictions have expressly mentioned their intention to withdraw the measures by means of a sunset clause once an international agreement on the taxation of the digitalized economy has been reached under the aegis of the BEPS Inclusive Framework.

3.4 Transfer Pricing Based Approaches

The latest policy brief released by the OECD in January 2019 introduced a marked emphasis on solutions based on a reform of transfer pricing rules either based on a reconsideration of a role of "user participation" or on the relevance of the deployment of "marketing intangibles".

3.4.1 User Participation

The overall objective and policy rationale of the "user participation" proposal is to identify whether there are significant sources or location-specific rent and, in the affirmative, attribute residual taxing rights to the jurisdiction in which they are established.

This goal gives the "user participation" proposal structurally narrow scope and application, which deviate from the current international tax rules on nexus and allocation only operates in respect of those activities in which the user actively contributes to value creation.

According to the "user participation" proposal, changes would affect social media platforms, search engines and online marketplaces, thus making this proposal the one with the strongest ring-fencing effect.

[857] The proposed withholding tax on royalties formed the subject of a public consultation, providing an in depth outline of the measure, which was launched by the HMRC on December 1st 2017 and closed on 23 February 2017. The text of the consultation is retrievable at the following link: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/663889/Royalties_Withholding_Tax_-_consultation.pdf

[858] The French digital service tax already seems to have ignited the sparks of trade war with the United States, although some compromise would seem to have been struck in the meanwhile. See E. Schulze, *US and France have reached a deal on digital tax, Macron says*, CNBC Online News (26 August 2019), retrievable at: https://www.cnbc.com/2019/08/27/france-and-us-reach-draft-compromise-on-french-digital-tax.html

The rationale of this proposal is consistent with the goal of bringing nexus and allocation of taxing rights in line with value creation, at least insofar as there is an adequate recognition of the contribution by the user.

Various digitalised business models allow algorithms and artificial intelligence to directly interact with users and derive business profits without an on-site human intervention in the market country. Consequently, this interaction facilitates a more targeted functioning of digitalized business, which can maximise profits by processing data provided by the users and approaching them also remotely.

The "user participation" proposal identifies the user contribution to value creation by means of a residual allocation of the non-routine profits at group and company levels, thus generating a structural minimisation of that contribution.

The significant cost of developing algorithms and artificial intelligence, especially if combined with the current low-profits policies of digitalized business, can therefore lead to an ultralow allocation of taxing rights to the country where the users are established, especially if one compares with the actual levels of contributions by the users to the profits derived by the companies.

Furthermore, the "user participation" proposal can give rise to specific problems insofar as digital platforms can give rise to new sources of location-specific rents in a country that may differ from that in which the consumer is based. For this reason, giving greater taxing right to the jurisdiction of "user value creation" need not always mean expanding the taxing rights of the consumer jurisdiction,[859] but to the rents arising from users residing in a given jurisdiction.[860]

In this and other respects, the case for reallocating taxing rights based on location-specific rent must be distinguished from a case based on "destination-based" apportionment, which would by contrast appear to be implicitly factored in by the marketing intangibles proposal, which would effectively lead to a hybridisation of the current system (which would be left in place for routine returns) and of a destination-based inspired approach which would come into play with regard to extra-returns arising from the deployment of marketing intangibles.

The key question underlying to design attempts based on the aim to capture location-specific rents would thus be when location-specific rents arise and how they may be conceptualised to be arising within the framework of the increasing digitalisation of the economy – which, prima facie, would seem to point precisely towards an opposite direction – and, in second instance, how such location specific rents could be measured.

Scholarly contributions that have engaged themselves with this perspective have formulated the conclusion that the main source of location-specific rents in the digitalized economy would be found in "digital platforms". Said digital platforms would be susceptible to the origination of location-specific rents in four main ways[861]:

- through their direct network effects;
- through their indirect network effects;

[859] W. Cui, *The Digital Services Tax: A Conceptual Defense*, Working Paper 26.10.2018, retrievable on the SSRN portal., 9.
[860] *Ibid.*
[861] *Ibid.*. See, *contra*, minimising the relevance of networks effects and "passive use", J. Becker, J. Englisch, *supra* n. 21, 166 and 171, where the Authors "[h]ave rejected the notion that users are co-producers in a tax-relevant way. Most user involvement is actually passive (…); if there is an activity, it is more appropriately classified as consumption which creates externalities (…). While this implies that the user – inadvertedly – creates value for the firm, this does not justify 'compensatory' taxation according to the intrinsic logic of the benefit principle."

- through taxation of online advertisement revenues, which effectively serves as destination-based formulary apportionment; and
- through the collection of data, which should be viewed as creating a tax base.[862]

Besides the case of digital platforms, the impact of the user's contribution proposal would be more limited on the reform of international taxation nexus and allocation.

Furthermore, also the very need to differentiate between routine and non-routine functions, required by the "user participation" proposal could significantly limit the effective recognition of value creation in the country of the users, thus correspondingly reducing the impact on nexus and allocation of taxing powers, which would be residual and limited to the source of unacknowledged location-specific rents.

This can easily be demonstrated by developments already taking place on the ground: in response to the BEPS package (including Action 7), some MNE groups with highly digitalised business models were able to establish local affiliates in market jurisdictions, especially in those jurisdictions constituting the businesses' larger markets. However, the local affiliates are commonly structured to have no ownership interest in intangible assets, not to perform DEMPE functions, and not to assume any risks related to such assets. Accordingly, only a modest return may be allocated to these "limited risk distributors," or LRDs. Thus, without effective changes to profit allocation rules, an MNE group may seek to sidestep the nexus issue by establishing local affiliates that are not entitled to an appropriate share of the group's profit.[863]

The relevance of transfer pricing rules for the perspective at stake would depend on how location specific rent is understood. In particular, once a platform technology is applied to a given country to generate profits, assuming that the deployment of that technology for the users of the concerned country does not exclude the deployment of the same technology elsewhere, the entire economic rent generated by the technology in respect of that country should be attributed thereto. By contrast, current transfer pricing rules emphasise managerial decisions, legal ownership of intellectual property rights, and the bearing of financial risks should entitle a company to residual (and extraordinary) profits outside the user jurisdiction[864].

By contrast, in light of the above outlined location-specific rents perspective, traditional transfer pricing and profit attribution rules will be relevant only for attributing normal returns to various business functions. This two-tier approach (also shared the "market intangibles" proposal) may increase the potential for controversies on the allocation of taxing powers and create more problems than it solves.

The British Government Proposal made the point believes that user participation can be distinguished from the role that customers serve in a business, and how user participation constitutes more than the collection of customer data, which is relevant to a broad range of digital and non-digital businesses.[865]

Therefore, it may be argued that the "user participation" proposal is not about addressing low taxed income or levelling an unlevelled playing field – the justifications given for rule changes in BEPS just a few years ago. Rather, as suggested in literature, the proposals are now clearly about

[862] Interestingly, the UK 2018 Paper explicitly claims that data collection should not be analogized to user participation and in itself does not create new taxing rights for the country from which user data is collected. See UK 2018 Paper, Paragraphs 2.33-2.41.
[863] See OECD, 2019 *Public Consultation Document*, Para. 13.
[864] Cui, *supra* n. 113, 25.
[865] See OECD, 2019 *Public Consultation Document*, 2.3.

a revenue shift to move tax revenue from jurisdictions of residence to the jurisdictions where digital companies have users.[866]

The "user participation" proposal maintains there is something distinctive about value creation in the digital economy. The proposal focuses on the example of a user uploading data on a social media platform to illustrate the importance of user participation in the digital space. However, under current rules, the profits are not necessarily taxed in the country of the user (and viewer of the advert), but rather in the country where the advertising algorithms has been developed, for example. This means that, under the current nexus and allocation rules, the user contribution to the profits is not taken into account when the company is taxed.

However, if user participation is a meaningful concept, it cannot be rationally limited to information communication technologies. Consider a clinical trial from a user participation perspective: such trials involve a corporation giving thousands of individuals free medicine over a period of years in exchange for those users providing deeply personal medical data, as well as a service to the company – the use of their bodies for purposes of experimentation. The resulting data is monetized by obtaining a patent and customizing products to specific diseases and patient populations. This user data is also required for regulatory approvals, without which the company may not sell anything at all.

Neither multisided business models nor network effects are new economic phenomena, nor are those phenomena limited to the digital platform businesses affected by user participation proposals. Multisided business platform are generally defined as businesses that a) offer distinct products or services, b) to different groups of customers, c) whom a "platform" connects, c) in simultaneous transactions. In simpler terms, they are market makers – businesses that help unrelated parties get together to exchange value. Network effects refer to the phenomenon whereby a product or service gains additional value as more people use it.

For this reason, we concur in that it does not seem intellectually defensible to suggest that users only meaningfully contribute to value creation in the context of certain digital platforms, or to think that the boundaries of the idea are clear enough to allow for anything approaching reasonable implementation.[867]

Yet, understanding the user participation perspective remains important. For one thing, the user participation proposal highlights the political angle much of Europe brings to the current digital tax debate. Even more importantly, HMRC and the European Commission have both suggested that when "active user participation" is present, "jurisdictions in which users are located should be entitled to tax a portion of those businesses' profits."[868]

The "user participation" proposal drafted by HMRC wishes to achieve this result using what is in effect a formulary system, but applying it only for the residual profit, thus presupposing a dividing line between routine and non-routine function, which can prove hard to draw in practice.

Either way, just like in the version presented by the EU Commission, the "user participation" proposal seeks to allocate some (although not all) of the excess return of a business to the destination jurisdiction. As earlier evoked, a pure destination-based approach would not necessarily resonate with the location-specific rent perspective to which long-term reform proposals should be inspired.

[866] Grinberg, *International Taxation*, *supra* n. 21, 7.
[867] Grinberg, *International Taxation*, *supra* n. 21,7
[868] HMRC, *Corporate Tax and the Digital Economy: Position Paper Update*, Para. 3.7

3.4.2 The "Marketing Intangibles" Approach

The marketing intangibles proposal is premised on the rationale that an enterprise, traditional or digital, can actively be present in another jurisdiction on a remote basis or through a limited local presence such as a limited risk distributor (hereinafter, "LRD") to develop existing or new marketing intangibles such as brands, trade names, customer data, customer lists and customer relationships[869].

This implies that when the non-resident is actively present through digital or traditional means in another jurisdiction it creates marketing intangibles. Due to the footprint created by marketing intangibles the market State is justified to exercise tax jurisdiction. This proposition seems to be a reasonable suggestion and it could comply with the benefit principle. Such an approach, arguably, also seems to be consistent with the value creation standard[870]. Therefore, it could be justified that the profit allocable to marketing intangibles be taxed to a certain extent in the market jurisdiction. As rightly pointed out by the proposal, only the profit allocable to marketing intangibles should be within the scope of this proposal. The profit allocable to trade intangibles should be carved out[871].

In fact, such a proposal could also mitigate profit-shifting concerns to a certain extent. It may well be true that under the post BEPS rules, profit attributable to marketing intangibles can be stripped out of the market State, i.e. the State where they are created. For instance, consider the situation of Company R in Country R (a low tax State), which operates in the business of branded products, sells its products to its related LRD, Company S a resident of State S, which then sells those products to clients. Under the current rules, the LRD is compensated with a routine margin and the residual profits move out of State S. Thus, this proposal seeks to reallocate a part of the residual profits that are attributable to the marketing intangibles to the market State[872].

The question arises as to how does one determine the residual profit allocable to marketing intangible, which could be reallocated to the market State. One approach would be to rely on a facts and circumstances analysis that is the current transfer pricing approach[873]. Under this approach the contribution of marketing intangibles to the overall profits needs to be determined. Thereafter, a portion of the profit linked to marketing intangibles will need to be reallocated to the market State. Clearly, this approach involves a high degree of subjectivity. This could also lead to tax uncertainty and a plethora of tax disputes. An alternate approach to evaluate the contribution of marketing intangibles would rely on costs (capitalized or not) incurred to develop marketing intangibles[874]. Once again, this approach involves a high degree of complexity.

Thus, in order to avoid these issues, it would be desirable to foresee the application of a simplified residual profit allocation mechanism that would use mechanical approximations[875]. The simplified mechanism could be based on deemed margins[876] and formulaic approaches both at a routine, residual and reallocation level. This would achieve ease of administration in that it would basically entail that the transfer prices would have to be determined based on a safe harbour approach, for instance, the profit margin for an intra-group supply of services would be set at a pre-determined rate instead of being set as a result of a transfer pricing analysis based on a "facts and circumstances" assessment; on the other hand, the distinction between base routine profits and residual profits, conceptually a very complex exercise to which traditional transfer pricing

[869] OECD, 2019 *Public Consultation,* Paras 30-31.
[870] OECD, 2019 *Public Consultation,* Para 33.
[871]OECD, *2019 Public Consultation,* Para 34.
[872]OECD, *2019 Public Consultation,* Paras 35-37.
[873]OECD, *2019 Public Consultation,* Para 45 and 46.
[874]OECD, *2019 Public Consultation,* Para 47.
[875]OECD, *2019 Public Consultation,* Para 47 and 48.
[876]For example, see https://mnetax.com/marketing-intangibles-solution-to-digital-tax-dispute-should-apply-only-to-consumer-facing-businesses-us-official-says-32441

analysis does not appear to have yet found a standardised approach, would be approximated based on some formulas aimed at splitting the residual profits from the routine ones . This aspect is important in light of high degree of complexity of the facts and circumstances approach attached to this proposal.

This proposal also does not create issues related to ring fencing as it applies across the board to all MNEs[877]. Appropriate nexus rule will need to be developed to implement this proposal. Our suggestion would be to design such rules on the basis of turnover only. This would imply that the additional factors reflected in the SEP proposal would not be reflected in this approach. Moreover, such nexus rules need to be designed independently from the permanent establishment framework.

4. Issues with the Proposed Solutions

An approach based on digital economic presence would basically raise some fundamental issues in that it would appear extremely difficult in the current framework of profit attribution rules to permanent establishments based on significant people functions performing significant functions in respect of assets and control of risks. The key peculiarity of the main digital business models lies precisely in the circumstance that it is possible to generate profits in a country without the need to locally deploy the above functions. There would thus be a fundamental incompatibility between the current approaches in profit attribution, which cannot be dismissed so easily, and the peculiarities of "digital Permanent Establishments".

Equalisation levies present several shortcomings both on the legal and policy plane. The main difficulty would seem to be observable with regard to the possibility of fitting these levies within our outside the scope of application of tax treaties.

In the former case, it is quite clear that the adoption of such a measure would constitute an instance of treaty override as a State could not levy a withholding tax on what would arguably constitute an item of business profits without infringing provisions based on Art. 7 of the OECD Model Tax Convention.

On the other hand, if equalisation levies are outside the scope of tax treaties, they may lead - as the Indian experience has already shown - to cases of international double taxation. This may for instance arise where the foreign entity is subject to the levy in the market jurisdiction and to corporate income taxes in its country of residence[878].

From a legal viewpoint , it is highly debatable whether an equalisation levy would really fall outside the scope of tax treaties[879]. The OECD commentaries state that the objective of art. 2 OECD Model Convention (hereinafter, also MC), dealing with taxes covered is: *"to widen as much as possible the field of application of the Convention by including, as far as possible, and in harmony with the domestic laws of the Contracting States, the taxes imposed by their political subdivisions or local authorities, to avoid the necessity of concluding a new convention whenever the Contracting States' domestic laws are modified, and to ensure for each Contracting State notification of significant changes in the taxation laws of the other State"* [880]. The method of levying the taxes is equally immaterial : by direct assessment or by deduction at the source, in the form of surtaxes or surcharges, or as additional taxes[881]. Art. 2(4) OECD MC also states that : *"The Convention shall apply also to any identical or substantially similar taxes that are imposed after*

[877] OECD, *2019 Public Consultation*, Para. 29.
[878] OECD, Action 1 - 2015 Final Report, p. 117, N 307.
[879] *See* further in this regard, R. Ismer, C. Jescheck, 'Debate: Taxes on Digital Services and the Substantive Scope of Application of Tax Treaties: Pushing the Boundaries of Article 2 of the OECD Model?' (2018) 46 Intertax 573.
[880] 2017 OECD Commentary, para. 1 ad art. 2.
[881] 2017 OECD Commentary, para. 2 ad art. 2.

the date of signature of the Convention in addition to, or in place of, the existing taxes. The competent authorities of the Contracting States shall notify each other of any significant changes that have been made in their taxation laws".

The question of whether the Indian equalisation levy would fall within the scope of a provision patterned upon art. 2 OECD MC is discussed in the literature. Some commentators have argued that the Indian equalisation is distinct from a genuine turnover tax. It is certainly the obligation of the payer to deduct the equalisation levy. However, it is the recipient of the payment which is carrying the burden of the tax as he gets a 6% lower consideration for his service. From this perspective, therefore, the equalisation levy is structured differently than a traditional turnover tax (typically VAT) in which it is by contrast the recipient of the services or products which is being burderned because he pays a higher price. In fact, if one is to refer to the case law of the Court of Justice of the European Union on the essential features of a value added tax, he may find that several divergences between an equalisation levy and value added tax could be observed. In fact, four characteristics would have to be met to for a tax to be considered a value added tax[882]: (i) the tax applies generally to transactions relating to goods or services; (ii) it is proportional to the price charged by the taxable person in return for the goods and services which he has supplied, (iii) it is charged at each stage of the production and distribution process, irrespective of the number of transactions which have previously taken place, (iv) the amounts paid during the preceding stages of the process are deducted from the tax payable by a taxable person, with the result that the tax applies, at any given stage, only to the value added at that stage and the final burden of the tax rests ultimately on the consumer.

The view according to which, since the equalisation levy actually seeks to burden the recipient, should be considered a tax levied on a particular element of the recipient's cross-border income making it a special form of source taxation[883], appears to have some merits, especially with regard to taxes such as the Italian one that would foresee a compulsory transfer of the tax burden to the service provider.

On the other hand, should the opposite assessment prevail and the equalisation levy be considered as a form of indirect tax or, more specifically, turnover tax, major issues would arise outside the scope of tax treaty law with regard to its compliance with international trade law obligations. Assuming that, as the current experience suggests, the concerned tax be levied only on the supply of services, the compatibility with the General Agremeent on Trade in Services should be assessed[884].

The GATS include two primary rules: national treatment ("NT") and most-favored-nation ("MFN"). The application of the former concerns discrimination among foreigners, and therefore it applies in cases of different treatment of residents of different countries. GATS Art. XVII prohibits a less favorable treatment of foreign service providers compared to domestic service providers (in the covered industries).

At the same time, GATS includes an exception in Art. XIV(d) for "difference in treatment … aimed at ensuring the equitable or effective imposition or collection of direct taxes," direct taxes defined

[882] See in this regard CJEU, 8 June 199, Case C-338/97, Pelzl and Others and CJEU, 3 October 2006, Case C-475/03, Banca Popolare di Cremona, Para. 28 – 38 where the "test" and the underlying reasoning is applied to a tax such as IRAP. At the same time, a more literal interpretation of the prohibition to introduce turnover taxes has recently been set forth by Advocate General Kokott in her Opinion of 5 September 2013 delivered in relation to the case C-385/12 on the special Hungarian retail tax. For the time being, however, the Court of Justice would appear to have upheld its narrower test.

[883] R. Ismer, C. Jescheck, *The Substantive Scope of Tax Treaties in a Post-BEPS World: Article 2 OECD MC (Taxes Covered) and the Rise of New Taxes*, 45(5) Intertax, 386 (2017).

[884] For a more in depth analysis of the issue, see *Input Statement by the International Observatory on the Taxation of the Digital Economy* (University of Lausanne, International Bureau of Fiscal Documentation, KU Leuven) submitted in relation to the OECD Request for Input on Work Regarding the Tax Challenges of the Digitalised Economy of 22 September 2017 (available at: http://www.oecd.org/tax/beps/tax-challenges-digitalisation-part-2-comments-on-request-for-input-2017.pdf), 279-304.

as “all taxes on total income, on total capital or on elements of income or of capital, including taxes on gains from the alienation of property, taxes on estates, inheritances and gifts, and taxes on the total amounts of wages or salaries paid by enterprises, as well as taxes on capital appreciation.”[885] As it can be appreciated, this definition of “direct taxes” is very much convergent with the standard formulation of Art. 2 of the OECD MC.

Thus, shall an equalization levy be considered to fall within the scope of application of tax treaties, it may be argued that, at least, it would satisfy the above mentioned GATS carve-out.

The struggle to comply with the different layers of international legal obligations would on the other hand make the adoption of an equalisation levy very troublesome from a policy perspective as it would force to twist it in a somewhat unnatural way.

For instance, shall States seriously aim at bypassing that the equalisation levy be caught in the scope of application of tax treaties, they should ensure, inter alia, that the levy be not creditable from income taxes. This would however generate instances of double taxation that may create distortion that could result difficult to address.

Similarly, in order to comply with international trade obligations (as well as EU law obligations, where applicable), care should be taken that the levy should not discriminate against foreign service providers. The “easy” way to achieve this goal, as witnessed in the Italian experience, would be to apply the tax to residents and non-residents alike. However, from a policy perspective, this extension is merely intended to serve as an alibi as the need to impose an additional tax burden on residents would otherwise not have arisen. Rather, the underlying policy objective of the equalisation levy is to provide market jurisdictions with a substitute to current tax treaty jurisdictional rules in order to effectively tax the revenues of non-resident entreprises[886]

A further example in this regard, may be found in whether the equalisation levy should be triggered only above a certain threshold, typically represented by the booking of a certain level of turnover in the market country. In the same way, it may make sense to foreseen the possibility for impacted businesses to be subject to differential tax treaty rates based on the prevailing margins arising as per the respective business models[887]. All these policy adaptations would however, at least from an EU perspective, potentially trigger State aid concerns, as the experience with the Hungarian “advertisement tax” suggests[888]. Thus, in one way or another, it would seem that the existing body of international and supranational rules posing counter-limits to the adoption of unilateral measures appear so pervasive that would be eventually implemented may actually appear as an “equalisation levy” in name only or as a type of levy with fairly concerning distortive effects.

Coming to the latest batch of the proposals emerged in the debate, revolving around what could be defined as “transfer pricing-based” approaches some general critical remarks can be set forth both with regard to the “user participation” proposal and the “marketing intangible” proposals, the former may raise issues of “ring-fencing by design” in the sense that it would seem to be conceived expressly to apply only to a fairly limited set of representative business models expressed by the digitalization of the economy, the “marketing intangible” proposal, by contrast, would have a much broader scope of application and would thus preserve neutrality with the various forms of remotely (digital and non-digital) operated as well as traditional businesses.

[885] See Art. XXVIII (o) GATS.
[886] Danon, *Can Tax Treaty Policy Save Us?*, *supra*. n. 21, Section 2.
[887] G. Kofler, G. Mayr and C. Schlager, *Taxation of the Digital Economy : « Quickfixes » or Long Term Solutions ?*, 57(12) Eur. Taxn., 526-528 (2017).
[888] In particular, on 4 November 2016, the European Commission delivered a negative State Aid decision in relation to the Hungarian Advertisement Tax (case ref.SA.39235). The Hungarian Advertisement Tax could be seen as close to the currently debated equalisation levies. It foresaw a progressive rate structure. This characteristic would found by the Commission to constitute a selective advantage for taxpayers with low turnover. It may be argued that similar issues may be raised in case of the introduction of a de minimis threshold based on turnover as well as by contemplating different rates depending on the different profitability of the concerned digital business models.

Considering the narrow scope and the limited changes in the allocation of taxing powers, in the view of this author, the impact of the "user participation" proposal on the necessary reform of international taxation will be fairly limited[889]. Yet, it seems to recognize that this proposal acknowledges the contribution of users to value creation as it would allow to capture location-specific rents that otherwise escape the current nexus and allocation international rules on income and consumption taxation. On the other hand, the main merit of the "marketing intangibles" proposal would lie in its preservation of neutrality among different industries and forms of conducting businesses. The main criticalities of this proposal would lie in concerns regarding its distributional outcomes vis-à-vis market countries and in the high degree of complexity of the proposal. These criticalities are however not foreign also to the "user participation" proposal and, between the two approaches, the one based on marketing intangibles would appear to be the one with the greatest potential for grounding principled reform and avoid ringfencing.

All sets of proposals examined so far are however associated with a great degree of uncertainty and would basically imply a thorough overhaul of the current pillars of the international tax regime: despite fundamental tax reform may sometimes be the only possibility, the current debate suggests that is currently at stake is not so much a dysfunctionality or an inherent shortcoming of the current rules vis-à-vis digital business models but, rather, the pretext – albeit perhaps not one deprived of its own merits – for a redefinition of the balance between source and residence or, better still, between supply jurisdictions and market jurisdictions, with the latter wanting to increase their share of the international tax base besides what is currently foreseen under applicable international treaty rules. All the proposals discussed so far seem to elude, to a greater or lesser extent, this "elephant in the rule": the user participation proposal does so by providing a radical solution that would however be applicable only to a very limited panoply of cases; the marketing intangibles approach does so based on the declare objective of minimizing the share of profits in relation to which the shift of taxing rights to the market country would occur; finally, the significant economy presence/digital permanent establishment approach is the only one where the distributional element is more clearly articulated, at the same time, the author argues that analogous results may be achieved in a more incremental way, without the need to intervene on the structure of the rules that govern the taxation of cross-border income under international law but, rather, by simply intervening on the distributional effects as the following section of this paper tries to depict.

5. Can a Compromise be Envisaged?

The debate surrounding the taxation of the digital economy and the difficulty to reach an international consensus in this area have so far seemed to entail a shift back to unilateral fiscal policies. The most illustrative example of this trend is the introduction of equalisation levies which, in substance, are designed provide market jurisdictions with a substitute to current tax treaty rules in order to effectively tax the revenues of non-resident digital entreprises.

This phenomenon signals that a real tension surrounding the unavoidable issue of the attribution of taxing rights between "source and residence" is immanent to the above policy debate. Any attempt to circumvent or sugarcoat the issues would lead to suboptimal and possibly distorting results such as the experiences outlined above.

In this regard, it appears that the preferred avenue for addressing such concerns should necessarily pass through the gates of tax treaties. For the reasons outlined in the previous sections, the implementation of an approach based on a "digital permanent establishment" would appear somewhat problematic: as mentioned in section 4, an approach based on digital economic presence would basically raise some fundamental issues as it would appear extremely difficult in the current framework of profit attribution rules to permanent establishments based on significant people functions performing significant functions in respect of assets and control of risks. The key

[889] In this sense, see also Pistone, Nogueira, Andrade, *The 2019 OECD Proposals, supra* n. 21*,* at 16.

peculiarity of the main digital business models lies precisely in the circumstance that it is possible to generate profits in a country without the need to locally deploy the above functions. There would thus be a fundamental incompatibility between the current approaches in profit attribution, which cannot be dismissed so easily, and the peculiarities of "digital PEs". At the same time, these concerns would appear to be transcended by the "second wave" of proposals centered upon transfer pricing approaches would seem to be ready to radically revisit the landscape in the area of profit attribution rules within Multinational Groups as well as in the relations between Head Offices and Permanent Establishments.

A possible alternative , as it could be implemented through tax treaties but would not require an amendment to transfer pricing and profit attribution rules, could be instead to favor the introduction of a new distributive rule dealing with digital supplies[890].

This new rule, which would take precedence over art. 7 OECD Model Convention and which would be structured similarly to art. 10, 11 and, for instance, 12A of the 2017 UN Model Convention[891], would provide that income from digital supplies may be taxed in the state of residence. The rule would however also stipulate that the other contracting state (in the current jargon, "the market jurisdiction") may tax income from digital supplies provided however that the non-resident enterprise derives gross revenues in excess of a certain turnover (ideally, weighted on the actual size of the concerned market) and in such case the tax so charged would not exceed a certain percentage agreed between the contracting states. As for the distributive rules currently in the Model Convention, this new provision would however not apply as regards income effectively connected to a permanent establishment in the market jurisdiction.

Although simpler than most other avenues for reform addressed earlier in this paper, the introduction of such option would however require numerous issues to be settled, such as defining the digital supplies and providing for instance for sourcing rules. Moreover, it is to be acknowledged that whenever new "categories" of income are created for tax treaty purposes, the risks of unforeseen conflicts of qualifications (that is, on which tax treaty rule should be applied to the concerned item of income) increase accordingly.

It is clear that the enactment of this option would require a political consensus that may currently not have consolidated yet but would at least have the merit of foreseeing a solution that would incrementally build upon the existing rules without the need of radically overhauling the transfer pricing criteria at the core of the current international tax regime. The emphasis on transfer pricing reform observable in the "second wave" or proposals may however signal a fundamental discontent with certain aspects of the current transfer pricing rules, so it cannot be excluded that, in the grand scheme of things, the current debate on the international tax challenges of the digitalized economy may turn into, if not a pretext, an occasion for challenging the foundations of the current international tax regime.

[890] See further in this regard, Danon, *Can Tax Treaty Policy Save Us?*, *supra* n. 21., Section 3.2.2.
[891] For an analysis of this new provision see T. Falcão, B. Michel, *Scope and Interpretation of Article 12A: Assessing the Impact of the New Fees for Technical Services Article, 4* BTR, 422 (2018).

E-commerce and effective VAT/GST enforcement: can online platforms play a valuable role?

*Luisa Scarcella**

Abstract

In recent years, the global volume of e-commerce sales has tremendously increased. At the same time, online sales have put the enforcement of traditional VAT/GST rules to the test, thus resulting in a higher risk of tax evasion and fraud. These types of risk are mainly associated with the qualification of taxable persons, the nature of transactions (C2C or B2C) and imports of low-value goods. The OECD has recently highlighted the potential role that platforms can play in the effective enforcement of VAT/GST rules in the e-commerce context. The OECD report includes the possibility to make e-commerce marketplaces liable for the VAT/GST on sales made through their platforms and other measures such as data sharing and enhanced co-operation between tax authorities and online marketplaces. However, even before the release of this report States around the world had already started to consider how to involve platforms in the effective VAT/GST enforcement. A comparative analysis of the legislations adopted in the UK, Germany, Australia and of the EU VAT e-commerce package aims at individuating which are the main benefits and limits of such new provisions. As it emerges, even if there is room for improvement, provisions strengthening the role of platforms for VAT/GST enforcement are in any case a valuable measure for States to adopt in order to create a level playing field for businesses and protect States' revenue.

Keywords: platforms – VAT/GST – distant sale of goods – tax enforcement

1. Introduction

Online platforms, such as e-commerce marketplaces, have massively contributed to the rapid growth of e-commerce. It has been estimated that global B2C e-commerce sales of goods alone are now to be worth in the region of USD 2 trillion annually with projections indicating they may reach USD 4.5 trillion by 2021, USD 1 trillion of which is estimated to be cross-border e-commerce. Moreover, statistics show that approximately 1.6 billion consumers are now buying online, and these numbers are estimated to grow to 2.2 billion consumers by 2022.[892]

Indeed, one of their main benefit of e-commerce is to allow customers and sellers to directly interact with each other. Online platforms facilitating online sales have enabled smaller businesses, to efficiently access millions of consumers in what is now a global marketplace. However, the possibility to connect non-entrepreneurs, smaller businesses and bigger ones with customers all over the world has certainly put to the test traditional VAT rules. Furthermore, the opportunity to exploit the online market and take advantage of the discrepancies among States allowed new tax evasion and avoidance schemes, which consequently resulted in a discrimination towards traditional businesses.

*Research Assistant at the Department of Tax and Fiscal Law of the University of Graz.

[892] European Commission, 'Impact Assessment Accompanying the Document Proposals for a Council Directive, a Council Implementing Regulation and a Council Regulation on Modernising VAT for Cross-border B2C e-Commerce', COM (2016) 379 final; Aaron Orendorff, 'Global Ecommerce Marketplaces: The Complete List by Region and Sales?' (Shopify, 2017) <https://www.shopify.com/enterprise/global-ecommerce-markets> accessed 09 September 2019; James Manyika and others, 'Global Flows in a Digital Age: How Trade, Finance, People, and Data Connect the World Economy' (McKinsey Global Institute, 2014) https://www.mckinsey.com/~/media/McKinsey/Featured%20Insights/Globalization/Global%20flows%20in%20a%20digital%20age/MGI%20Global%20flows%20in%20a%20digial%20age%20Executive%20summary.ashx accessed 09 September 2019.

VAT/GST regulation where not initially created having in mind an online arena. Consequently, all the enforcement provisions were mainly tackling tax evasion and fraud in VAT/GST focusing on features characterizing the physical world. However, the advent of the so-called digital economy, required the application of these rules in the digital world as well. Facing new challenges related to the online dimension of e-commerce, legislators around the world started focusing on how to improve their enforcement mechanisms.

At domestic, European and International level, the debate on how to enforce the applicable VAT rules started to look into the direction of online platforms facilitating the transactions. In order to cope with the challenges brought by e-commerce, State have already implemented some useful measures. For example, in the European Union, there has been a shift in cross border transactions and now VAT is to be applied to the destination country where the customer is based[893], according to the principle of applying VAT where consumption is deemed to take place.[894] Other solutions aim at deeming online platforms for VAT collection. According to a study from the European Commission of 2016, on which the new European legislation is based, 57% of cross-border supplies of goods are purchased via only the three biggest digital platform.[895] Thus, platforms would have an advantageous position since they would be the preferred channel for e-commerce. Moreover, the same report shows that approximately two in every three e-commerce supplies of goods are made via digital platforms with one out of three made through direct sales[896]. However, the need to include platforms in VAT collection could be challenged by a recent study EUROSTAT study published in 2019. Based on the results of the 2018 survey on 'ICT usage and e-commerce in enterprises, it emerge that during 2017, 87 % of EU enterprises with web sales used their own websites or apps, while only 40 % used an e-commerce marketplace.[897] Nonetheless, it could be argued that 40% is still a relevant portion of the market and it does not include entrepreneurs from third countries. Indeed, as showed in section 2 one of the most relevant enforcement issue concerns imported low value goods.

Based on these premises, this contribution focuses on VAT/GST enforcement in online distant sales of goods by looking at the two sides of the same medal. On one side, e-commerce taking place through online marketplaces represents new challenges for VAT/GST enforcement. Thus, it results on different tax burden on the market players. Moreover, ineffective enforcement and consequent VAT/GST evasion and fraud can lead to important revenue losses which cannot then be used for public expenditure to fulfil socio-economic rights under the light of the broader social justice value. On the other side, the presence of online platforms between plenty of small sellers and customers can represent a valuable resource for rendering VAT/GST rules more efficiently enforceable. Even if e-commerce brings up new challenges for correct tax enforcement, since

---

[893] Werner Doralt and Hans Georg Ruppe (edited by Tina Ehrke-Rabel), *Steuerrecht*, (8th edn.) 365; Marie Lamensch, 'European Commission´s New Package of Proposals on E-Commerce: A critical Assessment', (2017) 28 International VAT Monitor 2, 137-146; Aleksandra Bl, 'EU VAT Proposals to Stimulate Electronic Commerce and Digital Publishing', (2017) 28 International VAT Monitor 2, 132 -136; Madeleine Merkx and John Gruson, 'Definitive VAT Regime: Ready for the Next Step?' (2019) 28 EC Tax Review 3, 136–149; Laura Mattes, 'VAT Aspects of Cross-Border Transactions in the BEPS Era', (2016) 27 International VAT Monitor 3, 175 – 181; Elvire Tardivon-Lorizon and Amanda Z. Quenette, 'Indirect Taxation of E-commerce – Significant Recent Changes in the United States and the European Union', (2018) 29 International VAT Monitor 6, 215 -219; Han Kogels, 'Making VAT Fit for the Digital World', (2015) 26 International VAT Monitor 6, 365-366.

[894] When transactions take place in more than one jurisdiction (e.g. goods are moved from one country to another or services are rendered to a person established in another county), the place of supply rules are decisive to determine whether and which State will charge VAT. The "place of supply" or "place of taxation" rules are based on proxies which indicate where the goods or services supplied is expected to be used by a business or consumed. Proxies can be distinguished in tangible and in intangible proxies. The first ones are based on factors such as the location of goods or land, or the place of performance. Differently, the second ones are based on the location of the customers or consumers. Michael Lang, Peter Melz, Eleonor Kristoffersson and Thomas Ecker (Eds.), *Value Added Tax and Direct Taxation: Similarities and Differences* (IBFD, 2009); Sebastian Pfeiffer and Marlies Ursprung-Steindl (Eds.), *Global Trends in VAT/GST and Direct Taxation* (Linde, 2015) 200; Arthur Cockfield and others, 2013, *Taxing Global Digital Commerce*, (Kluwer Law International BV 2013); OECD 'Action Plan 1 Final Report', (2015) 44-45.

[895] European Commission, Op. cit.

[896] OECD, 'The Role of Digital Platforms in the Collection of VAT/GST on Online Sales', (2019), 6.

[897] Eurostat, 'E-commerce statistics', (2019) <https://ec.europa.eu/eurostat/statistics-explained/index.php/E-commerce_statistics> accessed 09 September 2019.

there is the presence of a third party which can be involved in the enforcement mechanism, this business model might even facilitate the enforcement of VAT/GST rules. Consequently, it can result in a more efficient enforcement system in order to avoid evasion and fraud in comparison to physical stores which need to be audited on a case by case basis.

Aiming at highlighting the role of platforms in VAT/GST enforcement, this contribution is structure in five sections. Section 2 offers an overview of how online distant sales can challenge the enforcement of VAT/GST rules. The risks of VAT/GST evasion and fraud emerging from these sections clearly represent a thread of revenue losses which shall be used for redistributive and welfare purposes. For example, in the European context, VAT represent one of the major resources for the European public finances. Section 3 will go through the role that e-commerce platforms can play in enforcement mechanisms. These roles have been recently indicated in an OECD report. Through the inclusion of platforms active in the e-commerce sector, the goal is to improve VAT/GST enforcement and consequently to reduce the risks of evasion and fraud and increase the possibility for the use of the resources for the welfare and social justice purposes. Finally, section 5 aims at offering a comparative analysis of the new legislations specifically targeting platforms facilitating distance sales of goods. The countries object of this analysis are: the United Kingdom, Germany, Australia. Moreover, this section includes a brief description of the new EU VAT e-commerce package and an overview of which are the benefits and limits of this approach in ensuing the equal application of enforcing provisions among businesses and in safeguarding public resources.

2. Contextualizing VAT Evasion and Fraud Risks in the Digital Arena

First introduced around less than 70 years ago, VAT is now a pivotal component of tax systems around the world.[898] In the European Union the common Value Added Tax (VAT) is major and growing source of revenue raising over 1 trillion euros in 2015, corresponding to 7% of EU GDP. Moreover, the EU's own resources are also based on VAT and VAT has been seen as an efficient way to raise revenues also in developing countries[899].

However, the development of the digital economy characterized by the lack of territoriality and a possible global reach for customers, means that companies will also incur in VAT liability in multiple countries and there will be uncertainty on their tax obligations related to the transactions they have carried out with or without the use of an online platforms. Indeed, distance sales and services supplied through digital platforms have offered the opportunity even to small entrepreneurs and non-entrepreneurs to connect and supply their own goods and services to a wider audience of potential customers located all over the world. However, this potential business use of the internet, has also undermined the possibility to apply and enforce VAT rules. For example, if a consumer who is already employed starts regularly to sell some of his/her creations as a hobby on eBay, it is questionable if he/she is still to be treated as a consumer or as an entrepreneur, since exemptions for small entrepreneur might find application. Even if outside the scope of this contribution, the same issue arises also in the case of platforms operating in the so-called "sharing economy" and facilitating peers to peers transactions between privates.[900] In this

[898] Liam Ebrill, Michael Keen, and Victoria Perry, *The Modern VAT*, (IMF, 2001).
[899] Many African countries have adopted it during the 90s, please see Sijbren Cnossen, 'Mobilizing VAT revenues in African countries', (2015) 22 International Tax and Public Finance 6 , 1077–1108. While in 2019 it was adopted by Bahrain and in 2018 by Saudi Arabia and during this year other gulf countries are expected to adopted it. Yass Alkafaji and Omaima Khanfar, 'Bahrain/Kuwait/Oman/Qatar/Saudi Arabia/United Arab Emirates/Gulf Cooperation Council - The New VAT Regime in the Gulf Cooperation Council Countries', (2017) 71 Bulletin for International Taxation 10; Howard Hull and Roberto Scalia, 'International/GCC - GCC VAT - International Goods', (2018) 29 International VAT Monitor 2.
[900] For more literature on the VAT/GST challenges of the sharing economy please see: Giorgio Beretta, 'VAT and the Sharing Economy', (2018) 10 World Tax Journal 3, 381-425; Ivo Grlica, 'How the Sharing Economy Is Challenging the EU VAT System', (2017) 28 International VAT Monitor 2, 124-131; Carrie Brandon Elliot, 'Taxation of the Sharing Economy: Recurring Issues, (2018) 72 Bulletin for International Taxation 4.

case, for example, there might be a consumer who gives rides through BlaBla car or rents his/her home on Airbnb. In the last years, many doctrinal contribution and institutional papers have been written to highlight the difficulties in determining whether the transaction taking place through an electronic interface and between two parties were to considered as taxable and if the person providing the service (e.g. renting a house) was to be considered as a taxable person.[901] Platforms facilitating "entrepreneurship" which enables individuals carrying out C2C transactions to transform themselves in entrepreneurs (consequently carrying out B2C transactions) made much harder to determine whether there is a taxable person or no.[902] Moreover, while in the occasional trade of goods or occasional supply of services there are no tax consequences, if the occasional trade or supply becomes a sustainable economic activity, VAT liability arises.[903] The line between acting as an entrepreneur/taxable person and as a private becomes blurred. On this point, relevant case law shows that buying and selling goods/services could quickly be seen as a sustainable economic activity for VAT and a source of income for income tax, including administrative obligations.[904] Moreover, as C2C supplies are out of the scope of VAT, this also led to competition distortions. Finally, these platforms continuously develop and become more and more hybridized by offering other services, such as debt collection, transportation and selling under their own name and for their own risk to consumers.

In such a kaleidoscopic economic environment, it is then questionable whether the conditions determining the tax treatment of transactions are clear to everyone. Private persons are not always aware of the consequences of their internet activities while other times they consciously choose not to register for tax purposes. This problem mainly arises in countries where the law does not provide a turnover-related VAT exemption for small businesses or provides a very low threshold for this exemption (e.g. the Netherlands and Sweden).[905] Online platforms, although not liable for VAT in these cases, have an information position from which the tax liability of the users of the platform could be traced.

Furthermore, in relation to distance sales involving third countries and low value goods (whose value is less than 22 EUR), there might be serious risks for VAT fraud to arise. In the example reported by van der Hel-van Dijk & Griffoen, a EU based consumer purchases via online trading platform based in the US a phone from a Chinese supplier and pays 300 EUR. The consumer pays via credit card or another payment system supported by the platform. After deducting its commission, the platform pays the remaining amount to the Chinese supplier. Next, a courier company transports the phone to the logistics centre of the platform which is located in another EU country different from the one of the consumer. From there, the phone is then sent to a distribution centre located in the country of the consumer and then delivered to him/her. The platform has included in its terms and conditions that the supplier of the goods is responsible for any tax obligations arising from the transaction.[906] In this case, the Chinese supplier, although he has received 300 EUR for the phone, might present the phone to customs as a shipment of 15 EUR. Consequently, this shipment would be subjected to the VAT exemption scheme for small consignment and import duties, consequently no VAT and import duties will be paid (the threshold for import duties is 150 EUR). In these cases, third countries supplies could be able to offer on the market cheaper products than the EU competitors and this will lead to distortions of

[901] Marie Lamensch, 'European Commission´s New Package of Proposals on E-Commerce', Op. cit.; Aleksandra Bl, 'EU VAT Proposals to Stimulate Electronic Commerce and Digital Publishing', Op. cit.; Madeleine Merkx and John Gruson, 'Definitive VAT Regime: Ready for the Next Step?', Op. cit.; Laura Mattes, 'VAT Aspects of Cross-Border Transactions in the BEPS Era', Op. cit.; Elvire Tardivon-Lorizon and Amanda Z. Quenette, 'Indirect Taxation of E-commerce', Op. cit.; Han Kogels, 'Making VAT Fit for the Digital World', Op. cit.

[902] Ehrke, Aspekte grenzüberschreitenden digitalen Wirtschaftens in der Umsatzsteuer, DStJG 42

[903] Ibid.

[904] Francesco Cannas, Calogero Vecchio and Davide Pellegrini, 'Policy note: A New Legal Framework Towards a Definitive EU VAT System: Online Hosting Platforms and E-Books Reveal Unsolved Problems on the Horizon', (2016) 46 Intertax, 8/9, 690–698.

[905] Lisette van der Hel-van Dijk and Menno Griffoen, 'Eu VAT Note: Online Platforms: A Marketplace for Tax Fraud?, (2019) 47 Intertax 4, 393.

[906] Lisette van der Hel-van Dijk and Menno Griffoen, Op. Cit., 394.

competition and discrimination. Thus, the platform information position is crucial in order to determine the tax liability.

Other similar risks with reference to VAT and import duties regard also the reverse-charge mechanism on import or customs procedure which are used in combination with an intra-community delivery existing only on paper to so-called "missing traders", while the goods after import are actually delivered to the consumer. This means that this type of transactions might look on paper as B2B transactions, but they are actually are B2C. [907]

These risks are clear examples of the possibility to commit VAT evasion and fraud by exploiting the loopholes created by the difficulties to apply traditional rules to these new business models. For this reason, both at EU level and at OECD level the debate on how to avoid possible VAT evasion and fraud due to new business models involving platforms has become extremely relevant. The increase use of e-commerce allows different schemes for tax fraud and a table showing these different possibilities has been drawn by van der Hel-vanDijk & Griffoen (2019). However, focusing on the issues rising from distance sales and the new roles attributed by the EU and the OECD to platforms in this context the main issues are the following:

1) First of all, private individuals might not acknowledge they are liable to pay tax. This means that there are many potential private individuals with a relatively small tax interest (depending on a case by case basis) who offer goods and/or services online, but who are presumably ignorant of the tax and administrative obligations. Moreover, since we are talking of many individuals and low amounts it might be difficult for tax authority (which does not have all the needed information) to find out the business carried out by that private person.
2) A second risk is associated with (foreign) entrepreneurs/taxable persons who do not register. In the case of distance sales of goods, the expectation is that small entrepreneurs/taxable persons who trade via online platforms will remain outside the view of tax authorities since supervision is quite difficult to implement.
3) Thirdly, issues might arise from online platforms with multiple activities and various agreements of VAT taxation of transactions (e.g. application of different tax rates, supplying of goods and services which are exempted together with other goods and services which are not)[908].
4) An additional risk concerns online sales of imported low-value goods.[909] According to a recent OECD report, because many jurisdictions exempt low value parcels from online sales from VAT/GST, administrative costs associated with collecting the VAT/GST on the goods is likely to outweigh the VAT/GST that would be collected.[910] The threshold varies among jurisdiction but due to the increasing volume of low-value goods on which no VAT/GST is collected, this has resulted in a decreased VAT/GST revenue and unfair competitive pressures on domestic retailers on which VAT/GST is applied.[911] Ultimately, this might lead to incentivization for domestic suppliers to relocate in offshore jurisdiction in order to sell value goods free of VAT/GST.[912]

Despite the challenges in VAT/GST enforcement which might arise in the context of online sales through platforms, the fact that online platforms act as an intermediary between supply and demand, as a debt collector and as a business themselves can be crucial in fighting VAT evasion

[907] Ibid.
[908] Lisette van der Hel-van Dijk and Menno Griffoen, Op. Cit., 396-397.
[909] OECD, Addressing the Tax Challenges, Op. cit.
[910] Ibid.
[911] Ibid.
[912] Ibid.

and fraud. Moreover, they have access to information regarding the transactions taking place through them, which is an additional reason to consider them a valuable resource.[913]

3. Which roles can Online Platforms play in the fight against VAT fraud?

The role of third parties in tax matters has tremendously increased in the last decades.[914] Indeed, third parties can be involved in different ways in the collection of taxes. Offline, third parties have been playing a pivotal role as tax withholders and as tax-related information providers. For example, in many states, financial institutions withhold taxes on financial capital gains. In this case, financial institutions are in a better position than the taxpayers to verify whether there have been capital gains on taxpayers' financial assets. Moreover, they might be obliged to provide relevant accounts' and transactions' information to the tax authorities. Another classic example of third party is the employer which withhold the tax on wage for the employees. Even, the same basic structure of the VAT is built in such a way that even if the consumer is the subject which borne the tax, the tax is paid by the entrepreneur which represents a third party between the consumer and the State. As emerges from these examples, tax systems have already been involving third parties in traditional tax collection activities. Thus, it is not surprise that a similar approach will be considered also in the case of the e-commerce taking into consideration the relevant players in this arena.

In the e-commerce context, there are different reasons why it is beneficial for States to involve platforms in VAT collection. Platforms have information on the overall transactions occurring through them. Moreover, in most of the cases, payments go through the platform. Finally, as stated in the OECD 2019 report the reliance on digital platforms for VAT/GST collection may also be motivated by the fact that digital supply chains are often long and complex, and that suppliers in this chain may not be aware of the roles of the various parties in the chain.[915] An approach that relies on the digital platform to collect and remit the tax that is due on the ultimate supply to the end customer may be expected to provide an efficient solution for tax administrations and the experiences of jurisdictions who have already adopted this model appear to support and confirm that expectation.[916]

The growth of international online B2C trade, both in volume and in numbers of participants, is one of the main factors which are now putting to the test for VAT/GST collection.[917] In this regard, in many countries one of the first reaction of tax legislators was to tackle the supply of digital services and digital goods. The VAT/GST on cross-border business-to-business (B2B) trade in services and intangibles, which also continues to grow, is generally collected through a reverse-charge or self-assessment mechanism. These self-assessment mechanisms generally work well in a B2B context – however, they are largely ineffectual in a B2C context and this is becoming more and more relevant in light of the exploding B2C online trade. For example, within the European Union, already in 2015, the so-called Mini One Stop Shop (MOSS) scheme entered into force with regard to digital services. However, issues regarding the online sales of goods through platforms, especially concerning low value goods were not yet solved. In the European

[913] E.C.J.M. van der Hel-van Dijk and M.A. Griffoen, Op. cit. 391–401.
[914] Michael Doran, 'Tax Penalties and Tax Compliance', (2009) 46 Harvard Journal on Legislation, 143; Leandra Lederman, 'Statutory Speed Bumps: The Roles Third Parties Play in Tax Compliance', (2007) 60 Stanford Law Review, 695; Tina Ehrke-Rabel, 'Third parties as suplementaries to transparency', in Funda Busran and Johanna Hey, *Transparency in taxation*, (EATLP Congress Papers, 2018).
[915] OECD, The Role of Digital Platforms, Op. cit. 56.
[916] OECD, The Role of Digital Platforms, Op. cit. 58.
[917] OECD, The Role of Digital Platforms, Op. cit. 4.

VAT context, this has recently led to two type of changes (additionally to the Mini One Stop Shop - MOSS system which is already in place from 2015). On one hand, the MOSS scheme will become a One Stop Shop (OSS) since it was extended to all type of goods and services. On the other hand, the attention started to be focused on platforms and their possible role in facilitating the collection and accounting of VAT.[918] Indeed, the important role in collecting and accounting VAT on behalf of foreign suppliers was already highlighted by OECD in a 2017 report.[919] In the same year, the European Union, as it will be further analyzed in section 4, through the adoption of the VAT e-commerce package, had introduced new provisions deeming the platform as liable for VAT. These provisions will enter to force in 2019 and in 2021, the OECD released a report containing guidelines for States which want to involve platforms in VAT collection.[920]

3.1 Full VAT Liability

The first role for online platforms individuated by the OECD is the possibility for States to introduce deeming provision under which online platforms are considered full VAT liable for sales occurring though their platforms. Under the full VAT/GST liability regime, the digital platform is in principle required to assess, collect and remit the VAT/GST to the tax authorities and comply with the VAT/GST reporting and other obligations as required under the VAT/GST rules in the taxing jurisdiction.

3.1.1 Indicators for the viability of a full VAT/GST liability regime

In order to design full VAT liability provisions, the OECD report analysis possible indicators based on the functions performed by the digital platforms in order to verify if this scheme can be viable for that type of platform.

The first suggestion concerns the term “digital platform” which, since we can expect to be evolving in the next years, should be used as a generic term referring to those actors which in the context of online sales carry out functions considered as essential for their enlistment. Despite the fact that the types of digital platforms and their business models are continuously evolving, they generally build their activities on key functions. Thus, focusing on indicators based on the functions performed by digital platforms has the advantage to be more flexible to future changes. Among the possible criteria that a tax authority might use when determining the digital platform it wishes to enlist in the collection of VAT/GST under a full VAT/GST liability, one consists in facilitating groups of customers (buyers and sellers) to interact directly with each other and enter into transactions through the platform. Another indicator of the viability to apply a full VAT/GST liability regime can be found in whether a digital platform is in the position to comply with this regime depending on its business and delivery model. In this case frequent consultation with individual platforms are seen by the OECD as an added value.[921]

However, more generally and according to the OECD report, it is reasonable to assume that a platform will be in a position to comply with these obligations if:

[918] Eli Hadzhieva, Impact of Digitalisation on International Tax Matters. Challenges and remedies, European Parliament, February 2019.
[919] OECD, Mechanisms for the Effective Collection of VAT/GST Where the Supplier Is Not Located in the Jurisdiction of Taxation (2017).
[920] First comments on the VAT e-commerce package can be found in: Marie Lamensch, ‘Adoption of the E-Commerce VAT Package: The Road Ahead Is Still a Rocky One’, (2018) 27 EC Tax Review 4, 186-195; Marie Lamensch, ‘European Commission’s New Package of Proposals on E-Commerce: A Critical Assessment’, (2017) 28 International VAT Monitor 2, 138-146; Madeleine Merkx, John Gruson, Naomi Verbaan, Bart van der Doef, 'Definitive VAT Regime: Stairway to Heaven or Highway to Hell?' (2018) 27 EC Tax Review 2, 74–82; Aleksandra Bal, 'Germany: New VAT Compliance Obligations for Online Platforms', (2019) 28 EC Tax Review 2, 114-119.
[921] OECD, The Role of Digital Platforms, Op. cit., 43-44.

- the platform holds or has access to sufficient and accurate information as required to make the appropriate VAT/GST determination; and

- the platform has the means (is able) to collect the VAT/GST on the supply.

Moreover, to address cases where more than one digital platform in a supply chain is eligible for a full VAT/GST liability regime, tax authorities could consider applying hierarchy rules. Certainly, any criteria used to define the digital platforms' eligibility for a full VAT/GST liability regime will need to be reviewed regularly in light of technological and commercial developments to ensure their efficiency and effectiveness.[922]

3.1.2 Scope of full VAT/GST liability regime

Regarding the scope of the full liability regime, with regard to the type of suppliers (foreign/domestic) we could assume that for tax authorities it might by harder to enforce compliance on potentially millions of foreign underlying suppliers. Nevertheless, distinguishing between foreign and domestic supplier might increase compliance complexities for the digital platforms which might have to carry out compliance activities in order to distinguish between the two types of suppliers. At the same time, also for tax administrations, their auditing activities to platforms might become challenging since they will have to assess the location of the underlying suppliers and in the case of domestic suppliers whether they have remitted the due local VAT/GST on the sales carried out online. Despite these difficulties, the OECD report still admits that certain countries might prefer to limit the scope to foreign underlying suppliers. However, in this case, there might be an agreement under which the platform agrees with the domestic underlying suppliers that it will be fully liable for VAT/GST obligations.[923]

With regard to the scope, another relevant distinction is the one between limiting it to services and/or goods. So far, certain jurisdictions have decided to limit the scope of the full VAT/GST liability regime to digital platforms acting as remote digital/electronic supplies by foreign suppliers. This choice, according to the OECD study might find its justification in ensuring the effective collection of VAT/GST on supplies in sectors where tax revenue was considered to be most at risk while aiming to avoid changes for suppliers and tax administrations in areas where there is no compelling need to deviate from existing collection regimes.[924]

Particularly important in the last year has been the possible introduction of a full VAT/GST liability regime for the collection of VAT/GST on the supplies of goods from online sales that are directly connected with an importation of these goods. As highlighted in section 2 many jurisdictions apply an exemption from VAT/GST for imports of low value goods because the administrative costs associated with collecting the VAT/GST on the goods would be outweighed by the collected VAT/GST. These exemption thresholds were generally established before the advent and growth of the digital economy. However, these might need to be reviewed under the light of recent developments of the digital economy since data suggest that such imports of low-value goods represent the vast majority of packages that reach the borders from online trade, and create increasingly significant logistical challenges for customs authorities to process.[925] Thus, many jurisdictions around the world are considering full VAT/GST liability regime for digital platforms as a potential approach to increase the efficiency and the effectiveness of VAT/GST collection on imported low-value goods. According to OECD, VAT/GST liability regimes on imports below the de minimis customs threshold is motivated by the consideration that if digital platforms collect and

[922] OECD, The Role of Digital Platforms, Op. cit., 46-48.
[923] OECD, The Role of Digital Platforms, Op. cit., 54 and 55.
[924] OECD, The Role of Digital Platforms, Op. cit., 56-58.
[925] As reported in the 2019 OECD report, parcel volume increased from 44 billion in 2014 to 65 billion in 2016 across 13 major markets33 and continues to increase at a growing rate that is calculated to be 17-28% each year between 2017 and 2021. See Pitney Bowes, Global Parcel Shipping Index (2017). http://news.pb.com/article_display.cfm?article_id=5784

remit the VAT/GST on imports of low value goods, customs authorities will have to intervene less or not at all in the collection processes for imports that are not subject to customs duties. In this way, the costs related to the collection of VAT/GST on imports of low-value goods will also decrease and allow customs authorities to use the saved resources for other tasks. VAT/GST on imports of goods above the customs threshold can then (continue to) be collected together with customs duties and taxes under normal customs procedures with imports of goods that are directly connected to online sales.[926]

Regarding the type of transactions (B2B and B2C), the full VAT/GST liability regime could apply for the collection of VAT/GST on both categories of supplies performed via a digital platform. However, where in a jurisdiction different VAT/GST rules are applied for B2B and B2C supplies, such as different rules for determining the place of taxation and for collecting the tax, knowing the status of the customer (business or non-business) is indispensable for determining the correct VAT/GST treatment of a supply. Consequently, in the case of implementing a full VAT/GST liability regime for digital platforms, these will have to be guided by tax authorities on how to distinguish between the B2B and B2C supplies where required.[927]

Normally, we can expect platforms to be allowed to rely on the basis of information to which they have access or to which they can be reasonably expected to have access when making such a distinction. Moreover, the latest OECD report states that where a digital platform, acting in good faith and having made reasonable efforts to obtain the appropriate evidence, is unable to establish the status of its customer, a presumption could be applied that the customer is a non-business customer, in which case the rules for B2C supplies would apply.[928] Efforts could include for example, the provision by the customer of a VAT/GST registration or identification number which has been proven to be invalid (e.g. by looking at the online register of entrepreneurs provided by tax authorities), the digital platform may presume that the customer is a non-business and apply the rules for B2C supplies. Differently, digital platforms may operate under the assumption that the underlying suppliers that are selling through their platform are businesses unless they have information to the contrary. Key information elements which can be considered for the determination of the tax treatment of the transaction might include the customer status (when jurisdictions differentiates between B2B and B2C), the nature of the supply, elements to determine the place of taxation and/or the applicable VAT/GST collection regime (e.g. IP address), the AT/GST exemption threshold for VAT/GST registration and/or collection purposes, the value of the supply and the applicable VAT/GST rate and the point at which VAT/GST liability arises.[929]

Under the full liability scheme, compliance burdens for platforms have substantially increased. It is important then that tax authorities give platforms access to updated information concerning their obligations and compliance processes in order to timely comply with tax obligations. Another measure which can be implemented to rebalance the compliance burdens on the platform is the adoption of a rule reducing or eliminating digital platforms' liability for mistakes resulting from reliance on inaccurate information, if they can supply evidence of their good faith and of their reasonable efforts (which depend on the ones individuated by tax authorities) to secure the accuracy and reliability of the information on the basis of which they have acted. Interesting is also the OECD encouragement to the possibility for jurisdictions to make the compliance information available in machine readable format in order to facilitate compliance in multiple jurisdictions by reducing the need for human intervention and manual input.[930]

[926] OECD, The Role of Digital Platforms, Op. cit., 63.
[927] OECD, The Role of Digital Platforms, Op. cit., 65-68.
[928] OECD, The Role of Digital Platforms, Op. cit., 69
[929] Ibid.
[930] OECD, The Role of Digital Platforms, Op. cit., 73.

3.1.3 Practical aspects of VAT/GST collection under the full VAT/GST liability regime

Another crucial element in the design of a full VAT/GST liability regime for online platforms is represented by the individuation of the taxing point, in other words the determination of the time at which the platform will have to account for the VAT/GST on the supplies carried out through its platforms for which it is now has VAT/GST liable.[931] Since under the full liability regime, digital platforms will have to account for the VAT/GST on supplies going through them but without being the actual underlying supplier, platforms may not always have all the needed information to determine the taxing point according to standard rules (e.g. time of actual supply, performance or delivery, time of receipt of payment). Moreover, even where platforms have this information, it is an heavy burden on them to proceed to individual determinations for each of the potentially millions of supplies for which it has VAT/GST liability. The OECD suggested solution for this type of issues is to define the taxing point at the time at which the confirmation of the payment is received by or on behalf of the underlying supplier which consists in the time at which the payment has been accepted or authorized by or on behalf of the underlying supplier even if it does not necessarily mean that the actual money transfer has been made or processed.[932] For imports of low- the taxing point usually is the time of confirmation of the payment, which is generally at a time prior to shipping or arrival of goods at the border, thus VAT collection will take place earlier on the supplies of imported goods from online sales away from the border (which is currently the general practice). Establishing the taxing point at the time of confirmation of payment both for the supply of services and for the supply of goods may simplify compliance under the full VAT/GST liability regime for platforms. Regarding the more practical process for the collection and remittance of VAT/GST, the OECD report refers to different scenarios. The main distinction among the different scenarios is the one where customer pays the VAT/GST inclusive price to the platform and the scenario where VAT is paid directly to the underlying supplier by the customer. In the first case, the platform will in principle remit the due VAT/GST to the tax authority and the balance consisting in the sales price minus fees and commissions to the underlying supplier. Differently, if the customer pays VAT/GST directly to the underlying supplier the platform will have to recuperate the due VAT from the supplier (in addition to possible fees and commissions). In this second case, the advice to tax authorities where this system is in place is to implement an appropriate bad debt relief arrangement to limit the potential risk of default by underlying suppliers in remitting the VAT/GST to the digital platform provided that reasonable efforts to ensure compliance have been made by the platforms. In this case there is an additional risk for cascading effect. Thus, it is essential to avoid the application of VAT/GST on the recovery of the VAT/GST amount by platforms from the underlying suppliers, while normal VAT/GST rules are applied to commissions and/or fees collected by the platform for its services from the underlying supplier.[933]

3.2 Other possible valuable roles for platforms

Beside the possibility to rely on platforms for VAT collection through full VAT liability schemes, the OECD recommends other valuable approaches which can be adopted by States and which involve platforms. In particular these roles include information sharing obligation, the education of the suppliers, formal agreements between tax authorities and platforms and providing the option for platforms to voluntarily take on the obligation as collector of VAT/GST.

3.2.1 Information sharing obligation

[931] OECD, The Role of Digital Platforms, Op. cit., 75 and ff.
[932] OECD, The Role of Digital Platforms, Op. cit., 76.
[933] OECD, The Role of Digital Platforms, Op. cit., 78-79.

In the case of information sharing obligation, the online platform would be required by law to provide information to tax authorities in order to assist them in the enforcement of VAT rules and the platform must not necessarily liable or have an active role in the collection and remittance of VAT/GST to the tax authorities. The digital platform could be asked to provide this information either on a regular basis, upon request or spontaneously, e.g. in cases of suspicious activity[934]. The tax authority will also need to identify the type of needed information for the efficient and effective VAT/GST collection on online sales and whether it has the human and technical resources to process the collected data.[935] Moreover, with relation to European taxpayers, tax authorities will have also to be compliant with the GDPR provisions. Information gathered in this way, can also be used to for advanced risk analysis as a means to target non-compliance. Nevertheless, tax authorities shall minimize risks of unnecessary duplication of information obligations for digital platforms, since the information is already being collected by other means or provided to other authorities (e.g. customs authorities).[936] Regarding the type of information/data which need to be shared, these should certainly be the ones which are already available to digital platforms in the normal course of their business activities and which are at the same time relevant for VAT/GST compliance purposes.

The information sharing obligation might also be adopted to complement to the full VAT/GST liability regime. In this case, the application of any additional information sharing obligations shall be limited to the digital platforms that are not covered by those other measures. If the information sharing obligation is adopted as a stand-alone measure it should then apply to all types of digital platforms such as the ones who take an integral role in the supply e.g. online marketplaces, transfer buyers to sellers (click-through or shopping referral platforms), contract or agree to listing or advertising items for sale in any forum or medium, receive a fee, commission and/or other consideration for listing of advertising items or process payments.[937] Even if this type of obligation seems to be beneficial for the efficient collection of VAT/GST on platforms, digital platforms may be located outside the taxing jurisdiction and the enforcement of such an obligation against foreign digital platforms represent a difficult challenge. Thus, if this type of obligation is not combined with already existing administrative co-operation mechanisms among jurisdictions, it might be vain.

According to the OECD report two broad options of the information sharing obligation could be considered:

1. Provision of information on request. In this case, a jurisdiction will require an online platform to keep records of the sales subjected to VAT/GST and that this information shall be available on request. Records might refer to a specific category of sales taking into consideration a given period or a particular supplier or in respect of a specific transaction
2. Systematic provision of information. In this case, a digital platform will systematically and periodically provide information on online sales carried out via the platform to the tax authority of the relevant jurisdiction of taxation which will specify the format and the information required. The obligation might also be limited to a specific sales e.g. goods above a certain value and the submission period will vary depending on the use of those data by the tax administrations.[938]

3.2.2. Education of suppliers using digital platforms

Very often VAT/GST obligations represent a challenge to businesses engaging in cross border e-commerce and many times businesses do not know the applicable VAT/GST rate for a particular good or service in the taxing jurisdiction, invoicing, recordkeeping and reporting obligations. This

[934] However, it should be clearly determined when the platform should consider an activity as suspicious.
[935] OECD, The Role of Digital Platforms, Op. cit., 117-119.
[936] OECD, The Role of Digital Platforms, Op. cit., 120.
[937] OECD, The Role of Digital Platforms, Op. cit., 123.
[938] OECD, The Role of Digital Platforms, Op. cit., 128.

issue is even amplified when a business activity is carried out in various countries. Earlier guidelines[939] and the Collection Mechanisms Report[940] have previously highlighted the importance of a proper communication strategy for achieving appropriate compliance levels by foreign suppliers in the taxing jurisdiction.[941]The availability of readily accessible and easily understood guidance for taxpayers has been proved to be beneficial for compliance levels by foreign suppliers and increasing proactive taxpayer engagement especially with reference to jurisdictions using simplified registration and compliance mechanisms for the collection of VAT/GST on inbound supplies.[942] Indeed, tax authorities might not be able to directly reach out to suppliers outside their jurisdiction to advise them of their obligations, particularly in the case of world active platforms where there might be millions of suppliers. Thus, platforms giving access the global market might be used as communication channels to provide accurate and timely information to underlying suppliers on their VAT/GST obligations.[943] On this point, the report underlined the capacity of digital platforms to communicate with the often large numbers of suppliers that sell through their platform offers as unique opportunity to tax authorities for disseminating information on these suppliers' VAT/GST obligations, such guidelines, direct messages concerning notifications of changes in obligations, the organization of webinars and advice from tax authorities via a platforms community forum.

3.2.3 Formal co-operation agreements

Tax authorities might also decide to enter into formal agreements with digital platforms, similarly to the co-operative compliance concept[944]. A variety of measures and approaches to involve digital platforms in maximizing VAT/GST compliance levels in online sales can be included in the content of the agreement. Often the agreement might include also information sharing (periodic and spontaneous) and education (including using the platform as a conduit to communicate with underlying suppliers on compliance obligations, etc.), as well as alerting the tax authorities and platforms to instances of fraud, and responding quickly to notifications by a tax authority where underlying suppliers are found to be in breach of their VAT/GST obligations.

This type of formal co-operation agreement could be useful especially in cases where the digital platform is not liable for or plays no role in collecting and remitting the VAT/GST. Moreover, this type of agreement can represent a first intermediary step before implementing a full VAT/GST liability regime for digital platforms. Through co-operation agreement tax authorities will be able to efficiently liaise with a digital platform and vice versa to support compliance by the underlying suppliers. Furthermore, OECD highlights that since the platform could reach out to a dedicated contact point in the tax authority and making this type of agreement public might induce consumers to consider the platform as "safe" with regard to VAT/GST obligations especially considering consumers purchasing goods online who may pay VAT/GST to a platform at the point of sale and consequently might expect that they will not face a further VAT/GST liability on importation. Indeed, in the draft of the agreement terms, conditions and the timeframe of the agreement should be clear, particularly in respect of any legal aspects (e.g. joint and several

[939] OECD, International VAT/GST Guidelines (2017).
[940] OECD, Mechanisms for the Effective Collection of VAT/GST where the Supplier Is Not Located in the Jurisdiction of Taxation, OECD publishing, (2017).
[941] OECD, International VAT/GST Guidelines, (2017).
[942] In this context, it has been sustained that proactive taxpayer engagement and education programs can help ensuring that taxpayers clearly understanding of their obligations. IMF/OECD Tax Certainty (2017) http://www.oecd.org/tax/tax-policy/tax-certainty-report-oecd-imf-report-g20-financeministers-march-2017.pdf Accessed 09 September 2019.
[943] It is notable that several digital platforms have spontaneously taken initiatives to communicate with their underlying suppliers on their VAT/GST obligations in the various taxing jurisdictions – this includes the operation of online forums for the platforms' communities of suppliers whereby information on general regulatory issues including taxation can be shared.
[944] The 2013 report "Co-operative Compliance: A Framework: From Enhanced Relationship to Co-operative Compliance", based on a detailed analysis of practical experiences of several countries recommended that tax authorities develop a relationship based on trust and co-operation. The report is based on a detailed examination of the practical experiences of countries that have established this type of relationship. OECD, Co-operative Compliance: A Framework: From Enhanced Relationship to Cooperative Compliance (2013).

liability provisions, response times for information requests, mutual contact details, etc.). Terms shall be realistic and proportionate since it is a voluntary and co-operative based agreement.

3.2.4 Platforms acting as a voluntary intermediary

Tax authorities might also allow platforms to act voluntarily as a third-party service provider on behalf of underlying suppliers. This scheme might be especially relevant in the case where a platform is liable only for certain supplies and it can also operate as complementary to the full VAT/GST liability regime for transactions which are not covered by that obligation.[945] Jurisdictions might foster platforms to act as voluntary intermediaries also to extend the scope of liability beyond the statutory requirement. In the case of full liability applicable only to imports of goods below the customs de minimis threshold, allowing the digital platform to voluntarily collect and remit the VAT/GST on behalf of the underlying supplier above this de minimis threshold could be also considered.[946] This type of agreement has the potential to increase compliance, and it becomes even more beneficial in the cases where platforms are not located within the borders of the relevant jurisdiction. Consequently, this might push jurisdictions to establish a simplified registration and compliance regime to facilitate compliance and convince more platform to act as voluntary intermediary.[947]

4. Interim Conclusions

As it has emerged in these previous sections, e-commerce is rapidly increasing. However, at the same time challenges related to the effective enforcement of VAT/GST rules in the online world are also arising. There are four main risks which have been highlighted in literature and by the OECD. Firstly, the fact that individuals might not be aware of their tax liability. This is the case, for example of private individuals which offer goods and/or services online and is not aware that when the threshold has been exceeded, VAT/GST obligations arise. Secondly, foreign entrepreneurs do not register. Thirdly, the possibility of VAT/GST evasion in the case of businesses offering multiple activities which are subject to different rules regarding the tax rates and possible exemptions. Finally, VAT/GST evasion and fraud risks together with possible distortions in the market between local and foreigner businesses, in relation to the exemption of imported low-value goods.

In order to minimise these risks, the OECD has individuated five different ways in which platforms can have a valuable role in VAT/GST enforcement mechanisms, namely: through the adoption of a full liability regime for platforms facilitating the distant sales of goods; by imposing information sharing obligations on platforms; by Using digital platforms as educators for taxpayers; through the conclusion of co-operation agreements between tax authorities and platforms; through the willingness of platforms to act as an intermediary.

On the basis of this last and previous reports from the OECD emphasising the possible ways in which platforms can be included in the enforcement phase and taking into account previous experiences regarding the supply of digital services, different States around the world have implemented new legislation targeting platforms also in the case of distant sales.

In the following section, a comparative analysis of these new legislation aims at offering an overview of how these rules were designed and how platforms were involved in the enforcement mechanisms while at the same time highlighting benefit and limits of such provisions.

[945] OECD, The Role of Digital Platforms, Op. cit., 147-148.
[946] OECD, The Role of Digital Platforms, Op. cit., 149-151
[947] OECD, The Role of Digital Platforms, Op. cit., 153

5. Case studies: UK, Germany, Australia and the new European VAT e-commerce package

As it emerged from section 3, platforms can play different roles in the VAT/GST enforcement procedures. The advantageous position they have as third party between businesses and customers led some jurisdictions to explore the possible inclusion of platforms in their VAT/GST enforcement strategy. In the following sections, the aim is to provide an overview of how different jurisdictions have been adopting new legislation addressing the VAT/GST enforcement by including provisions directly targeting platforms facilitating the sale of goods. The states that will be object of the following comparative analysis are the United Kingdom, Germany and Australia.

Lastly, starting from 2021, new provisions concerning e-commerce and platforms facilitating the online sales of goods will also enter into force in the European Union. Indeed, these provisions, to be implemented in all the Member States, will have a wider geographical scope and might influence other States around the world.

5.1 United Kingdom

Within the European context, the UK was the first State to implement a VAT rules enforcement regime involving platforms. The UK has introduced in September 2016 and strengthened in March 2018 an approach targeting overseas traders. The UK approach, introduced through the Finance Bill 2016, consists in checking overseas traders' VAT obligations and obliging marketplace operators to perform validity checks concerning VAT ID numbers given by the registered traders. First of all, the HMRC will contact the overseas business which is not VAT compliant. If this business will not follow the directions given by the HMRC, then the HMRC will contact the platform through which the overseas business is trading in the UK and inform it that the platform itself might be held jointly and severally liable for the VAT in respect of the overseas business' future taxable sales through that platform.[948] The platform will not be held jointly and severally liable for the non-compliance of the overseas business if the platform will be able to secure the overseas business compliance or will remove it from its online marketplace.[949] Typically, the platform will be given a time frame of 30 days to comply by taking one of this two actions.

After the Finance Bill 2016 introduced these provisions focusing on overseas businesses, the Finance Bill 2017/2018 extended the scope of this new liability rules to online market places in cases where:

a) A UK business selling goods via online platforms fails to account for after being notified by the HMRC;
b) An overseas business which sells good via the online platform fails to account for and the online platform knew or should have known that the business should be registered for VAT in the UK.

Consequently, in the latter case, the platform will be held jointly and severally liable not only from the moment in which it has received the notice from the HMRC but from the moment in which it knows about the overseas business not being compliant to UK VAT law. In order to avoid this liability regime, the platform has 60 days from the moment in which it is aware of the overseas business' non-compliance to ensure that this business will not sell goods to UK consumers via its

[948] Aleksandra Bal, 'Managing EU VAT Risks for Platform Business Models', (2018) 72 Bulletin for International Taxation 4; HMRC, VAT: Extending joint and several liability for online marketplaces and displaying VAT numbers online, Guidance note, 2018, chrome-Extension://oemmndcbldboiebfnladdacbdfmadadm/https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/661739/7183_guidance_note_-_extending_joint_and_several_liability.pdf
[949] HMRC, VAT: Extending joint and several liability for online marketplaces, Op. cit.

online marketplace. During this time, the platform can inform the overseas business about its UK VAT obligations.[950]

Regarding the second type of obligation, the Finance Bill 2017-2018 has also required platforms to verify the VAT numbers displayed on their websites, checking the validity of those numbers and the correct relation to a seller.[951]

5.2 Germany

In December 2018, Germany has adopted new rules providing electronic market places to be held liable for paying VAT if traders using their platforms do not comply with their VAT obligations. There are two key aspects characterizing the new German rules: the record-keeping obligations for operators of electronic marketplaces (par. 22f of the German UStG) and the liability for traders.

Before going in details with the content of these two elements contained in the new German VAT law, marketplaces will be affected by the new provisions only if they "can be used by third parties to "legally effect" transaction with goods"[952]. According to literature, marketplaces dealing with services are outside the scope of this law and also in the case where the purchase contract is not concluded on the marketplace (e.g. in the case of price comparison websites), new rules shall not find application.[953] Moreover, these new rules make no distinction between German and foreigner marketplaces. Since traders, when applying for the VAT certificate, will have to indicate the name of marketplace and the accounts names used for trading, this implies that the German tax authorities will get to know also about non-established marketplace operators active within Germany.

Regarding the record-keeping obligations, under these new rules, marketplaces' operators must keep records regarding the trader itself and the transactions which it carries out through the marketplace. The record keeping obligation is based on a VAT certificate which the marketplace must archived. This certificate is issued by the competent tax office of the trader upon request of the trader itself and which certifies the correctness of the VAT registration. In addition to the certificate, from January 2019, the marketplace must also keep information such as:

a) The full name and address of the trader;
b) The traders' tax number which has been issued by the competent German tax office and if available the German VAT ID number which has been issued by the central tax office;
c) The start and end dates of the validity of the VAT certificate;
d) In relation to the relevant sales (sales for which either the place of the beginning of transport/dispatch or the place of destination is in Germany) made by the trader through the marketplace, information regarding the place of the beginning of the transport or dispatch, the place of destination, the time of supply and the amount of sales.

Moreover records must be kept even for sales which will not be subject to German VAT at all, such in the cases where the trader is a private individual, the trader benefits from the small entrepreneur's scheme (for Germany the scheme is applicable to German residents which have a turnover under the registration threshold of EUR 17.500), or when the place of supply is outside Germany, when the sale is zero rated. Thus, even if the marketplace targets only private individuals as traders, the operator even if it will not be affected by the liability rule will still have to comply with the book keeping obligations. Finally, with regard to book keeping obligations, the non-compliance is an offence against the German General Tax Code which is penalized.

[950] Ibid.
[951] Ibid.
[952] Hans-Martin Grambeck, Electronic Marketplace may be held liable for German VAT – New Rules Entered into Effect on 1 January 2019, (2019) 30 International VAT Monitor 1, 9.
[953] Ibid.

The German VAT liability[954] for the marketplace arises for the VAT which would have been declared and remitted to the tax office by the traders. Nevertheless, there are three circumstances where the marketplace might not be held liable:

1. When the market operator is able to present a valid VAT certificate for the trader and it did not know (or could not have known) that the trader was not complying with German VAT law;
2. In cases where the trader is registered as a private individual and there are no indicators (e.g. business volume, number and regularity of trades) which can be an alarm for the non-compliant behaviour of the trader;
3. When the non-compliant trader has stopped trading on the marketplace after the tax office in charge of the trader has contacted the marketplace operator and required him to take action within an appropriate period.[955]

Regarding possible limits of the German new legislation, scholars have pointed out that the effectiveness of book keeping obligations could be challenged in cases where the marketplace is not involved in the logistics (an example is the Fulfilment by Amazon (FMA) Programme).[956] Indeed, in this case it will be hard for the marketplace providers to be aware of the place of dispatch and destination of the sold goods and the time of supply). Thirdly, it seems hard for German tax authorities to enforce these rules to marketplaces operators from outside Germany or the European area (which might also not be aware of these new rules).

5.3 Australia

Outside Europe, similar rules on platforms operators' liability have already entered into force in Australia. In 2017, the Australian government passed a new set of rules on goods and services tax (GST) aiming to level the playing field between Australian goods and imported products. Consequently, starting from 1 July 2018, Australian GST have been applied to sales of low value goods imported and sold to Australian consumers. According to the Australian legislation, the low value goods are the ones below the AUD 1.000 threshold.[957] Before the entrance into force of the new rules, imported goods below this threshold were exempt, from 1 July onwards they became subject to a 10 % GST.[958] Regarding the new liability regime of platforms operators, according to the new rules, the operator of an electronic distribution platform (EDP) will be treated as the supplier of low-value goods if goods are purchased through the platform by Australian based consumers and imported with the assistance of either the supplier or the operator.[959] Consequently, platforms operators will be required to register and return GST on the sales which took place via the platform in cases where the GST turnover has been above the AUD 75.000 registration threshold in a 12-months period.

However, the new regime provides for three exceptions where the obligation to charge GST will not apply:

1. It is a B2B transaction and the recipient business has an Australian business number (ABN) and is GST registered;

[954] This liability regime will enter into force after an interim period which varies depending on whether the trader is established within or outside the European Union or the European Economic Area. In the first case, the liability regime will apply as from 1 October 2019 while in the second case from 1 March 2019. This interim period aimes at giving the possibility (in a reasonable time) to the traders to apply for the VAT Certificate at their competent tax office. Hans-Martin Grambeck, Op. cit., 8.
[955] As indicated by Grambeck, the law does not provide for an indication of the relevant time in which actions shall be taken by the marketplace operator. Hans-Martin Grambeck, Op. cit., 8.
[956] Ibid.
[957] Aleksandra Balm, Managing EU VAT Risks for Platform Business Models, Op. cit;
[958] ATO, GST on low value imported goods, https://www.ato.gov.au/General/New-legislation/In-detail/Indirect-taxes/GST/GST-on-low-value-imported-goods
[959] Ibid.

2. There is a multitude of goods with a combined customs value of more than AUD 1.000 that are going to be shipped in Australia;
3. The goods sold through the platforms are exempt.[960]

Moreover, the Treasury Amendment (GST Low Value Goods) Act 2017 has introduced the treatment of re-deliverers as the suppliers of low-value goods if the goods are delivered outside of Australia as part of the supply and the re-deliverer assisted with their delivery into Australia as part of a shopping or mailbox service that it provides under an arrangement with the consumer. Finally, according to the new legislation non-resident suppliers of low-value goods that are connected with Australia can elect to access the simplified registration and reporting system.

5.4 The VAT liability regime for platforms as contained in the new EU VAT e-commerce package

As mentioned in the introduction, in 2016, a European Commission study showed how at European level, most of the distance sales of goods, both supplied from one Member State to another and from third territories or third countries to Member States are facilitated through the use of an electronic interface and often resorting to fulfilment warehousing arrangements.[961] Electronic interface can operate in the form of a marketplace, platform, portal or similar means. Despite the fact that Member States may provide that a person other than the person liable for the payment of VAT is to be held jointly and severally liable for payment of VAT, according to the European legislator this approach insufficient to ensure effective and efficient collection of VAT.[962] For this reason and in order reduce the administrative burden for vendors, tax administrations and consumers, new provisions deeming that taxable persons who facilitate distance sales of goods through the use of electronic interfaces in the collection of VAT on those sales by providing that they are the persons who make those sales, have been adopted.

Going into details with this new liability regime, this is restricted to sales of goods imported from third territories or third countries to the EU which are dispatched or transported in consignments of an intrinsic value not exceeding EUR 150 (nevertheless, a full customs declaration upon importation is required for customs purposes). On this point, the new Article 14a(1) of the VAT Directive establishes that in cases where platforms facilitate B2C imports of goods into the EU and the value of those goods is below EUR 150, platforms will be deemed to receive the supply from the original seller. Therefore, this transaction will be considered as a B2B supply and the transaction between the platform and the customer as a B2C supply. Similarly, the new Article 14a(2) provides the same tax treatment to platforms facilitating intra-EU sales of goods made by non-EU businesses. This provision is applicable to both domestic sales and sales involving transport from another Member State. Moreover, with the new Directive proposal presented by the EU Commission in December 2018, the new version of Article 14a of the VAT Directive splits a single supply into two supplies. Due to the necessity to determine the supply to which the dispatch or transport of the goods should be linked to and in order to identify the correct place of supply, the European Commission suggests, in its new proposal, to ascribe the dispatch or transport of goods to the B2C supply from the online platform to the customer. Consequently, the B2B supply from the seller of the goods to the online platform should be zero rated.[963]

[960] Aleksandra Balm, Managing EU VAT Risks for Platform Business Models, Op. cit.
[961] Council Directive (EU) 2017/2455 of 5 December 2017 amending Directive 2006/112/EC and Directive 2009/132/EC as regards certain value added tax obligations for supplies of services and distance sales of goods, Recital n. 7; European Commission, Impact Assessment Accompanying the Document Proposals for a Council Directive, Op. cit.
[962] Council Directive (EU) 2017/2455 of 5 December 2017 amending Directive 2006/112/EC and Directive 2009/132/EC as regards certain value added tax obligations for supplies of services and distance sales of goods, Recital n. 7.
[963] A. Bal, 'Germany: New VAT Compliance Obligations for Online Platforms', 28 EC Tax Review 2 (2019), p. 115; Pratica Fiscale e Professionale, 21 / 2019, p. 77; Francesco D'Alfonso, Regime One Stop Shop (OSS) per le vendite a distanza di beni intra-UE, (2019) 27 Pratica Fiscale e Professionale, 33; Corte dei Conti, L'E-commerce e il sistema fiscale, Deliberazione 24 maggio 2018, n. 8/2018/G, 113-116;

Regarding the wording of the new Directive, Art. 14 a defines the platforms as the one "facilitating" distance sales of goods. The definition of the term "*facilitates*" is now contained in Art. 5b of the proposal for the new implementing regulation[964] establishing that the word facilitate shall be understood as allowing the entrance into contact between customers and suppliers which will result in a supply of goods to the customer through the electronic interface. This new explicative provision tries to answer some of the issues related to the VAT e-commerce package scholars had previously highlighted.[965] More specifically, without the new definition of Art. 5b, the word "facilitate" would be much broader than the definition provided in the old implementing regulation where the deeming provision was applied only to platforms "taking part" in telecommunication, broadcasting and electronic services. Thus, it has been argued that all sort of platforms facilitating sales in one way or another would have fallen under this term even though some platforms do not have enough control on the supplies taking place through them (e.g. payment, conditions etc.).[966]

In establishing the scope of the deeming provision, the already adopted art. 14a (2) also lists the conditions under which a taxable person shall not facilitate a supply of goods, namely and either directly or indirectly:

(a) when the platform does not set the general terms under which the supply of goods is made;
(b) when the platform is not involved in charging the customer in respect of the payment made;
(c) when the platform is not involved in the ordering or delivery of the goods.

The scope is even more limited by the introduction in Art. 5b of the proposal for a new implementing regulation, of a list of cases where Art. 14a of Directive 2006/112/EC shall not find application where a taxable person only provides:
(a) the processing of payments in relation to the supply of goods;
(b) the listing or advertising of goods;
(c) the redirecting or transferring of customers to other electronic interfaces where goods are offered for sale, without any further intervention in the supply.

In this way, platforms which do not have enough control on the suppliers to effectively collect VAT shall not be VAT liable.

Furthermore, Art.5c (2) of the proposal for the implementing regulation states that for the application of Article 14a of the Directive it is presumed that the person selling goods through an electronic interface is a taxable person and that the person buying those goods is a non-taxable person. Nevertheless, the taxable person deemed to have received and supplied the goods himself can still rebut the presumptions referred to in the first subparagraph if he has information demonstrating the contrary. According to Bal, this provision was adopted in order to reduce the administrative burden on online platforms, since in this way, they don't have to fulfill the obligation of having to prove the status of the supplier and customer.[967]

With specific regard to the liability of the platform for the sales of goods imported from third countries, the relevant rule is contained in the first comma of Art.14a. In particular, this provision establishes that in the cases of distance sales of goods imported from third territories or third countries in consignments of an intrinsic value not exceeding EUR 150, the taxable person facilitating those transactions through the platform shall also be deemed to have received and supplied those goods himself. In this context, another important article is Art. 369n of the Directive,

[964] Aleksandra Bal, 'Germany: New VAT Compliance Obligations for Online Platforms', Op. Cit. 115.
[965] Marie Lamensch, 'Rendering Platforms Liable to Collect and Pay VAT on B2C Imports: A Silver Bullet?', Op. cit. 48-49.
[966] Ibid.
[967] Aleksandra Bal, 'Germany: New VAT Compliance Obligations for Online Platforms', Op. Cit. 115.

stating that supply of goods imported from third territories or third countries under this special scheme will become chargeable when the payment has been accepted. Art. 61b clarifies that the payment shall be considered accepted upon the receipt of a payment confirmation by or on behalf of the taxable person making use of the import scheme or by the supplier selling goods through the electronic interface, regardless of when the actual payment of money is made. The payment confirmation can be in the form of an authorization message or a commitment for payment from the customer. Regarding the chargeable event, the new Art. 66a of the Directive[968] establishes that supply of goods under the scope of Art. 14a will become chargeable at the time of the payment acceptance, which according to Art. 66a the implementing regulation[969] is the time when the payment is confirmed through an authorization message received or on behalf of the supplier selling goods through the electronic interface regardless of when the actual payment of money is made.

Scholars have highlighted some of the difficulties which arise from this new system and they mainly regard the One Stop Shop (OSS) system itself[970]. With the new Directive in fact, the EU decided to bring the distance sales under the old Mini One Stop Shop (MOSS) scheme which now goes under the name of OSS. By extending the MOSS, which was introduced for telecommunication, broadcasting and electronic services, the supplier of goods in a cross-border B2C relationship will, as from 2021, no longer require multiple VAT identification numbers for selling goods online to customers located in different countries. The supplier will now charge local VAT in the Member State where the customer is located and the invoicing rules of the Member State where the supplier is established will apply. Indeed, even if the OSS registration is optional, the expectation from this simplified scheme is that the platforms will register. Consequently, the imports of good made by them will be exempt and the VAT will be paid on a monthly basis through the OSS. Customs authorities will have to verify the validity of the OSS registration number as indicated in the import declaration and after this assessment the good will be delivered to its customer. This means that anyone which have the OSS number of platforms like Amazon or eBay can exempt import of goods but then the same platforms will be expected to declare and pay the corresponding VAT. These OSS registration numbers will be available to suppliers when selling through the platforms, but issues might arise when the suppliers use that number for selling directly to private customers, for which sales platforms shall not be held liable.
The introduction of Art. 5c (1) of the proposal for the new implementing regulation offers indications of when a platform shall not be held liable for the payment of VAT in excess of the VAT which it has declared and paid on the supply when the following conditions are met:
(a) the taxable person is dependent on information provided by suppliers selling goods through an electronic interface or by other third parties in order to correctly declare and pay the VAT on those supplies;
(b) the information received by the taxable person is erroneous;
(c) the taxable person can demonstrate that he did not and could not reasonably know that this information was incorrect.
However, it would be reasonable, in light of the above described issue, to read this provision in the sense that the platform shall not be held liable also in the case where it can demonstrate that the supplier used the platform's OSS registration number in cases where the sales took place off the platforms. Nevertheless, how the platform can effectively verify these types of transactions occurring off the platform but with its OSS registration number keep being an open question.[971]

[968] Council Directive (EU) 2017/2455 of 5 December 2017 amending Directive 2006/112/EC and Directive 2009/132/EC as regards certain value added tax obligations for supplies of services and distance sales of goods.
[969] Art.41a, Proposal for a Council Implementing Regulation amending Implementing Regulation (EU) No 282/2011 as regards supplies of goods or services facilitated by electronic interfaces and the special schemes for taxable persons supplying services to non-taxable persons, making distance sales of goods and certain domestic supplies of goods, 11.12.2018, COM(2018) 821 final.
[970] Marie Lamensch, 'Rendering Platforms Liable to Collect and Pay VAT on B2C Imports: A Silver Bullet?', Op. cit. 48-49.
[971] As also underlined by Lamensch, Op. cit., 48-49

Moreover, another issue which is still not solved regard the monthly listings of imports of goods under the OSS scheme which Member States need to prepare. The introduction of these listings which also contains value declarations is to compare import figures with the data of OSS returns. However, it has been already pointed out that this goals can be fulfilled only if data are correct since it would be possible for a third country supplier to import a goods and declaring a EUR 5 value, remitting VAT rated to the sale of the value of EUR 5 via OSS, thus the data would be matching, but the real value might be different.[972] Moreover, there is no specific provision addressing returned goods which do not leave the EU and are just sent to outlet centers or destroyed there. Consequently, the effectiveness of these listings which represent an extra burden for States is questionable.[973]

Finally, Art. 242a of the Directive[974] provides for the obligation of keeping of records for a period of at least 10 years in respect of supplies by taxable persons facilitated by an electronic interface. This period has been identified as the necessary one in order to assist Member States in the assessment of whether VAT has been accounted for correctly on those supplies. Account should be taken of what information is available to such taxable persons, is relevant to tax administrations and is proportionate to the purpose of the provision, as well as of the need to comply with the General Data Protection Regulation (EU) 2016/679 (see Article 1, points (4), adding a new Section 1a and Article 54c to the Regulation).

5.5. Benefits and limits of new VAT/GST liability regimes for platforms

As it emerges from the previous sections, the new provisions concerning VAT/GST on distant sales of goods adopted in UK, Germany, Australia and contained in the new VAT e-commerce package at European level have in common a new liability regime for platforms. These new provisions take advantage for VAT/GST enforcement purposes of the position held by online marketplaces where distant sales take place.

The choice to involve a third party in the enforcement of tax provisions is not a new idea in the broader taxation system. Indeed, there are many examples where third parties acting as intermediaries are included in the assessment and/or collection processes (e.g. the employers, financial intermediaries, notaries. Also in respond to studies concerning the so-called VAT gap[975], there has been a constant trend recently in strengthening and improving VAT/GST enforcement[976], both online and offline. However, considering the use of platforms for online distance sales, enforcing VAT/GST rules by holding platforms liable instead of many different suppliers can even be more costly-efficient for VAT/GST enforcement purposes than enforcing those rules in the offline world.

Nevertheless, while these new provisions represent a new opportunity for tax administration to efficiently collect VAT/GST, they also increase the bureaucratic burdens on platforms. Indeed, in order to be compliant, they will have to carefully check every type of transactions taking place through their online interface, every supplier taxable situation and will have to collect the relevant information which they might need to use in the future in order to demonstrate they did not know or could not reasonably know that the information provided by the supplier was incorrect.

While the possible introduction of ICT tools might facilitate the compliance with the increasing bureaucratic burdens, unless these tools are provided by the same tax agencies, this will not

[972] Ibid.
[973] Ibid.
[974] Council Directive (EU) 2017/2455 of 5 December 2017 amending Directive 2006/112/EC and Directive 2009/132/EC as regards certain value added tax obligations for supplies of services and distance sales of goods.
[975] The latest study on the VAT gap dates 4 September 2019. European Commission, Study and Reports on the VAT Gap in the EU-28 Member States: 2019 Final Report, (2019) TAXUD/2015/CC/131.
[976] One example is the wider use of the e-receipt and e-invoices.

alleviate the compliance costs for businesses. Consequently, this new regime could act as a deterrent for new businesses willing to enter this market.

Finally, the extension of the scope of the special schemes to cover also distance sales of goods and all services, will considerably increase the number of transactions to be reported in the VAT/GST return. Moreover, the extension of such schemes also to distance sales of goods imported from third territories or third countries, will require the customs authority of the Member State of importation to identify imports of goods in small consignments for which VAT/GST is to be paid through one of the special schemes. Thus, for this new mechanism to be successful (especially from a European single market perspective) there is certainly the need for stronger cooperation between customs and tax authorities of the different member States.

6. Conclusions

In the last years, the massive development of e-commerce has revealed both its opportunities and its risks. In the taxation field, issues do not only arise with reference to corporate income tax strategies adopted by e-commerce platforms to lower their tax burdens, but they also arise in the field of VAT/GST. The risks that arise in this context are many and plenty, such as the difficulties in determining whether there is a taxable person or in qualifying a transaction as C2C or B2B or in assessing whether taxpayers are sufficiently aware of tax consequences of their online activities. Moreover, in the context of imports of low-value goods, third-countries suppliers might misuse currently enforceable provisions.

Effective enforcement of VAT/GST need to be ensured for granting equal treatment to the concerned market operators and because possible VAT/GST evasion and fraud can lead to revenue losses. In this way, public revenues which shall be used for welfare policies and to safeguard social justice values are jeopardized. This has led domestic legislators, the European Union and the OECD to reflect on how effective enforcement could take place online since the old VAT/GST enforcement procedure were not introduced or designed for the virtual world in the first place. These risks have drawn the attention of legislators and regulators at single jurisdiction, European and International level. In March 2019, the OECD has released its framework of guidelines which will help States in designing the necessary policies to involve platforms in VAT/GST collection.[977] In particular, this report shows how platforms can be held liable for VAT/GST collection, can be obliged to share information due to their advantageous position, can be used as educator for the businesses operating through them, can be partners in co-operative agreements and finally, can voluntary operate as an intermediate.

Thus, the welcomed solution was a higher level of inclusion of online platforms in the enforcement mechanism. However, targeting a third party for tax assessment and/or collection purposes is not a new mechanism in the taxation field. Indeed, the impossibility for tax authorities to assess the tax situation of every single taxpayers has led to relying on third parties as tax collectors or information providers.[978] Examples are employers, financial institutions, notaries.

However, looking at the different new legislations which have been adopted in the UK, Germany, Australia and in the new EU VAT Package, it is possible to see the main benefits and limits of these new provisions.

Indeed, platforms have an advantageous position since they have access to relevant information about the type of transactions occurring through them. Because of this characterizing element, it is possible to collect VAT/GST through them more effectively than by assessing and collecting

[977] OECD, The Role of Platforms, Op. cit.
[978] Michael Doran, Op. cit. 143; Leandra Lederman, Op. cit. 695; Tina Ehrke-Rabel,Op. cit.; Tina Ehrke-Rabel, 'Big data in tax collection and enforcement', in Werner Haslehner and others (eds.) *Tax and the Digital Economy: Challenges and Proposals for Reform* (Kluver Law International, 2019).

VAT/GST on a single supplier basis. Nonetheless, these new regimes increase the bureaucratic burdens and costs on platform. Even if there are provisions which limit the liability by considering the functioning of the platform and the fact that platforms might be given wrong information and could not reasonably know that the given information was correct, this burden of proof could still act as deterrent for new businesses. Nevertheless, the technical implementation of this provision could be achieved with relevant ICT tools, which might facilitate the compliance.

Finally, notwithstanding the fact if these measures are far from perfect, they still represent a step forward in trying to reach a higher level of VAT/GST compliance, thus avoiding possible VAT/GST evasion and fraud, and at the same time safeguarding public revenues. Undoubtedly, platforms can count on more precise information on the transactions taking place on their online marketplaces and the exploitation of such advantage by tax authorities is perfectly in line with the development of modern new tax system increasingly relying on third parties.

ANNEX:

Best Practices on Platforms' Implementation of the Right to an Effective Remedy

IGF COALITION ON PLATFORM RESPONSIBILITY |OUTCOME DOCUMENT N°3 | 2019

1. Coordinators

Luca Belli and Nicolo Zingales

2. Background

This document represents the collective output of the *ad hoc* working group of the Dynamic Coalition on Platform Responsibility[979] (DCPR) on the implementation in the context of online platforms of the right to an effective remedy, enshrined inter alia in article 2.3 of the International Covenant on Civil and Political Rights, and several regional Human Rights instruments. The interest in elaborating this document emerged as a clear outcome of the 4th annual meeting of the DCPR, held during the 12th Internet Governance Forum, in December 2017. Many session participants expressed interest in advancing the discussion on platform responsibility, pivoted by the 2017 DCPR official outcome book[980] and building on the solid ground laid by the 2015 **DCPR Recommendations on Terms of Service and Human Rights** (hereinafter the "**Recommendations**")[981] which are the 2015 DCPR official outcome.

Based on the expression of interests expressed at the IGF 2017 meeting, DCPR Coordinators shared a call for participation to an ad hoc DCPR Working Group (WG) tasked with the analysis of reviewing the existing mechanisms for alternative dispute resolution offered by a selection of platforms, scrutinising due process requirements, and to identify best practices. WG members provided inputs to form a proposed Template[982] to be used for review of existing dispute resolution mechanisms. At the RightsCon 2018 meeting of the DCPR the composition of the WG was further expanded[983] and it was agreed to open an additional request for comments on the draft Template, to allow all DCPR members, besides the existing WG members, to provide comments for two additional weeks.[984]

The WG members agreed to work towards the identification of best practices, with a view to promoting due process in the context of alternative dispute resolution mechanisms offered by online platforms. The first draft was grounded on the analyses[985] developed by the WG members and was shared on the public DCPR mailing list to collect feedback. A consolidated version was

[979] DCPR is a multistakeholder group of the United Nations Internet Governance Forum, dedicated to the analysis of online platforms. DCPR is commonly referred to as the IGF Coalition on Platform Responsibility.
[980] Specifically, the edited volume "**Platform regulations: how platforms are regulated and how they regulate us".** The book is freely available at http://bibliotecadigital.fgv.br/dspace/handle/10438/19402
[981] The **Recommendations** can be accessed at https://www.intgovforum.org/cms/documents/igf-meeting/igf-2016/830-dcpr-2015-output-document-1/file
[982] To encourage and facilitate the inclusion of inputs and comments from WG and DCPR members, the DCPR Coordinators utilised a shared online document available at https://docs.google.com/document/d/1T-bMKnFBtsDQ_AycHjI-dlzwpAletMBWJilWRyD-4lM/edit#
[983] The list of contributing WG member is the following (in alphabetical order): Christina Angelopoulos; Luca Belli (DCPR Coordinator); Maria Bjarnadottir; Marta Cantero Gamito; Giovanni De Gregorio; Luã Fergus; Rosalie Gillett; Agnieszka Janczuck; Cynthia Khoo; Chiara Poletti; Roxana Radu; Nicolas Suzor; Ilana Ullman; Rolf Weber; Chris Wiersma; Richard Wingfield; Nicolo Zingales (DCPR Coordinator).
[984] The Report of the DCPR meeting at RightsCon 2018 is available at http://www.intgovforum.org/multilingual/index.php?q=filedepot_download/4905/1255
[985] WG members analised the mechanisms described in the Terms of Service (ToS) of the selected platforms. WG members considered the ToS publicly available in July 2018. All analyses performed by the WG members are available at https://docs.google.com/spreadsheets/d/11NJr2dQvTSoHs6ZubtQvbwf4Z-h8o7FaNTzR8Qk3UFI/edit#gid=1224846873

developed and shared with the wider IGF community to collect a broader range of comments, between October and 31 December 2018.[986] This final version of the Best Practices was shared on the DCPR mailing-list to receive final comments, over the month of February 2019, and verify that the text represented a consensus document, before publishing it. No objection was raised. However, we acknowledge that the Best Practices should be considered as a living document that could be updated in the future.

3. Introduction

In accordance with the approach adopted by the **Recommendations**, this document utilises the term "**shall**" when practices correspond to minimum standards for the respect of due process by platform operators (standards that "shall" be met), while it utilises "**should**" to suggest practices which are recommended, or "should" be followed to facilitate the most "responsible" adherence to due process principles in the definition and implementation of alternative dispute resolution mechanisms.

The document is structured in four sections exploring the safeguards (a) prior to the adoption dispute resolution measures; (b) in connection with the adoption dispute resolution measures; (c) relating to dispute resolution mechanism; (d) and relating to the implementation of the remedy. Best practices have been identified by merging together solutions that appear most suitable to protect platform users' rights, at the same time attending to considerations of viability of online platforms' business models. Quotations of the contractual clauses that inspired the practices are included. When best practices were not identifiable, this document has suggested formulations that maximise the protection of user rights while striking a fair balance between stakeholder interests.

This document was based primarily on the analysis of the contractual agreements that Internet users are required to adhere to in order to become platform users. Platform operators typically detail in these agreements, broadly defined as "Terms of Service" (Tos),[987] the rules and mechanisms applicable to alternative dispute resolution mechanisms. Moreover, analysts where asked to verify, to the extent possible, the concrete implementation of those mechanisms by simulating a dispute in the platforms of choice.

A. Safeguards prior to the adoption of dispute resolution measures

1. Platforms should require registration in order for users to actively interact with others and to create content, within the platform. However, they should not impose the use of real name as public user login. While requiring complete and accurate information about users at the moment of registration, platforms shall not oblige users to make that information public.[988]
Furthermore, platforms should not permit registrations with the effect to:

a) Creating public reliance on someone else's name, image, or other personal information, if that is liable to deceive third parties as to a user's identity. No deception arises, however, in case of clearly parodic impersonification of public figures.
b) Misleading third parties as to a user's authority to represent a particular natural or legal person.

User information shall be shared with third parties, including state actors, only when this is justified by a court order.

[986] See https://www.intgovforum.org/multilingual/content/dcpr-best-practices-on-due-process-safeguards-regarding-online-platforms%E2%80%99-implementation-of
[987] These Best Practices utilise the same definition of ToS provided by the Recommendations, thus covering not only contractual agreements available under the traditional heading of "Terms of Service" or "Terms of Use", but also any other platform's policy document (*e.g.* Privacy Policy, Community Guidelines, etc.) that is linked or referred to therein.
[988] See Recommendations, Section I.5

Twitter

If you do choose to create an account, you must provide us with some personal data so that we can provide our services to you. On Twitter this includes a display name (for example, "Twitter Moments"), a username (for example, @TwitterMoments), a password, and an email address or phone number. Your display name and username are always public, but you can use either your real name or a pseudonym.

Linkedin

Members cannot: a) impersonate others on the Services or mislead, confuse, or deceive others. Pretending to be someone else or to be representing a business in a way that is not truthful is not allowed. b) use someone else's name, image, or other personal information to deceive others into thinking you are someone other than the member or associated with a business or organization when the members are not. c) use or attempt to use another individual's LinkedIn account or create a member profile for anyone other than the member (a real person). d) misrepresent their identity or information or mislead, confuse, or deceive others. When choosing a profile picture, members may not use an image that is not their likeness or a head-shot photo for their profile. Also, members may not manipulate identifiers in order to disguise the origin of any message or post transmitted through the Services.

2. In case platforms aim at restricting the type of content deemed as acceptable, their terms of service shall set out detailed rules, clearly explaining what type of content can be considered acceptable.[989] In doing so, platforms shall bear in mind their responsibility to respect human rights, including freedom of expression. Categories of content that could deemed as unacceptable and shall be clearly defined in the terms of service include spam, shocking and pornographic content, content instigating violence or discriminating against individuals based on race, ethnicity, national origin, sex, gender, gender identity, sexual orientation, religious affiliation, disabilities, or diseases, or content deemed as illegal in specific jurisdictions.

LinkedIn (applicable to disputes concerning: Intellectual property infringement; Revenge porn; Fake news; Terrorism-inciting contente; Hate speech; Right to erasure/ right to object to processing/ right to rectify or restrict processing; Defamation; Child pornography)

Honesty and Authenticity [...] You may not use the Services to share false content or information, including news stories, that presents untrue or unverified facts or events as though they are true or acts or events as though they are true or likely true. [...] Adult Content It's not acceptable to post content containing nudity, sexually explicit material, or pornography. Some adult content may be allowed in an educational, medical, scientific, or professional artistic context so long as it is not gratuitously graphic. The Services are never to be used for sexual exploitation of children. You also may not post content that threatens sexual violence or sexual assault. You may not use the Services to engage in or promote escort services, prostitution, or human trafficking. Bullying and Harassment Bullying or harassment that targets individuals or groups to degrade or shame them is not allowed. This includes, but is not limited to, abusive or humiliating language, sexual advances and innuendo, revealing others' personal or sensitive information (aka "doxing") or posting content about them without consent, or inciting or engaging others to do any of the same. Hate, Violence, and Terrorism We do not allow organizations or groups that engage in or promote violence or property damage, organized criminal activity, prejudice, or hate. Also, you may not use our Services to express support for such groups or to post content or otherwise use the Services to incite violence or hatred against particular individuals or

[989] See Recommendations, Section III.1

groups. Content that depicts terrorist activity, that is intended to recruit for terrorist organizations, or promotes or supports terrorism in any manner, is not tolerated on the Services. Harmful Content and Shocking Material You may not post violent or graphic content or otherwise use the Services with the intent to shock or humiliate others. We do not allow activities that promote, organize, depict or facilitate criminal activity. We also do not allow content depicting or promoting instructional weapon making, drug abuse, and threats of theft. Content or activities that promote or encourage suicide or any type of self-injury, including self-mutilation and eating disorders, is also not allowed. Spam Untargeted, irrelevant, unwanted, unsolicited, unauthorized, inappropriately commercial or promotional, or gratuitously repetitive messages and other similar content are considered spam and are not allowed on the Services. You may not use our invitation features to send messages to people who don't know you or who are unlikely to recognize you as a known contact. Please make the effort to create original, professional, relevant, and interesting content in order to gain popularity, instead of trying ways to artificially increase the number of views, re-shares, likes, or comments.

3. As a general rule, platforms should only store personal data for as long as necessary for the purpose(s) for which they were originally collected.[990] This should include retention for a period that is reasonably necessary to comply with legal obligations (e.g. law enforcement requests), meet regulatory requirements and resolve disputes, or to protect the safety or integrity of the platform. Examples of the latter are where storage helps to prevent spam and detect fraud or malicious behaviour aimed at service disruption, or to explain why platform operators removed specific content or accounts from the platform.

Airbnb

Airbnb generally retain personal information "for as long as is necessary for the performance of the contract between you and us and to comply with our legal obligations". Users/members can request the erasure of personal information.

4. Platforms should provide meaningful notice of any changes in their ToS at least 30 days before the changes go into effect.[991] Platforms shall provide users with the opportunity to review the changes before they become effective and changes cannot be retroactive. Notification of changes shall be communicated both via email, where practicable, and through the platform.

WordPress

WordPress uses posts/email/other communication in advance of changes - see "13. Changes." in ToS, including statement that "any dispute that arose before the changes shall be governed by the Terms (including the binding individual arbitration clause) that were in place when the dispute arose." - AND it keeps change logs –

Wikipedia

Wikipedia provides Terms of Use, as well as any substantial future revisions of these Terms of Use, to the community for comment at least thirty (30) days before the end of the comment period. If a future proposed revision is substantial, we will provide an additional 30 days for comments after posting a translation of the proposed revision in multiple languages

[990] See Recommendations, Section I.
[991] See Recommendations, Section II.1

5. Platforms shall offer mechanisms to report behaviours categorised as abusive by the ToS, by flagging contents and or by filing predefined forms. For instance, when prohibited by the platform's ToS, users should be able to flag:

- Spam,
- Content categorised as inappropriate by the terms of service
- Profiles or groups engaging in activities forbidden by the terms of service
- Phishing and or fraud attempt
- Safety concerns

Specific notice-and-counter-notice mechanisms should be established for

- Intellectual property infringements
- Law enforcement requests for account information (routine and emergency)
- Content removal requests, based on ToS infringement
- Reporting of hacked account.

Where relevant, the abovementioned form shall include at least the following elements

- The email address of the claimant
- The description of the violation type
- The username of violating account
- The URL of post
- Any supporting material in attachments

Linkedin

Linkedin provides mechanism to report abusive behaviours by flagging contents or filing forms according to its Community Guidelines and User Agreement. In general, the following contents could be flagged by users: - Spam, inappropriate, and offensive content - Inappropriate profile photos - Inaccurate profiles - Fake profiles - Inappropriate groups - Phishing or suspicious messages - Safety concerns A specific mechanism based on notice and counter notice is established for copyrights contents (https://www.linkedin.com/legal/copyright-policy). Moreover, a member can report also by flagging or filling a form: - trademark infringements (see the "Trademark Infringement Form"). - fake profiles - hacked accounts (see the form "Reporting Your Hacked Account") - scams

Medium

Medium's rules state: How to report a violation If you find a post or account on Medium that violates these rules, please flag it. You can use this form to provide more detail or to report other conduct you believe violates our rules. Additionally, you can send us an email to yourfriends@medium.com. The report form asks for the following details: How can we help you? (drop down menu features: "report a rules violation.") Your email address Description Violation type Medium username of violating account URL of post Attachments Medium also provides information on filing a DMCA notice: How To File a DMCA Notice To submit a notice of claimed copyright infringement, you will need to provide us with the following information: 1. A physical or electronic signature (typing your full name will suffice) of the copyright owner or a person authorized to act on their behalf; 2. Identification of the copyrighted work claimed to have been infringed (e.g., a copy of or link to your original work or clear description of the materials allegedly being infringed upon); 3. Identification of the infringing material and information reasonably sufficient to permit Medium to locate the material on our website or services (e.g., a link to the infringing post); 4. Your contact information, including your address, telephone number, and an email address; 5. A statement that you have a good-faith belief that the use of the material in the manner asserted is not authorized by the copyright owner, its agent, or the law; and 6. A statement that the information in the notification is accurate, and, under

penalty of perjury, that you are authorized to act on behalf of the copyright owner. You can report alleged copyright infringement by emailing the above information to copyright@medium.com. You can also mail a copyright notice to: Designated Copyright Agent A Medium Corporation 760 Market Street, Suite 900 San Francisco, CA 94102

6. Platform users shall have the right to initiate litigation and take part in class actions in their own jurisdiction.[992] Such rights shall always be available in jurisdictions that are targeted by the platform services (e.g. by using local language, currency or country code domain name).

B. Safeguards in connection with the adoption of dispute resolution measures

7. As a general rule, platforms shall notify affected individuals prior to the adoption of any adverse measures, explaining the specific grounds on which such measure is taken.[993] Exceptions to user notification should be narrowly circumscribed and explained in the terms of service.

Twitter

By default, Twitter will attempt to notify the reported account holder(s) of the existence of a legal request pertaining to the account(s) if we are not otherwise prohibited from doing so. Exceptions to user notice may include exigent circumstances, such as emergencies regarding imminent threats to life, child sexual exploitation, or terrorism. Twitter attempts to notify the user(s) about the legal request through a notification in the Twitter app and by sending a message to the email address associated with the account(s), if available. If we are not permitted to notify the user(s) at this step in the process (e.g., because the legal request is accompanied by a non-disclosure order), we may notify the user(s) about the existence of a legal request after Twitter has withheld the reported content or disclosed information associated with the Twitter account(s).

8. Platform should always allow affected individuals to contest a notified measure *before* adoption.[994] Measures should be implemented immediately, on a temporary basis, when this is justified by particular urgency e.g. when content shall be removed before it incites others to do harm, or in case of child abuse imagery.

Medium

If you break the rules If it looks like you've violated our rules, we may send you an email and ask you to explain what you're up to and why. Context is important, and we want to understand the big picture. If you don't adequately explain yourself or fix the problem, we may suspend your account or remove your content. We strive to be fair, but we reserve the right to suspend accounts or remove content, without notice, for any reason, particularly to protect our services, infrastructure, users, or community. If you attempt to evade suspension by creating new accounts, we will suspend your new accounts.

9. Platforms shall always notify affected individuals after the adoption of the measure, explaining the specific grounds based on which the measure was taken.[995]

[992] See Recommendations, Section II.2
[993] See Recommendations, Sections III.1 and III.2
[994] See Recommendations, Section III.2
[995] See Recommendations, Section III.1

Youtube
If a strike is issued, you'll get an email and see an alert in your account's Channel Settings with information about why your content was removed (e.g. for sexual content or violence).

10. Furthermore, platforms shall always allow affected individuals to contest a measure *after* adoption.[996]

Twitter
Violators can appeal permanent suspensions if they believe we made an error. They can do this through the platform interface or by filing a report. Upon appeal, if we find that a suspension is valid, we respond to the appeal with information on the policy that the account has violated."
"File an appeal and we may be able to unsuspend your account. If you are unable to unsuspend your own account using the instructions above and you think that we made a mistake suspending or locking your account, you can appeal. First, log in to the account that is suspended. Then, open a new browser tab and file an appeal.

Instagram
Instagram complies with the notice and takedown procedures defined in section 512(c) of the Digital Millennium Copyright Act ("DMCA"), which applies to content reported and removed for infringing United States copyrights. If your content was removed under the notice and counter-notice procedures of the DMCA, you will receive instructions about the counter-notification process, including how to file a counter-notification, in the warning we send you. When we receive an effective DMCA counter-notification, we promptly forward it to the reporting party. If the reporting party does not notify us that they have filed an action seeking a court order to restrain you from engaging in infringing activity on Instagram related to the material in question within 10-14 business days, we may restore or cease disabling eligible content under the DMCA". "Similarly, if the content was removed based on U.S. trademark rights, and if you believe the content should not have been removed, you will be provided an opportunity to submit an appeal. In these cases, you'll receive further instructions about this process in the notification you receive from Instagram.

11. To ensure the effectiveness of contestation, time limits to contest any measure shall be clearly specified.

Twitter
A time limit is mentioned only in the copyright procedure but not for the contestant, only for the original claimant. "What Happens After I Submit a Counter-notice? Upon receipt of a valid counter-notice, we will promptly forward a copy to the person who filed the original notice. If we do not receive notice within 10 business days that the original reporter is seeking a court order to prevent further infringement of the material at issue, we may replace or cease disabling access to the material that was removed.

C. Safeguards relating to the dispute resolution mechanism

[996] Id.

12. Platforms should have in place a specific mechanism in their websites allowing users to resolve disputes arising between them in relation to their platform activity,[997] besides the mechanisms allowing users to solve disputes between the platform and its users, as specified in paragraph 16.

Airbnb (only concerning claims on security deposits)
The procedure for claims on security deposits proceeds as follows: - Airbnb will ask for documentation from the host, and as soon as it is received, Airbnb will ask the host to contact the guest through Airbnb's Resolution Center to discuss the claim. - When the host sends a request, the guest will be notified by email and through an alert on Airbnb Dashboard. - The guest will have to reply to the host's request in the Resolution Center within 72 hours. The guest's response will depend on whether or not the guest agrees to the amount requested by the host: o Agree to the amount:

- *Click Accept in the Resolution Center. In such case, Airbnb will process the payment and send it to the host (usually within 5 to 7 business days).*
- *Don't agree to the amount: Click Involve Airbnb in the Resolution Center. The guest must provide reasons the invalidity of the host's claim. In such event, Airbnb will contact the guest and provide 72 hours to respond so that Airbnb can mediate.*

The Help Center signals that, in any case, they will make sure both guest and host are represented fairly and gather any details and documentation needed to reach a resolution. It is states that most security deposit claims will be resolved within one week.

13. Platforms should provide detailed and clear explanations to users on the significance of any requests for initiation of disputes that is notified to them, and actions that may be taken in response to those.[998] Platforms should also offer additional assistance, for example by providing a channel for interaction with customer service, or listing contact information of the relevant non-governmental organisations.

Twitter (general guidance)
In case of suspension of account, they describe the procedure to unblock/unsuspend the account and explain the possible reasons (e.g. Your account has been locked for security purposes, Your account is limited because it may have violated the Twitter Rules)

- *"You may be able to unsuspend your own account. If you log in and see prompts that ask you to provide your phone number or confirm your email address, follow the instructions to get your account unsuspended." https://help.twitter.com/forms/general?subtopic=suspended - "Are you seeing a message that your account is locked? Your account may also be temporarily disabled in response to reports of spammy or abusive behavior. For example, you may be prevented from Tweeting from your account for a specific period of time or you may be asked to verify certain information about yourself before proceeding. Get help unlocking your account. File an appeal and we may be able to unsuspend your account. If you are unable to unsuspend your own account using the instructions above and you think that we made a mistake suspending or locking your account, you can appeal. First, log in to the account that is suspended. Then, open a new browser tab and file an appeal. Source (https://help.twitter.com/en/managing-your-account/locked-and-limited-accounts) - Help with locked or limited account We may lock an account or place temporary limitations on certain account features if an account appears to be compromised or in violation of the Twitter Rules or Terms of Service. If you log in or open your app and see a message that your account is locked or that some of your account features have been limited, follow*

[997] See Recommendations, Section II.2
[998] Id.

the instructions to restore it or continue reading for more information. In case of legal requests in the US they offer the contact of two NGOs specialised in freedom of expression (ACLU and EFF). "Unfortunately, we cannot provide you with any legal advice and cannot provide any further information beyond what we provided in our notice. If you wish to seek legal counsel, here are some resources that may help. For U.S. legal requests, you might consider contacting the American Civil Liberties Union (http://www.aclu.org/affiliates, +1 212-549-2500) or the Electronic Frontier Foundation (https://www.eff.org/pages/legal-assistance, info@eff.org, +1 415-436-9333). In other countries For non-U.S. legal requests, you might consider contacting a local attorneys' association or law school, which may be able to provide you with contact information for specialised legal assistance on free expression issues or reduced-cost legal aid services available in your location
Twitter also has a social media account (@Twittersupport) which is the official source for 24/7 Twitter support.

14. Platforms should inform complainants of counternotices and other defenses raised in response to their requests, so as to enable a meaningful contestation.[999]

Linkedin (only for copyright)
Yes, Linkedin has included a note in the counter-notice form which explains the time for the complainant to commence a formal judicial action upon receipt of a copy of the counter-notice (https://www.linkedin.com/help/linkedin/ask/TS-CNRCCI?lang=en§). "Note: Under the Digital Millennium Copyright Act, upon receipt of a copy of this Counter-Notice, the Complainant has 10 business days to commence a formal judicial action against the User in relation to the User's infringing activity. If such action is filed, the allegedly infringing content will be removed or will remain removed from the LinkedIn and/or SlideShare site until the matter is resolved. If no action is filed, we will re-post, or allow you to re-post, the content 10-14 business days after receipt of this Counter-Notice.

15. Platforms that receive requests for content removal shall only implement permanent deletion after an internal (human) review. Users shall always have the possibility to challenge automated deletion and the right to have the deletion decision reviewed by an independent expert or a panel of experts.

Youtube
Reported content is reviewed along the following guidelines: Content that violates our Community Guidelines is removed from YouTube. Content that may not be appropriate for all younger audiences may be age-restricted." However, in its most recent transparency report, YouTube stated that 74.2% of videos are removed before any views thanks to automated flagging.

16. Platforms shall provide an alternative dispute resolution mechanism, designed in a flexible way based on generally accepted procedural rules, for disputes between a user and the platform.[1000]
The rules for such a mechanism should encompass at least the following elements:

- Procedure for the appointment of the adjudicator
- Necessary independence and qualifications of the adjudicators
- Choice between 1 adjudicator and panel of 3 adjudicators
- Procedural principles of such mechanisms should enshrine the right to be heard, equal treatment of parties, access to information, acting in good faith.

[999] See Recommendations, Section II.2
[1000] Id.

Wikimedia
We hope that no serious disagreements arise involving you, but, in the event there is a dispute, we encourage you to seek resolution through the dispute resolution procedures or mechanisms provided by the Projects or Project editions and the Wikimedia Foundation.

Tumblr
The Terms of Service state that: "You and Tumblr agree that we will resolve any claim or controversy at law or equity that arises out of this Agreement or the Services in accordance with this Section or as you and Tumblr otherwise agree in writing. Before resorting to formal dispute resolution, we strongly encourage you to contact us to seek a resolution.

Snapchat (only for businesses)
Not if the user is an individual, Yes if the user is a business. Then the dispute will be settled under LCIA Arbitration Rules. "One arbitrator (to be appointed by the LCIA), the arbitration will take place in London, and the arbitration will be conducted in English. If you do not wish to agree to this clause, you must not use the Services.

17. Platforms should offer alternative dispute resolution mechanisms as an option, but not as an inderogable pre-requisite or substitute for litigation.[1001] Platform users shall always have a meaningful opportunity to opt out from the use of such mechanisms.

Amazon (only for small claims)
Any dispute or claim relating in any way to your use of any Amazon Service, or to any products or services sold or distributed by Amazon or through Amazon.com will be resolved by binding arbitration, rather than in court, except that you may assert claims in small claims court if your claims qualify. The Federal Arbitration Act and federal arbitration law apply to this agreement.

Reddit (informal process, not specified)
Yes. In their User Agreement, par. 13. Governing Law and Venue they specify that " if you have an issue or dispute, you agree to raise it and try to resolve it with us informally. You can contact us with feedback and concerns here or by emailing us at contact@reddit.com.

eBay (opt out available)
Opt-Out Procedure IF YOU ARE A NEW USER OF OUR SERVICES, YOU CAN CHOOSE TO REJECT THIS AGREEMENT TO ARBITRATE ("OPT-OUT") BY MAILING US A WRITTEN OPT-OUT NOTICE ("OPT-OUT NOTICE"). THE OPT-OUT NOTICE MUST BE POSTMARKED NO LATER THAN 30 DAYS AFTER THE DATE YOU ACCEPT THE USER AGREEMENT FOR THE FIRST TIME. YOU MUST MAIL THE OPT-OUT NOTICE TO EBAY INC., ATTN: LITIGATION DEPARTMENT, RE: OPT-OUT NOTICE, 583 WEST EBAY WAY, DRAPER, UT 84020.

Uber (small claims, & equitable relief against possible IP infringement)
However, you and Uber each retain the right to bring an individual action in small claims court and the right to seek injunctive or other equitable relief in a court of competent jurisdiction to prevent the actual or threatened infringement, misappropriation or violation

[1001] Id.

of a party's copyrights, trademarks, trade secrets, patents or other intellectual property rights.

18. Platforms should set a reasonable time limit (e.g. 30 days) for the resolution of any controversy, with the possibility to extend such period upon mutual agreement between the disputing parties. Furthermore, platforms should only set a time limit (e.g., 1 year) to the initiation of claims that have arisen in the past.

Lyft
Before initiating any arbitration or proceeding, you and Lyft may agree to first attempt to negotiate any dispute, claim or controversy between the parties informally for 30 days, unless this time period is mutually extended by you and Lyft.

Tumblr
Time Limitation on Claims and Releases From Liability | You agree that any claim you may have arising out of or related to this Agreement or your relationship with Tumblr must be filed within one year after such claim arose; otherwise, your claim is permanently barred.

Tumblr (for copyright)
The original Notifying Party (or the copyright holder he or she represents) will then have ten (10) days to notify us that he or she has filed legal action relating to the allegedly infringing material. If Tumblr does not receive any such notification within ten (10) days, we may restore the material to the Services.

19. Platforms shall ensure that adjudication of disputes conforms to established standards of independence and impartiality,[1002] for example by reference to rules and procedures adopted by recognised mediation or arbitration associations.

Ebay
The arbitration will be conducted by the American Arbitration Association ("AAA") under its rules and procedures, including the AAA's Consumer Arbitration Rules (as applicable), as modified by this Agreement to Arbitrate. The AAA's rules are available at www.adr.org or by calling the AAA at 1-800-778-7879. The use of the word "arbitrator" in this provision shall not be construed to prohibit more than one arbitrator from presiding over an arbitration; rather, the AAA's rules will govern the number of arbitrators that may preside over an arbitration conducted under this Agreement to Arbitrate.

User Privacy Notice

If you have an unresolved privacy or data use concern that we have not addressed satisfactorily, please contact our U.S.-based third party dispute resolution provider (free of charge) at https://feedback-form.truste.com/watchdog/request. eBay is committed to your privacy. This privacy notice explains our collection, use, disclosure, retention, and protection of your personal information.

Amazon
The arbitration will be conducted by the American Arbitration Association ("AAA") under its rules, including the AAA's Supplementary Procedures for Consumer-Related Disputes. The AAA's rules are available at www.adr.org or by calling 1-800-778-7879. Payment of

[1002] See Recommendations, Section II

all filing, administration and arbitrator fees will be governed by the AAA's rules. Amazon will reimburse those fees for claims totaling less than $10,000 unless the arbitrator determines the claims are frivolous. Likewise, Amazon will not seek attorneys' fees and costs in arbitration unless the arbitrator determines the claims are frivolous. You may choose to have the arbitration conducted by telephone, based on written submissions, or in person in the county where you live or at another mutually agreed location.

20. Platforms shall provide sufficient reasons to appreciate the rationale of the decision taken by the appointed adjudicator, and should provide an updated list of factors elucidating the application of their terms of service (i.e., their implementation criteria).[1003]

Twitter

Our enforcement philosophy

We empower people to understand different sides of an issue and encourage dissenting opinions and viewpoints to be discussed openly. This approach allows many forms of speech to exist on our platform and, in particular, promotes counterspeech: speech that presents facts to correct misstatements or misperceptions, points out hypocrisy or contradictions, warns of offline or online consequences, denounces hateful or dangerous speech, or helps change minds and disarm.

Thus, ***context matters****. When determining whether to take enforcement action, we may consider a number of factors, including (but not limited to) whether:*

- *The behavior is directed at an individual, group, or protected category of people;*
- *The report has been filed by the target of the abuse or a bystander;*
- *The user has a history of violating our policies;*
- *The severity of the violation;*
- *The content may be a topic of legitimate public interest.*

Is the behavior directed at an individual or group of people?

To strike a balance between allowing different opinions to be expressed on the platform, and protecting our users, we enforce policies when someone reports abusive behavior that targets a specific person or group of people. This targeting can happen in a number of ways (for example, @mentions, tagging a photo, mentioning them by name, and more).

Has the report been filed by the target of the potential abuse or a bystander? *Some Tweets may seem to be abusive when viewed in isolation, but may not be when viewed in the context of a larger conversation or historical relationship between people on the platform. For example, friendly banter between friends could appear offensive to bystanders, and certain remarks that are acceptable in one culture or country may not be acceptable in another. To help prevent our teams from making a mistake and removing consensual interactions, in certain scenarios we require a report from the actual target (or their authorized representative) prior to taking any enforcement action.*

Does the user have a history of violating our policies?

We start from a position of assuming that people do not intend to violate our Rules. Unless a violation is so egregious that we must immediately suspend an account, we first try to

[1003] Id.

educate people about our Rules and give them a chance to correct their behavior. We show the violator the offending Tweet(s), explain which Rule was broken, and require them to delete the content before they can Tweet again. If someone repeatedly violates our Rules then our enforcement actions become stronger. This includes requiring violators to delete the Tweet(s) and taking additional actions like verifying account ownership and/or temporarily limiting their ability to Tweet for a set period of time. If someone continues to violate Rules beyond that point then their account may be permanently suspended.

What is the severity of the violation?

Certain types of behavior may pose serious safety and security risks and/or result in physical, emotional, and financial hardship for the people involved. These egregious violations of the Twitter Rules — such as posting violent threats, non-consensual intimate media, or content that sexually exploits children — result in the immediate and permanent suspension of an account. Other violations could lead to a range of different steps, like requiring someone to delete the offending Tweet(s) and/or temporarily limiting their ability to post new Tweet(s).

Is the behavior newsworthy and in the legitimate public interest?

Twitter moves at the speed of public consciousness and people come to the service to stay informed about what matters. Exposure to different viewpoints can help people learn from one another, become more tolerant, and make decisions about the type of society we want to live in.

To help ensure people have an opportunity to see every side of an issue, there may be the rare occasion when we allow controversial content or behavior which may otherwise violate our Rules to remain on our service because we believe there is a legitimate public interest in its availability. Each situation is evaluated on a case by case basis and ultimately decided upon by a cross-functional team.

Some of the factors that help inform our decision-making about content are the impact it may have on the public, the source of the content, and the availability of alternative coverage of an event.

Public impact of the content: A topic of legitimate public interest is different from a topic in which the public may be curious. We will consider what the impact is to citizens if they do ***not*** *know about this content. If the Tweet does have the potential to impact the lives of large numbers of people, the running of a country, and/or it speaks to an important societal issue then we may allow the the content to remain on the service. Likewise, if the impact on the public is minimal we will most likely remove content in violation of our policies.*

Source of the content: Some people, groups, organizations and the content they post on Twitter may be considered a topic of legitimate public interest by virtue of their being in the public consciousness. This does not mean that their Tweets will always remain on the service. Rather, we will consider if there is a legitimate public interest for a particular Tweet to remain up so it can be openly discussed.

Availability of coverage: Everyday people play a crucial role in providing firsthand accounts of what's happening in the world, counterpoints to establishment views, and, in some cases, exposing the abuse of power by someone in a position of authority. As a situation unfolds, removing access to certain information could inadvertently hide context and/or prevent people from seeing every side of the issue. Thus, before actioning a potentially violating Tweet, we will take into account the role it plays in showing the larger story and whether that content can be found elsewhere.

D. Safeguards relating to the implementation of the remedy

21. Platforms should clarify both in their ToS and in the implementation of their practices the territorial scope of any remedy that can be sought or imposed.

Twitter (global remedy unless it is a request by a government or third party in which case it is local)

If content violates their ToS they remove the content from the platform (globally) otherwise if content are removed on the basis of legal requests they remove it only on the country. For content removal requests, this may mean the reported content violates Twitter's Terms of Service or Rules, and the content will be removed from the Twitter platform. Or, perhaps the content is determined to be illegal in a particular jurisdiction and Twitter will withhold access to the identified content in the location in which it is alleged to be in violation of local law. For information requests, Twitter may file or serve objections for requests that are legally defective, overly broad, and/or appear to impermissibly burden free expression. Twitter also checks whether the user(s) filed any objections with the appropriate court. For valid and properly scoped information requests where there has not been a successful objection by Twitter or the user(s), a Twitter agent will assemble the required account records and produce them electronically through our secure LRS site to the requester. Once the records have been produced, the case is considered completed and closed unless we're able to provide delayed notice to affected users after the expiration of an associated non-disclosure order. Source: https://help.twitter.com/en/rules-and-policies/twitter-legal-faqs

22. Platforms should offer the possibility to request the adoption of temporary measures prior to resolution of a dispute. The provisions of set out in paragraph 16 apply for such procedures mutatis mutandis.

Wordpress (only useful answer, based on the analyst's personal experience)
Occasionally, WordPress responds to reports by suspending a blog(-post);

23. Platforms should give users the opportunity to request a review of any implemented measures.[1004] This includes the right to appeal against the assessment of the factual context in which a decision was taken and its consistency with the factors laid out in the platform's ToS (i.e. the enforcement philosophy referred to in C9). Platforms should also provide the possibility to request a review to account for supervened circumstances, as well as representative examples of the types of circumstances (e.g. court decisions) that qualify for the granting of such requests.

Twitter

[1004] See Recommendations, Section II.2

If content that was withheld in response to a legal request becomes allowed in the future, where we can, we will restore access to it so anyone in the world can view it. Some circumstances in which we have un-withheld content in the past include: An objection filed by Twitter against a court order deeming certain content was illegal was accepted by a higher court. An objection filed by a user against a court order deeming certain content was illegal was accepted by a higher court. The validity period of a court order prohibiting publication of certain material expired. An official judicial body expressed an opinion that a request made by an administrative authority was invalid*.

Airbnb (yes, about the facts- but do not allow to challenge their interpretation of standards and expectations)

Following Airbnb's Standards and expectations (https://www.airbnb.com/help/article/1199/what-are-airbnb-s-standards-and-expectations): enforcement teams are made up of dedicated professionals, "but they're still human". Therefore, they acknowledge potential incorrect decisions. ("So, in rare cases, enforcement decisions may be incorrect"). In the event of disagreement with a decision, users are invited to contact Airbnb directly, and then the platform commits to "re-review the decision carefully". However, as it is specified, the definitions of the standards and expectations themselves aren't subject to review.

24. Platforms should have flexible rules allowing for different types of arrangements regarding the allocation of costs in relation to the implementation of a remedy. These rules may include an indication of the amount of claim below which a platform will reimburse users for filing, administration, and arbitrator fee; and should include penalties in case a claim is established to be frivolous.

eBay

Costs of Arbitration Payment of all filing, administration and arbitrator fees will be governed by the AAA's rules, unless otherwise stated in this Agreement to Arbitrate. If the value of the relief sought is $10,000 or less, at your request, eBay will pay all filing, administration, and arbitrator fees associated with the arbitration. Any request for payment of fees by eBay should be submitted by mail to the AAA along with your Demand for Arbitration and eBay will make arrangements to pay all necessary fees directly to the AAA. If (a) you willfully fail to comply with the Notice of Dispute requirement discussed above, or (b) in the event the arbitrator determines the claim(s) you assert in the arbitration to be frivolous, you agree to reimburse eBay for all fees associated with the arbitration paid by eBay on your behalf that you otherwise would be obligated to pay under the AAA's rules.

Amazon

Payment of all filing, administration and arbitrator fees will be governed by the AAA's rules. We will reimburse those fees for claims totaling less than $10,000 unless the arbitrator determines the claims are frivolous.

Lyft

Lyft attributes costs in the event of passenger cancellations (by charging a fee, which the driver receives). Drivers are not charged a fee for cancelling on passengers, but are penalized on performance or ratings: Passengers: https://help.lyft.com/hc/en-ca/articles/115012922687-Cancellation-policy-for-passengers "Cancel fees | You may be charged a fee for cancelling a ride when both of the following occur: - 2 minutes or more

pass since a driver accepts your ride request - Your driver is on time to arrive within 5 minutes of the original estimated arrival time In most cities, you'll be charged $10 for cancelling a scheduled ride." "No-show fee | No-show fees are charged under these circumstances: 1. Your driver arrived to pick you up 2. Your driver waited 5 minutes or more 3. Your driver tried to contact you" Drivers: https://help.lyft.com/hc/en-ca/articles/115012922847 "Cancellation and no-show fee policy for drivers | As consideration for your time and effort, drivers receive cancellation and no-show fees. Fees are based on your region and ride type, so use our cities page to see specific amounts." Damage Fee is attributed to passengers: "Damage Fee. If a Driver reports that you have materially damaged the Driver's vehicle, you agree to pay a "Damage Fee" of up to $250 depending on the extent of the damage (as determined by Lyft in its sole discretion), towards vehicle repair or cleaning." https://www.lyft.com/terms In the event of a dispute going to arbitration, Lyft will compensate users for all but $50 of filing fee, unless claim is for $5000 or more (section 17(e)), if user initiates, or compensate entirety of filing and arbitration fees if Lyft initiates: https://www.lyft.com/terms. Lyft also agrees not to seek attorneys' fees and non-filing expenses if it wins in arbitration (section 17e(6)), but will also not pay user's legal fees in any event.

25. Platforms should set out rules mentioning the possible consequences of repeated infringement of ToS, specifying any significant variations in those consequences depending on the type of violation. They should also make clear that such consequences may only arise in case of established, rather than merely asserted, violations.

YouTube (for copyright)
If you receive more than one strike in the same three-month period, here's what happens: Second strike: If your account receives two Community Guidelines strikes within a three-month period, you won't be able to post new content to YouTube for two weeks. If there are no further issues, full privileges will be restored automatically after the two-week period. Each strike will remain on your account and expire three months after it was issued. Each strike expires separately. Third strike: If your account receives three Community Guidelines strikes within a three-month period, your account will be terminated.

Wikimedia
There are detailed policies relating to blocking users from editing content, and banning users from the platform.

In an unusual case, the need may arise, or the community may ask us, to address an especially problematic user because of significant Project disturbance or dangerous behavior. In such cases, we reserve the right, but do not have the obligation to:

- *Investigate your use of the service (a) to determine whether a violation of these Terms of Use, Project edition policy, or other applicable law or policy has occurred, or (b) to comply with any applicable law, legal process, or appropriate governmental request;*
- *Detect, prevent, or otherwise address fraud, security, or technical issues or respond to user support requests;*
- *Refuse, disable, or restrict access to the contribution of any user who violates these Terms of Use;*
- *Ban a user from editing or contributing or block a user's account or access for actions violating these Terms of Use, including repeat copyright infringement;*
- *Take legal action against users who violate these Terms of Use (including reports to law enforcement authorities); and*

- *Manage otherwise the Project websites in a manner designed to facilitate their proper functioning and protect the rights, property, and safety of ourselves and our users, licensors, partners, and the public.*

In the interests of our users and the Projects, in the extreme circumstance that any individual has had his or her account or access blocked under this provision, he or she is prohibited from creating or using another account on or seeking access to the same Project, unless we provide explicit permission. Without limiting the authority of the community, the Wikimedia Foundation itself will not ban a user from editing or contributing or block a user's account or access solely because of good faith criticism that does not result in actions otherwise violating these Terms of Use or community policies.

The Wikimedia community and its members may also take action when so allowed by the community or Foundation policies applicable to the specific Project edition, including but not limited to warning, investigating, blocking, or banning users who violate those policies. You agree to comply with the final decisions of dispute resolution bodies that are established by the community for the specific Project editions (such as arbitration committees); these decisions may include sanctions as set out by the policy of the specific Project edition.

Especially problematic users who have had accounts or access blocked on multiple Project editions may be subject to a ban from all of the Project editions, in accordance with the Global Ban Policy. In contrast to Board resolutions or these Terms of Use, policies established by the community, which may cover a single Project edition or multiple Projects editions (like the Global Ban Policy), may be modified by the relevant community according to its own procedures.

The blocking of an account or access or the banning of a user under this provision shall be in accordance with Section 12 of these Terms of Use.

Section 12: Though we hope you will stay and continue to contribute to the Projects, you can stop using our services any time. In certain (hopefully unlikely) circumstances it may be necessary for either ourselves or the Wikimedia community or its members (as described in Section 10) to terminate part or all of our services, terminate these Terms of Use, block your account or access, or ban you as a user. If your account or access is blocked or otherwise terminated for any reason, your public contributions will remain publicly available (subject to applicable policies), and, unless we notify you otherwise, you may still access our public pages for the sole purpose of reading publicly available content on the Projects. In such circumstances, however, you may not be able to access your account or settings. We reserve the right to suspend or end the services at any time, with or without cause, and with or without notice. Even after your use and participation are banned, blocked or otherwise suspended, these Terms of Use will remain in effect with respect to relevant provisions, including Sections 1, 3, 4, 6, 7, 9-15, and 17.

Twitter

Note: If your account appears to have engaged in repeated violations of the Twitter Rules, or has aggressively engaged with other accounts, you may not be presented with the option to verify by phone. In this case, you will only be able to use Twitter in a limited state for the specified time listed." https://help.twitter.com/en/managing-your-account/locked-and-limited-accounts - If someone repeatedly violates our Rules then our enforcement actions become stronger. This includes requiring violators to delete the Tweet(s) and taking additional actions like verifying account ownership and/or temporarily limiting their ability to Tweet for a set period of time. If someone continues to violate Rules beyond that point then their account may be permanently suspended.

26. Platforms should have in place mechanisms allowing to complement the above-mentioned measures with e.g. public apologies, commitments to review internal policies and processes, which may be more effective and suitable to redress, in fulfillment of their corporate social responsibility to:

- make a policy commitment to the respect of human rights
- adopt a human rights due-diligence process to identify, prevent, mitigate and account for how they address their impacts on human rights
- have in place processes to enable the remediation of any adverse human rights impacts they cause or to which they contribute