This document is a DRAFT for comments, prepared by a working group of the IGF Coalition on Platform Responsibility. Please share your comments on the mailing list of the Coalition or send them via email to luca.belli[at]fgv.br or nicolo.zingales[at]fgv.br

Glossary of

Platform Law and Policy Terms

Developed by the IGF Coalition on Platform Responsibility

Status: DRAFT

This document is a DRAFT for comments, prepared by a working group of the IGF Coalition on Platform Responsibility, a multistakeholder group under the auspices of the UN Internet Governance Forum. The elaboration process is documented at this link.

This document is an official outcome of the IGF Coalition on Platform Responsibility and was elaborated thanks to the contribution of the Coalition's Glossary Working Group. The members of the working group are: Luca Belli, Vittorio Bertola, Yasmin Curzi de Mendonça, Giovanni De Gregorio, Rossana Ducato, Luã Fergus Oliveira da Cruz, Catalina Goanta, Tamara Gojkovic, Cynthia Khoo, Stefan Kulk, Paddy Leerssen, Laila Neves Lorenzon, Chris Marsden, Enguerrand Marique, Michael Oghia, Courtney Radsch, Konstantinos Stylianou, Rolf H. Weber, Chris Wiersma, Monika Zalnieriute and Nicolo Zingales.

This document is a DRAFT for comments, prepared by a working group of the IGF Coalition on Platform Responsibility. Please share your comments on the mailing list of the Coalition or send them via email to luca.belli[at]fgv.br or nicolo.zingales[at]fgv.br

This document is a DRAFT for comments, prepared by a working group of the IGF Coalition on Platform Responsibility. Please share your comments on the mailing list of the Coalition or send them via email to luca.belli[at]fgv.br or nicolo.zingales[at]fgv.br

Why a Glossary of Key Terms on Platform Law and Policy?

Luca Belli and Nicolo Zingales

At the 2019 Internet Governance Forum (IGF), during the customary stocktaking meeting of the Coalition on Platform Responsibility[1], hereinafter "the Coalition", taking place after the annual session, the main suggestion emerging from participants as a next step in the Coalition work has been the elaboration of a **Glossary of Platform Law and Policy Terms**, so as to provide a common language for academics, regulators and policy-makers when discussing issues of platform responsibility.

The Coalition relies on spontaneous contributions of its members and thus, the Glossary initiative was launched issuing a request for suggested term, in order to shape the Glossary structure, based on the Coalition collective intuition of which list of terms may be most useful. Stakeholders agreed that the Glossary would be a "living document" that could be updated over time, while aiming at bringing together contributions from a heterogeneous range of disciplines and vocabularies. It was also agreed that the definitional efforts should recognize as much as possible the existence of competing/alternative views on the topic. For this reason, contributors were encouraged to conceived of definitions as a springboard for learning more about those views, through links and references to external sources. Links to external sources will be added in the version of the Glossary that will be uploaded on the Coalition´s website, after having received and incorporated any comments arising in the discussions at the IGF 2020.

The following action plan was shared for feedback, and subsequently implemented between May and October 2020:

1) reception of expressions of interest for the development of the Glossary and participation to the Coalition session
2) consolidation of the proposed terms and circulation of a draft list of terms to be used to compose the Glossary
3) reception of feedback on the draft list and suggestion of further terms
4) creation of a multistakeholder working group dedicated to the elaboration of the glossary (the Glossary Working Group) including all the individuals who expressed interest in the initiative[2]
5) elaboration of draft entries describing the proposed terms

[1] For further information on the Coalition, please visit the dedicate section of the Internet Governance Forum website https://www.intgovforum.org/multilingual/content/dynamic-coalition-on-platform-responsibility-dcpr

[2] The members of the working group are: Luca Belli, Vittorio Bertola, Yasmin Curzi de Mendonça, Giovanni De Gregorio, Rossana Ducato, Luã Fergus Oliveira da Cruz, Catalina Goanta, Tamara Gojkovic, Cynthia Khoo, Stefan Kulk, Paddy Leerssen, Laila Neves Lorenzon, Chris Marsden, Enguerrand Marique, Michael Oghia, Courtney Radsch, Konstantinos Stylianou, Rolf H. Weber, Chris Wiersma, Monika Zalnieriute and Nicolo Zingales.

6) consolidation of the draft entries into a first draft version of the glossary and request for comment on the first draft
7) consolidation of the updated version into a consolidated draft to be circulated at the IGF 2020 for further feedback from the IGF community
8) discussion of the draft at the 2020 session of the Coalition, during the IGF, and elaboration of a strategic approach aimed at maximising the impact of the Glossary.

While this may not be the first attempt to create a glossary of platform-related terms[3], the above illustrates the uniqueness of the open and transparent bottom-up process that was followed to achieve these results, encapsulating at its core the IGF´s principles of multistakeholder collaboration. We hope that this provides a basis for much needed mutual understanding and enables more meaningful and inclusive discussion among academics, policymakers, journalists, and other stakeholders with a keen interest in platform governance. To be continued!

1. About the IGF Coalition on Platform Responsibility

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.

To start, it should be acknowledged that a core achievement of the Coalition, well beyond the IGF's community of stakehoders, is to have coined and promoted the concept of "Platform Responsibility."[4] Such concept aims on the one hand to highlight the impact that private ordering regimes designed and implemented by platforms have on individuals' capability to enjoy their fundamental rights, and on the other hand, to interrogate the moral, social and human rights responsibilities[5] that platforms bear when setting up such regimes. Indeed, 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]

[3] See for instance http://cyberlaw.stanford.edu/blog/2018/01/glossary-internet-content-blocking-tools

[4] See L. Belli, P. De Filippi, N. 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/

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

[6] Idem.

[7] See L. Belli, P. De Filippi, N. Zingales (eds.), Recommendations on terms of service & human rights, Outcome Document n°1, 2015, tinyurl.com/toshr2015

[8] See Zingales and Belli (2014).

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, in line with the IGF mandate, which prescribes that the Forum shall "find 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 year-long participatory process, leading to the Best Practices Platforms' Implementation of the Right to an Effective Remedy[14]. 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.

Before reaching this latest phase of the coalition' work, we discussed the nuances of the **"Platform Values"** debate, with a special issue of the Computer Law and Security Review, dedicated to "Platform Values: Conflicting Rights, Artificial Intelligence and Tax Avoidance."[15] This volume aimed at promoting a discussion on the multiform notion of platform value(s) and the term "value" was

[9] See Belli, De Filippi and Zingales (2015).

[10] See Tunis Agenda (2005) para. 72.k and 72.g.

[11] See Venturini et al. (2016)

[12] See Belli and Zingales (2017)

[13] See http://bit.ly/CoEinternetintermediaries

[14] The Best Practices can be also found on the IGF website https://www.intgovforum.org/multilingual/index.php?q=filedepot_download/4905/1550

[15] A non-edited version of the Special Issue can be accessed at https://www.intgovforum.org/multilingual/index.php?q=filedepot_download/4905/1900

construed broadly 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 play a central role in the digital ecosystem, shaping the structure of online as well as offline activities. They have acquired a predominant role in digital policy circles and amongst Internet scholars, due to the enormous impact that their choices, activities and self-regulatory initiatives can have on the lives of several billion individuals. This impact is poised to increase over the incoming years [16], and for this reason, we hope that this Glossary, as well as the previous work of the Coalition will positively contribute to a better understanding of the operation of digital platforms and, consequently, more accurate and convergent policy initiatives.

References

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>

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/>

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

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

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

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>

[16] See, as an instance, Crémer J., de Montjoye Y.A. and Schweitzer H. (2019). Competition Policy for the digital era. European Commission Directorate-General for Competition; Eyler-Driscoll S., Schechter A. and Patiño C. (2019). Digital Platforms and Concentration. ProMarket and Chicago Booth Stigler Center. BRICS Competition Law and Policy Centre. (2019). Digital Era Competition Law: A BRICS Perspective. https://cyberbrics.info/digital-era-competition-brics-report/

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>

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>

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>

List of proposed terms and contributors:

1. (Digital) Access
2. Accountability
3. Aggregator
4. Amplification
5. Appeal
6. Application Program Interface
7. Arbitration
8. Automated decision making
9. Bot
10. Child Pornography / Child Sexual Abuse Material
11. Common Carrier
12. Content creator/influencer
13. Content
14. Content/web monetization
15. Coordinated flagging
16. Coordinated Inauthentic Behavior
17. Co-regulation
18. Content Curation
19. Dark patterns
20. Data Portability
21. Defamation
22. Deindexing
23. Demonetization
24. Deplatforming
25. Device Neutrality
26. Digital Rights
27. Disinformation
28. Dispute resolution (online)
29. Due diligence
30. Duty of care
31. End to End encryption
32. Federated Service
33. Fact-checking
34. Flagging
35. Filter
36. (Digital) Gatekeeper
37. Governance
38. Harassment
39. Harm (online harm)
40. Hash/Hash database
41. Hate Speech

42. **Human exploitation**
43. **Human Review**
44. **Incitement to violence**
45. **Inclusive Journalism**
46. **Information Fiduciary**
47. **Infrastructure**
48. **Intermediary Liability**
49. **Internet Safety/Security**
50. **Interoperability**
51. **Liability**
52. **Marketplace**
53. **Media Pluralism**
54. **Microtargeting**
55. **Moderation**
56. **Must-carry**
57. **Non-discrimination**
58. **Notice**
59. **Notice-and-notice**
60. **Notice-and-takedown**
61. **Notice-and-staydown**
62. **Nudging**
63. **Online Advertising**
64. **Open Identity**
65. **Open Standard**
66. **Optimization**
67. **Platform**
68. **Platform Governance**
69. **Platform Neutrality**
70. **Pornography**
71. **Prioritization**
72. **Proactive measures**
73. **Recommender systems**
74. **Red Flag Knowledge**
75. **Regulation**
76. **Remedy**
77. **Repeat Infringer**
78. **Responsibility**
79. **Revenge pornography / Non-consensual intimate images**
80. **Right to explanation**
81. **Right to be forgotten**
82. **Safe harbor**
83. **Self injury**
84. **Self-preferencing**

85. Self-regulation
86. Shadowban / Shadow banning
87. Sharing economy
88. Social network
89. Spam
90. Technological protection measures
91. Terms of service
92. Terrorist content
93.Transparency
94. User
95. User-generated content
96. Utility
97. Violence
98. User warning (of graphic content, etc.)
99. Wilful Blindness

Guidelines provided to Authors

Definitions could vary in length and focus (and the more nuance the better), but a typical definition would be between 100 and 1000 words per concept. In addition to collating the definitions in the printed booklet we'll publish for the IGF, we would aim this to be a "living document" where we can bring together contributions from a range of disciplines and vocabularies. So, by all means, do not hesitate to put your name if you are inclined to provide a definition for a particular term in the list, even if someone else has already indicated their availability to do so. We will then help coordinate to ensure that the inputs from multiple contributors are complementary, rather than duplicating efforts.

Timeline:

- 1) by 15 April: we receive your expressions of interest for the development of the Glossary and participation to the session, so that we can submit a request for an IGF 2020 Session of the Coalition (the deadline is the 22of April)
- 2) by 15 May: we consolidate your proposed terms and share a first draft of the list of terms that will compose the Glossary, to which you can add further terms or express your feedback until the 30th of May.
- 3) by 30 June: those who have expressed interest regarding the elaboration of specific terms will add their proposed definitions (ideally between 100 and 1000 words) to the shared document
- 4) by 30 July: we will allow all coalition members to share further updates and/or add alternative definitions to the list of proposed definitions
- 5) by 30 August we will have a finalised version of the Glossary that we can proofread and send to our designer to print a booklet that will be circulated at the IGF with acknowledgment of contributors.

One intrinsic limitation of this exercise is that it will be extremely hard to make a comprehensive and detailed glossary, especially considering the limited time and resources that people are likely to be able to put into this commitment in these hectic times. We can try to assuage those concerns by:

- (1) making sure the list of terms features some of the most pressing and debated issues in the debate over platform law and policy. It is probably a good choice not to get into the definition of specific types of illegal content, as that would be heavily dependent on the jurisdiction in question, but it may be a remiss not to attempt a definition of the policy issues that are used to claim/justify some form of platform responsibility beyond what the applicable law requires. So, in this sense, defining disinformation, trolling and "inauthentic coordinated behaviour" (a concept that Facebook uses to prevent coordinated forms of "misuse" of their service in violation of community standards) would appear more useful than defining terms like hate speech, defamation, violent extremism, terrorism, bullying, revenge pornography.

- (2) in the definitions, recognizing as much as possible the existence of competing/alternative views on the topic, possibly also providing links to sources where those views are more fully explained. While our goal is definitely to be schematic and clear, as you suggest, I'd like to think that this group can add value by referencing some of the cases/official documents of public authorities that address those concepts in more detail. In other words, we do not have to entirely sacrifice nuance for the sake of clarity and simplicity, if we can add links/references. Perhaps one challenge there is how to maintain as much as possible an impartial and objective perspective while also recognizing the diversity of approaches, but the two are not irreconcilable if we maintain a sufficiently abstract and high-level approach. After all, it is not uncommon for glossaries to have more than one entry for each concept.

1. (Digital) Access

In the words of Ribble (2011, 16), Digital Access is defined as "full electronic participation in society". In this context, the digital divide means inequality of access.

For Carpentier (2007), digital access includes:

- (1) access to media technology
- (2) access to skills to use the technology,
- (3) access to content that is considered relevant,
- and (4) access to the content producing organisation.

To the above, it seems essential to add access to applications and services of one's choice as well as to the technology (both hardware and software) enabling the development of applications and services (Belli & De Filippi, 2015; A4AI, 2020). These factors should be taken up together for describing 'meaningful connectivity (id.)

Access barriers related to basic connectivity ('having an internet connection') as well as (basic and advanced) digital skills are being progressively monitored nowadays, for example in the European Commission's Digital Economy and Society Index (DESI). The latter element is especially emphasised in the DESI as a major factor for the improvement of human capital related to digital access. "Digital skills" include not only 'basic usage skills' but also 'advanced skills and development' that can be used for the development of new digital goods and services (DESI 2020, p. 51). The lack of relevant skills also limits awareness of potential benefits from digitisation. Recent COVID-19 confinement measures made these two challenges more visible (e.g. children not being able to connect to remote schooling due to the lack of connection to digital infrastructure, the hardware or digital skills). Besides academia, digital literacy has been debated in policy as well (see for example Council of Europe 2016, and the forthcoming new (updated) "Digital Education Action Plan" of the European Commission, which is announced for 2020).

One of the media technology access barriers is the unequal availability of the (broadband) internet. By mid-2020, 59.6% of the world population use the internet (see Internet World Stats) with North America and Europe leading (over 85%). The access barriers to infrastructure are especially visible in rural and remote areas and developing countries without high-speed networks. For example, according to the State of Broadband report (2019), poor connectivity was a major barrier to using the internet for 43.5% respondents. According to some scholars, it is estimated that the bandwidth gap is still evolving and will be difficult to close (Hilbert, 2016).

While all the elements within the definitions that are introduced above are closely associated to the role of online platforms, the third and fourth elements - in connection to "**content**" - are especially emphasised in the **platform governance** communities. As a key term, digital access thus refers to issues of facilitating access to content and content-producing organisations. This area of law and policy is linked to public policy opportunities for developing valuable protections of online experiences. For example, in UNESCO's "Internet Freedom"-series, platforms are intermediaries that could foster freedoms online, substantively based on several principles, holding that "the Internet should be human rights-based, open, accessible for all and governed by multi-stakeholder participation" (R.O.A.M-principles; UNESCO 2014). In this way, for example, the Office of the Special Rapporteur for Freedom of Expression of the Inter-American Commission on Human Rights (Edison Lanza 2016, 15) has repeated these principles and linked them to the Rights to Freedom of Thought and Expression, the Rights to Access to Information and the Right to Privacy and Protection of Personal Data. In this line can be mentioned also the increasing number of court rulings holding that government shutdowns of the internet are illegal, such as the judgment of the Community Court Of Justice of the Economic Community Of West African States in *Amnesty International Togo VS The Togolese Republic* (ECOWAS Court of Justice, June 25, 2020; see also Pollicino, 2020). Similarly, the Council of Europe's "Human Rights Guidelines for Internet Service Providers" (2008), which sought to raise awareness for those entities providing access, stressed "the importance of users' safety and their right to privacy and freedom of expression and, in this connection, the importance for the providers to be aware of the human rights impact that their activities can have".

In July 2020, in a correspondence between Romano Prodi and David Sassoli, the President of the European Parliament supported the idea of establishing digital access as a human right (Sassoli 2020; Prodi 2020). Similar to previously developed statements, through the opinions of several courts and other institutions throughout the world (see Pollicino 2020), it points to a growing demand for establishing access in a way that grants internet users one or more separate fundamental (**digital) rights.**

References

Internet World Stats. Available at: https://www.internetworldstats.com/stats.htm

Alliance for Affordable Internet, ' Meaningful Connectivity: A New Target To Raise the Bar for Internet Access. ' (A4AI 2020) < https://a4ai.org/meaningful-connectivity/ >

Luca Belli & Primavera De Filippi, *Net Neutrality Compendium. Human Rights, Free Competition and the Future of the Internet.* (1st, Springer , 2015) <https://www.ohchr.org/Documents/Issues/Expression/Telecommunications/LucaBelli.pdf>

Nico Carpentier, ' Participation, Access and Interaction: Changing Perspectives.' [2007] New Media Worlds. Challenges for Convergence 214, 228. Edited by V. Nightingale & T. Dwyer. OUP Australia & New Zealand.

Council of Europe, ' Council of Europe Strategy for the Rights of the Child' (Council of Europe 2016) <https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=090000168066cff8>

Council of Europe, in co-operation with the European Internet Services Providers Association (EuroISPA)., ' Human rights guidelines for Internet service providers. Directorate General of Human Rights and Legal Affairs' (Council of Europe 2008) < https://rm.coe.int/16805a39d5>

Digital Economy and Society Index (DESI), ' ' (DESI 2020) < https://ec.europa.eu/digital-single-market/en/desi>

European Commission, ' Digital education action plan (updated).' (European Commission 2020) < https://ec.europa.eu/education/education-in-the-eu/digital-education-action-plan_en>

Martin Hilbert, ' The bad news is that the digital access divide is here to stay: Domestically installed bandwidths among 172 countries for 1986–2014' [2016] Telecommunications Policy 567, 581. <http://doi.org/10.1016/j.telpol.2016.01.006> (also available at <https://escholarship.org/content/qt2jp4w5rq/qt2jp4w5rq.pdf>).

Edison Lanza/Office of the Special Rapporteur for Freedom of Expression of the Inter-American Commission on Human Rights. ‘Standards for a Free, Open and Inclusive Internet.’ [2017] OEA/Ser.L/V/II, CIDH/RELE/INF.17/17 Available at <http://www.oas.org/en/iachr/expression/docs/publications/INTERNET_2016_ENG.pdf>.

Oreste Pollicino. ‘The Rights to Internet Access: Quid Iuris?’ [2020] In T*he Cambridge Handbook of New Human Rights: Recognition, Novelty, Rhetoric* edited by Andreas von Arnauld Kerstin von der Decken, and Mart Susi, 263-275. CUP. Available at SSRN: <https://ssrn.com/abstract=3397340>.

Romani Prodi. ‘La connessione sia un diritto umano (letter to Sassoli).’ [16 July 2020] *La Repubblica.* Available at <https://rep.repubblica.it/pwa/generale/2020/07/16/news/prodi_scrive_a_sassoli_la_connessione_sia_un_diritto_umano_-262146830/>.

Mike Ribble. ‘Digital Citizenship in Schools, Second edition.’ [2011] Washington, DC: International Society for Technology in Education. (see also No 1 of the “Nine Elements of Digital Citizenship, at <https://www.digitalcitizenship.net/nine-elements.html>, and: Council of Europe 2019 Digital Citizenship Education Handbook. Available at <https://rm.coe.int/16809382f9>.)

David Sassoli. ‘Il diritto al web sia una battaglia europea (reply to Romano Prodi).’ *La Repubblica*, [19 July 2020]. Available at <https://www.repubblica.it/politica/2020/07/19/news/sassoli_il_diritto_al_web_sia_una_battaglia_europea_-262314742/>.

UN Broadband Commission. ‘The State of Broadband 2019.’ [2019] Availlable at <https://broadbandcommission.org/publications/Pages/SOB-2019.aspx>.

UNESCO. ‘Fostering freedom online: the role of Internet intermediaries.’ [2014] UNESCO Series on Internet Freedom. Available at <http://www.unesco.org/new/en/communication-and-

information/resources/publications-and-communication-materials/publications/full-list/fostering-freedom-online-the-role-of-internet-intermediaries/>.

Case law:

Amnesty International Togo VS The Togolese Republic, Application No. ECW/CCJ/APP/61/18IN, ECOWAS Court of Justice (June 25, 2020). Available at <https://africanlii.org/node/7135>.

Websites:

https://www.internetworldstats.com/stats.htm

2. Accountability

Accountability refers, in the simplest conception of the term, to the condition of subjecting oneself to external oversight and control. This is a general concept which has widely different implications depending on the context in which it is used (e.g. governmental organizations, private companies, and even computer algorithms). However, as it can be evinced from this general definition it typically includes a transparency component, and a component of submission to external control, both of which can be manifested in different forms depending on the content and the target of accountability. For instance, the control component of accountability of an institution can be exercised through budgetary control and judicial review. Thus, it is crucial to understand when talking about accountability *who* is accountable *for what*, and *to whom*.

Traditionally, accountability has been structured as a bidimensional principles. In well-functioning institutions, the executive is subjected to both vertical and v horizontal accountability. (O'Donell, 1999). The latter is imposed upon organisations by individuals (vertical) through their collective monitoring and actions, while the latter is imposed by public bodies that are specifically tasked to control and – when necessary – restrain the undue actions of a given institution. Transparency and freedom to access information is essential for both dimensions. Vertical accountability is maximised when individuals can act via civic organizations ("civil society") or media. Horizontal accountability is maximised when public entities created to check potential abuses and inefficiencies are well resourced and can act independently.

The primary consequence of accountability is responsiveness, meaning that the entity in question effectively responds to the demands of transparency and external control. This typically presupposes the existence of tools and procedures that allow the exercise control and oversight. The form of accountability can be prescribed with some level of specificity by law, as it is the case for instance in the case of data protection law. The EU General Data Protection Regulation (GDPR) and the Brazilian General Data Protection Law, for instance, explicitly contain a principle of accountability. Such principle requires that organisations put in place appropriate technical and organisational measures to be able to demonstrate compliance, such as: adequate documentation on what personal data are processed, how, to what purpose, how long; documented processes and procedures aiming at tackling data protection issues at an early state when building information systems or responding to a data breach; and the appointment of a Data Protection Officer.

As far as platforms are concerned, the issue of accountability has acquired a particular connotation in the context of injunctions. This is because, under virtually every regime of intermediary liability, injunctions can be imposed against intermediaries regardless of the existence of a primary or even secondary duty to undertake a certain action. For instance, in Europe such injunctions are permitted by article 12(3), 13 (3) and 14(3) of the E-Commerce Directive, based on the rationale that every intermediary that gets entangled with harm can be required to provide assistance. In Germany, this rationale led to the doctrine of *Störerhaftung,* i.e. 'disturber' or 'interferer' liability, allowing injunctions against persons who causally contribute to an infringement in violation of a reasonable duty to review. Despite terminological differences, the essence remains the same: although there may be no liability for the damages caused by the infringing activity, failure to execute the injunctions will lead to a different kind of liability, potentially of criminal nature, for contempt of court.

References

Martin Husovec, Injunctions Against Intermediaries in the European Union: Accountable But Not Liable? (1st, Cambridge University Press, 2017)

O'Donnell, Guillermo, ' Horizontal Accountability in New Democracies' in Andreas Schedler, Larry Diamond, Marc F. Plattner: (eds), *The Self-Restraining State: Power and Accountability in New Democracies* (1st, Lynne Rienner Publishers, London 1999).

3. Aggregator

Definition proposed after the second round of comments. To be included by November 2020

4. Amplification

In the context of platform governance, ‘amplification’ refers to actions (typically by platforms) that magnify the visibility or reach of certain content, viewpoints, or speakers. The phrase is most commonly used to refer to platform content recommendations and other algorithmic rankings, in cases where particular content is seen to be given an unfair or otherwise unwarranted ranking. Other instances of ‘amplification’ can include the use of bots or astroturfing to disseminate content, and the use of platform advertising services to similar ends. The concept has gained traction due to the growing attention for non-illegal forms of online harm, such as disinformation, where the speech as such is unlikely to be prohibited and removed, but its rapid spread is nonetheless seen as a source of concern and a target for regulation.

Amplification is not a legal term, but it is increasingly used in associated policy debates. Perhaps most notably, the European Commission’s Communication “Tackling Online Disinformation: A European Approach” discusses amplification at length (European Commission, 2018). It identifies three different kinds of amplification: (1) “algorithm-based” amplification, which relates recommender systems, (2) “advertising-driven amplification”, which relates to platform advertising services, and (3) “technology-enabled amplification”, which refers to the use of bots and the use of fake accounts. In the Commission’s diagnosis of disinformation, the problem is thus not merely that disinformation is created and disseminated, but that it is amplified by various factors to reach a disproportionate audience.

The UN Special Rapporteur on Freedom of Expression, David Kaye, displays a similar understanding of the term in an open letter to Mark Zuckerberg regarding the Facebook Oversight Board, dated 1 May 2019 (Kaye 2019a). He proposes that the Board should have access to information about “and factors that may amplify the content at issue (e.g. recommendation algorithms, bot accounts, ad policies).” In a note to the UN General Assembly, the Special Rapporteur also suggested that tools be developed to combat hate speech *inter alia* through “de-amplification” (Kaye 2019b).

The recent White House Executive Order On Preventing Online Censorship does not address amplification in the same length, but does allege that online platforms have “amplified China’s propaganda” and offers this as a ground for further regulation, although it does not define amplification or further elaborate on the claim (US Executive Order on Preventing Online Censorship 2020).

A key challenge in identifying “amplification” in online platforms is that it implies a baseline of non-amplified treatment, which may not be available. For recommender systems, it is often alleged that hate speech or disinformation are amplified by algorithms that prioritize attention and engagement, but this begs the question what an appropriate (i.e. non-amplified) ranking for this content would be instead. For instance, a hateful website may be considered ‘amplified’ if it is ranked as the first result on Google Search, but what if it is the second? The tenth? The 100th? Thus, although claims of ‘amplification’ can seem objective and analytical, they may conceal an ultimately subjective and political assessment about the appropriate configuration of recommender systems, and about media diversity in general.

A narrower conception of 'amplification' is possible in which it only singles out direct positive discrimination of content, such as Facebook's prioritization of trusted news sources and YouTube's prioritization of coronavirus victims. But this narrower conception does not correspond with current usage, as outlined above, as this tends to also include attention-optimizing systems that benefit hate speech and disinformation indirectly. In this light, amplification should be understood as a broad term that can refer to a wide range of factors in the online media environment which facilitate the spread of certain content, whether as an intentional design feature or as an unintentional by-product.

References

European Commission, ' Communication on Tackling Online Disinformation: A European Approach.' (White House 2018) < https://www.whitehouse.gov/presidential-actions/executive-order-preventing-online-censorship/>

David Kaye, ' Mandate of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression' (OHCHR 2019) < https://www.ohchr.org/Documents/Issues/Opinion/Legislation/OL_OTH_01_05_19.pdf>

David Kaye, 'UN General Assembly: Promotion and protection of the right to freedom of opinion and expression' (OHCHR 9 Oct 2019) <https://www.ohchr.org/Documents/Issues/Opinion/A_74_486.pdf>

The White House, ' US Executive Order on Preventing Online Censorship' (White House 2020) < https://www.whitehouse.gov/presidential-actions/executive-order-preventing-online-censorship/>

5. Appeal

This entry: (I) provides an understanding of the notion of appeal in legal doctrine; (II) elucidates the basic functions of an appeal mechanism; (III) offers examples of when an appeal mechanism may be needed; iv) and highlights the recommendations put forward by the IGF Coalition on Platform Responsibility with regard to appeal mechanisms.

i) The concept of appeal

The concept of appeal is grounded on the necessity of correcting error, which may always occur when decisions are taken. In this perspective, appeal mechanisms that allow for error correction are an essential feature of due process and rule of law principles at the core of any well-functioning legal systems. An appeal is therefore a mechanism thanks to which defendants that deem to have been victim of a wrongful judgement may have these concerns addressed and, eventually, corrected. The fundamental goal of any appeal mechanisms is making sure that decisions are taken observing procedural fairness, correcting errors, including arbitrary or irrational applications of existing rules and procedures.

Appeals are crucial for ensuring that justice is done in each case and, for this reason, they are included in modern human rights instruments. Indeed, modern legal systems provide appeal mechanisms for correcting errors as historical evidence demonstrate that errors about examining specific facts or about applying existing rules to frame those facts are expected to occur regularly.

Appellate procedures vary substantially among legal systems and the scope of appellate review is generally limited to claims and defences addressed in the proceeding that is challenged (usually defined as “first-instance proceeding”). (ALI and UNIDROIT, 2006:27)

There are three general standards of review by appellate bodies: questions about the application of substantial rules (so-called “questions of law”), questions regarding how the facts have been analysed ((so-called “questions of fact”), and matters of procedure or discretion.

In general terms, appeals can:

(i) constitute a repeat exercise of the lower body's decisionmaking, both in terms of fact-finding and the application of the relevant law/rules to those facts;

(ii) accepting the factual determinations of the lower court (particularly when the lower court heard evidence but the appellate court didn't) but reviewing whether the relevant laws/rules were applied correctly (or even, in common law countries at least, whether the relevant laws/rules need re-interpreting);

(iii) reviewing the procedural aspects of the lower court/body (e.g. was the process flawed in some way by admitting irrelevant evidence, or ignoring relevant evidence).

The so-called "*de novo*" review describes a review of a lower body (usually a court) by a superior body (i.e. an appellate court). De novo review is used in questions of how specific normative provisions were applied or interpreted. In this type of review, the appellate court can repeat in its entirety the fact-finding exercise of the lower body or court. De novo judicial review can reverse the decision that is challenged, and, for this reason, this type of review is qualified using the Latin expression "de novo" which means "over again" or "anew." As the appellate body re-examines the issue from the beginning, this type of review is defined as "nondeferential review" because the decision is taken anew without deferring to the lower body's decision.

Review standards can focus on both questions of fact and questions of law. The former are based on a more deferential approach. This means that the appellate body will limit its analysis to the facts – such as re-evaluating "clearly erroneous" the evidence – and subsequently defer the case to the body that took the contested for a new application of the rules in light of the factual scrutiny conducted by the appellate body.

Lastly the "nuclear option" amongst the standard of review most is the so-called "arbitrary and capricious" standards. This is the most deferential type of review, as the appellate body determines that a previous decision is invalid because it was made on unreasonable grounds or without any proper consideration of circumstances.

ii) The function of appeal mechanisms

The possibility of an error occurring is an unavoidable feature of any decision-making system. Appeals allow to correct possible errors, thus serving several types of functions. As tellingly explained by Marshall (2011), the primary function of the modern right of appeal is to protect against miscarriages of justice and, indeed, appeals aim at mitigating the risks and consequences of wrongful decisions (or, even worst, convictions, in case of criminal law). Wrongful decisions are always possible and arise either when a defendant - or anyone bringing a claim in civil proceeding - is wrongly judged or when a defendant does not receive a fair trial.

The core function of an appeal mechanism is therefore to provide redress form a wrongful judgement that may be the result of an extremely ample spectrum of possible reasons, including failure to accurately assess evidence; mislead or deception by irrelevant or fabricated evidence; or lack of consideration of exculpatory evidence.

A second core function of appeal mechanisms is to remedy the lack of a fair trial. Such situation may occur when decisions are taken applying existing (procedural) rules in an anomalous way or when clear and foreseeable procedural rules are missing.

Importantly, the existence of appeals mechanisms *per se* provides legitimacy to a system, while stimulating trust in such system. When efficient appeal mechanisms exist, all individuals and entities subject to a specific juridical system will know that rules are applied in a fair, transparent, and consistent fashion.

iii) When an appeal mechanism is needed

An appeal mechanism is needed to challenge erroneous decisions based on procedural or substantial ground. Such situations may occur in the following circumstances:

a. When the body that took the decision had no jurisdiction (or, more generally speaking, no competence) to take such decision or when the powers of have been utilised improperly.
b. When the procedure was applied unfairly.
c. When the decision is not reasonable.
d. When the decision is not proportional.
e. When the decision is not compatible with Human Rights obligations.
f. When the decision contradicts the legitimate expectations of an individual or entity subject to a given set of rules.
g. Or when the decision does not provide sufficient reasons, justifying why it has been taken.

iv) Recommendations put forward by the IGF Coalition on Platform Responsibility

All platforms should offer their users the possibility to appeal any decisions concerning them. Appeal systems shall respect the core minimum of the right to be heard, including: (1) a form of process, which is made available to users in clear and explicit an easily-comprehensible terms, mandating the respect of the guarantees of independence and impartiality; (2) the right to receive notice of the allegations and the basic evidence in support, and comment upon them; and
(3) the right to a reasoned decision.

References

ALI (The American Law Institute) and UNIDROIT, *UNIDROIT Principles of Transnational Civil Procedure* (1st, , 2006)

Peter D. Marshall, ' A Comparative Analysis of the Right to Appeal' [fall 2011] Duke Journal of Comparative & International Law Volume 22, Number 1.

6. Application Program Interface

An Application Program Interface (also known as API, or "middleware") is any well defined interface which identifies the service that one component, module or application provides to other software elements (de Souza et al, 2004). APIs can be grouped into two types: those which are more intensively computational, based on an execution engine and those which are declarative, based on presentation engines. In Oracle v Google (N.D. Cal. June 20, 2012), ECF No. 1211), on the scope of copyright protection for interface specifications, the US District Court distinguished three categories of information provided by these interfaces, namely (a) declaration or method header lines; (b) the method and class names; and (c) the grouping pattern of methods. In essence, these interfaces contain the information and instructions which enable third-party applications to run atop existing computer programs without a loss of functionality.

In the context of platform regulation, APIs are being advanced as a possible solution to give more control to individuals, both on how their personal data is collected and used (see My Data Declaration, 2017), and on how the platform moderates their feed (Keller, 2019). There are technical challenges, however, in how to operationalize this model, including the technical standards on which such APIs should be based, the legal safeguards to preserve the protection of individuals' personal data (in particular when the API allow the transferring of data involving third parties) and the limits to regulation which may impose an API obligation to private entities (including the interference with the right to conduct business and freedom of expression).

References

C. R. B. de Souza et al, ' Sometimes You Need to See Through Walls- A Field Study of Application Programming Interfaces' [2004] Computer supported cooperative work, ACM Press 63, 67

Ashwin Van Rooijen, ' The Software Interface between Copyright and Competition Law: A Legal Analysis of Interoperability in Computer Programs' [2010] Kluwer Law, Alphen aan den Rijn

Daphne Keller ' Platform Content Regulation – Some Models And Their Problems' (The Center for Internet and Society, 6 May 2019) < http://cyberlaw.stanford.edu/blog/2019/05/platform-content-regulation-—-some-models-and-their-problems>

MyData, 'Declaration of Principles' [2017] <https://mydata.org/declaration>

7. Arbitration

The World Intellectual Property Organization (WIPO) defines arbitration as a "procedure in which a dispute is submitted, by agreement of the parties, to one or more arbitrators who make a binding decision on the dispute. In choosing arbitration, the parties opt for a private dispute resolution procedure instead of going to court." A simpler definition given by the Cambridge Dictionary states that the arbitration process is a way of solving an argument between people by helping them to consent and commit to a common and acceptable solution. It is important to highlight that both sides in the dispute have to agree to pursue an arbitral solution, that is to have the matter solved through the mediation of an arbitrator.

Arbitration is a type of dispute/conflict resolution method. In its process, the parties that have previously agreed to arbitration can settle the dispute outside of the courtroom. That way, it is usually much faster than legal procedures, for its informality and privacy, and also the reason why this procedure is often chosen rather than the litigation process. The disputes are resolved by an impartial third party, known as the arbitrator, whose decision is legally binding for all parties.

As for the online process of arbitration, its premise is the same as the real life one. The difference is that the resolution of conflicts can be done entirely online, using video calls for hearings and online systems for uploading multiple types of evidence (documents, photos, videos, etc) making it possible to resolve disputes without having to appear in person, and by that minimizing the costs of the process.

The most prominent example of the use of online arbitration is in disputes over Internet domains (Mania, 2015). A parallel can be made between domain names in the Internet and the system of business identifiers that are protected by intellectual property rights and has existed long before the arrival of the Internet, and conflicts regarding both issues can be resolved through arbitration process. The most common reason for disputes over Internet domain names comes from the practice known as "cybersquatting" (WIPO, nd), which is when a random person register a domain name under famous people or business trademarks and offers them for sale at prices far beyond the cost of registration to give them the rights of the domain name. Therefore, "any institution or person who considers that a registered domain name conflicts with the legitimate rights or interests of that institution or person may file a complaint with any of the competent domain name dispute resolution providers" (CCPIT, nd). If there is a consent between parties to solve their domain name dispute by an arbitration institute, the details of the dispute need to be submitted to the chosen institution.

After the Complaint is made and the entity against whom the Complaint was made files a Response, the dispute resolution service provider is chosen by all parties and "shall implement a system whereby panels of experts are responsible for the resolution of disputes" (CCPIT, nd). The number of panelists who form the panels is one to three, and they all have to be experts listed on an online file available for the complainants and respondents to select from. The panelists must be impartial and independent during the whole arbitration procedure and have no material interest with parties of either sides. After the conclusion of the process, the issuance of the panels of experts decision must be notified to all parties and the decision must be implemented, meaning that the domain name in question may be cancelled or transferred (WIPO, nd).

The WIPO - which is mandated to promote the protection of intellectual property worldwide - conducted extensive consultations with members of the Internet community around the world, after which it prepared and published a report containing recommendations dealing with domain name issues. Based on the report's recommendations, ICANN adopted the Uniform Domain Name Dispute Resolution Policy (UDRP), and under the standard dispute clause of the Terms and Conditions for the registration of a gTLD domain name, the registrant must submit to the UDRP proceedings. Also, the Protocol on Cybersecurity in International Arbitration ("Cybersecurity Protocol") provides guidance on reasonable information security measures that the parties and arbitrators can take, particularly in light of increasingly virtual hearings and paperless document transfer.

An example of alternative online arbitration practices for resolution of domain name conflicts is the Sistema de Administração de Conflitos de Internet, or the SACI-Adm, created by the Internet Steering Committee in Brazil (CGI.br). This method is specifically made for the resolution of disputes between the holder of a domain name in .br (brazilian ccTLD) and any third party that disputes the legitimacy of the domain name registration made by the holder. The scope of the SACI-Adm procedures is limited to requests for cancellation and transfer of domain - therefore, any claim related to obtaining indemnities cannot be dealt with in this context. The aspects of the SACI-Adm procedure are similar to the ones presented in the UDRP, the main difference between them is that in the brazilian regulation the .br Information and Coordination Center (NIC.br) doesn't allow the transfer of the domain name in conflict from the beginning of the arbitration with the SACI-Adm procedure until its termination (Angelini, 2012).

There's also the "baseball-style" arbitration process, that was used in the case between Google and news publishers in Australia. In this form, an arbitration attorney is selected to decide the main issues that are in dispute between two or more parties. "The arbitration attorney, contrary to popular belief, is not usually free to decide anything he or she pleases, instead, each party participating in the arbitration must submit a proposal to him or her in advance". This kind of arbitration is named as the "baseball-style" due to the discretion exercised by the arbitration attorney to these proposals. It is also sometimes called "final-offer or "either/or" arbitration because of the limits imposed upon the arbitration attorney.

It is important to state that there's a difference between Online Dispute Resolution (ODR) and online arbitration. The ODR is a wide field and emcompasses many types of dispute resolution practices that use online methods and tools to explore the convenience and efficiency of internet communications and make the resolution process easier and faster. The term addresses every aspect from electronic filing of resolution process submissions and transfer of documents to online hearings. The variety of ODR can envelop interpersonal disputes, "including consumer to consumer disputes (C2C) or marital separation, to court disputes and interstate conflicts" (Petrauskas & Kybartiene, 2011) . On the other hand, online arbitration is an essential part of ODR by which two or more parties can solve any disagreement originated from their contractual relationship online and it is mostly used for domain names conflict resolution, Business to Business ('B2B') cross-border e-commerce disputes, and in some degree used for the resolution of traditional cross-border commercial disputes (Amro, 2019).

References

World Intellectual Property Organization (WIPO). 'What is Arbitration?' Available at: <https://www.wipo.int/amc/en/arbitration/what-is-arb.html>

Cambridge Dictionary. 'Arbitration Meaning'. Available at: <https://dictionary.cambridge.org/pt/dicionario/ingles/arbitration>

Mania, Karolina. 'Online dispute resolution: The future of justice'. [2015] *International Comparative Jurisprudence*, *1*(1), 76-86. Available at: <https://www.sciencedirect.com/science/article/pii/S2351667415000074>

WIPO. 'Frequently Asked Questions: Internet Domain Names'. Available at: <https://www.wipo.int/amc/en/center/faq/domains.html#5>

CCPIT. 'Resolution of Internet Domain Name Disputes'. Available at: <https://www.ccpit-patent.com.cn/node/1105>

WIPO. 'WIPO Guide to the Uniform Domain Name Dispute Resolution Policy (UDRP)'. Available at: <https://www.wipo.int/amc/en/domains/guide/#b1>

WIPO. 'WIPO Internet Domain Name Process.' [1999]. Available at: <https://www.wipo.int/amc/en/processes/process1/report/index.html>

ICCA Report. 'ICCA-NYC Bar-CPR Protocol on Cybersecurity in International Arbitration'. [2020] Available at: <https://www.arbitration-icca.org/media/14/76788479244143/icca-nyc_bar-cpr_cybersecurity_protocol_for_international_arbitration_-_print_version.pdf>

Angelini, Kelli. 'SACI: o Sistema Administrativo de Conflitos de Internet implementado para domínios no ".br".' [2012] Available at: <https://www.politics.org.br/edicoes/saci-o-sistema-administrativo-de-conflitos-de-internet-implementado-para-dom%C3%ADnios-no-%E2%80%9Cbr%E2%80%9D>

Petrauskas, Feliksas & Kybartiene, Eglè. 'Online dispute resolution in consumer disputes'. *Jurisprudencija*, *18*(3). [2011] Available at: <https://www.mruni.eu/upload/iblock/0c4/8_Petrauskas_Kybartienuy-1.pdf>

Amro, Ihab. 'Online Arbitration in Theory and in Practice: A Comparative Study in Common Law and Civil Law Countries'. [2019] Available at: <http://arbitrationblog.kluwerarbitration.com/2019/04/11/online-arbitration-in-theory-and-in-

practice-a-comparative-study-in-common-law-and-civil-law-countries/?doing_wp_cron=1592416143.9425508975982666015625>

Arbitration.com. What is Baseball Arbitration?. [9 June 2011] Available at: <http://www.arbitration.com/articles/what-is-baseball-arbitration.aspx>

8. Automated decision making

Automated decision-making (ADM) generally refers to a process or a system where the human decision is supported by or handed over to an algorithm.

ADM are increasingly used in several sectors of our society and by different actors (both private and public). For instance, ADM can be embedded in a standalone software that produces a medical **recommendation** for a patient, an online behavioural advertising system that shows a certain content to a specific target, a credit score system to determine whether one can get a loan, an algorithm that selects the most interesting CV for a position, a **recognition filter** that scans and bans a **user-generated content** from a **platform**, an automated ticketing system which fines drivers exceeding speed limits, an algorithm to assess the recidivism risk, smart contracts (Finck, 2019), etc.

Given the spread of ADM and the potential impact on individuals, the Council of Europe has issued the Recommendation CM/Rec(2020)1 on the human rights impacts of algorithmic systems, promoting a lawful and human-centric design of ADM.

Otherwise, ADM is not generally regulated as such, but its deployment can be captured by a wide spectrum of laws.

In Europe, for instance, when an ADM processes personal data, the General Data Protection Regulation (GDPR) applies. The GDPR does not expressly define the concept of ADM, but it provides additional rules in case that a solely ADM process produces legal effects concerning the data subject (i.e. the person to whom data refers) or similarly affects him or her (Article 22 GDPR). In the literature, several doubts have been raised with reference to the exact meaning of "decision", "solely automated", "legal effects" and "similarly affect" (Mendoza and Bygrave, 2017; Bygrave, 2020). The Working Party Article 29 (WP29, now European data Protection Board) has provided an interpretation of these concepts, suggesting that: 1) the ADM can be fed with any kind of data (whether they are provided directly from the individual, observed or otherwise inferred); 2) a "solely" automated decision means there is no human involvement at any stage of the processing; 3) with "legal effects" entails that the decision must affects the legal rights and freedoms of individuals (e.g. a system that automatically refuse the admission to a country); 4) "similarly affects" intends to include other possible negative effects which may seriously impact the behaviour of individuals, e.g. potentially leading to discrimination (for example, a system denying someone an employment opportunity) (WP29, 2018). However, the provision does not seems to entail an evaluation of the negative impact on groups (Veale and Edwards, 2018), but several authors have argued in favour of expanding the data protection framework from the individual level to the collective one (Taylor, Floridi, and van der Sloot, 2016; Mantelero, 2018; Brkan, 2019).

As a general rule, the GDPR prohibits such kind of processing, unless: 1) it is necessary for entering into, or perform, a contract between the data subject and the controller (i.e. the entity leading the processing); 2) it is authorized by the law, which lays down appropriate safeguards for the rights and legitimate interests of the data subject; 3) the data subject explicitly consent to it. When exceptions 1 and 3 apply, the data controller can carry out the ADM, but it has to implement suitable measures to protect individuals' rights and freedoms (Article 22(3) GDPR). Among them, the GDPR lists three main rights that have to be guaranteed to individuals: 1) to obtain human intervention on the part of the controller; 2) to express his or her point of view; 3) to contest the decision. The implementation of such measures has been differently embraced by Member States (Malgieri, 2019).

Finally, in case the ADM involves the processing of particular categories of data (defined at Article 9 GDPR), such as health data or data revealing ethical or political opinions, the GDPR provides a specific discipline. In particular, ADM cannot be performed, unless there is the explicit consent of the data subject or the processing is necessary for a substantial public interest. In both cases, the controller must adopt suitable measures to protect data subjects' rights, freedoms, and legitimate interests.

Another important legal issue concerning ADM in the framework of GDPR relates to the **transparency** of the system, i.e. the possibility to understand the logic involved in the algorithm performing a decision according to Article 22. There has been a lively debate in the literature about the existence of the so-called "**right to explanation**" in the GDPR (Goodman, Flaxman, 2016; Malgieri, Comandé, 2017; *contra* Wachter, Mittelstadt, Floridi, 2017). Whether it can be envisaged directly or indirectly in the black letters of the GDPR, there is a convergence toward the elaboration of solutions that can promote the transparency of ADM and "XAI", i.e. explainable AI (Wachter, Mittelstadt, Russell, 2017; Edward, Veale, 2017; Kaminski, Malgieri, 2019; Brkan, Bonnet, 2020). The importance of explainability has been stressed by the High-Level Expert Group on Artificial Intelligence among the requirements for trustworthy AI (High-Level Expert Group on AI, 2019).

A similar - although not identical - provision on ADM is included at Art. 11 of Directive (EU) 2016/680 (Law enforcement Directive). Being a Directive, it is not self-executive, therefore Member States have to implement it in their national law. The Law Enforcement Directive explicitly forbids the use of ADM in criminal matters where the decision produces an adverse legal effect concerning the data subject or significantly affects him or her. Such a prohibition can be overcome only by Union or Member State law to which the controller is subject and which provides appropriate safeguards for the rights and freedoms of the data subject, at least the right to obtain human intervention on the part of the controller.

Data protection law is probably the most comprehensive framework tackling the phenomenon of ADM. However, ADM is also regulated by other branches of law.

For instance, when the ADM is likely to produce discriminatory results, the protection granted by anti-discrimination law kicks in. Both direct and indirect discrimination are prohibited by the European Convention on Human Rights and EU Law. For example, an ADM leading to the exclusion of a member from an online platform would be deemed to be illegal if based on race or proxies for it, such as African-American names (direct discrimination, see Edelman, Luca, Svirsky, 2017). Similarly, it would be considered indirect discrimination if a supposed neutral measure is likely to impact on a protected category. For instance, to anchor the earnings of platform's drivers to the distance and time they travel appears to be a neutral decision. However, studies show that women drive at a lower average speed, therefore they are likely to take less rides, and, as a consequence their pay is substantially lower than their male colleagues (Cook et al., 2018).

Nevertheless, the anti-discrimination legal framework suffers important limitations, since the digital discrimination brought by ADM transcends the traditional protected attributes (Borgesius, 2018; Xenidis, Senden, 2020). In online behavioural advertising, for example, people might be discriminated because the inferential analytics associates them with a certain group (even if they are not part of it), exposing them to price discrimination or exclusions from lucrative job ads (Wachter, 2020).

In the field of consumer protection, ADM has been recently taken into account in relation to transparency of online **marketplaces**. The "Omnibus Directive", amending the Consumer Right Directive (Directive 2011/83/EC), established that when the price is personalized on the basis of an ADM, the consumer must be informed about it. However, the provision imposes to disclose the "whether" but not the "how" of the ADM (Jabłonowska 2019). It must be said, though, that if the system processes personal data and the price personalization falls within the notion of Article 22 GDPR, the explainability and corresponding remedies (Article 22(3) GDPR) will apply to this situation. Such a transparency requirement, in any case, does not extend to 'dynamic' pricing which depends on real-time market demands. Differently, in case of rankings, the Omnibus Directive requires to inform the consumers about the main parameters and their weighting behind the "relative prominence given to products, as presented, organised or communicated by the trader". The concept of ranking is constructed in a technological neutral way, therefore it might consist in an ADM.

Another sector regulating ADM is medical devices. When a software stand-alone can be used for medical purposes, i.e. provide information to support diagnostic or therapeutic decisions or monitor vital physiological parameters, it will have to comply with Regulation (EU) 2017/745 that establishes the steps to bring a medical device for human use on the market.

ADM is also addressed in the field of **content recognition technologies**. For instance, the new Copyright in the Digital Single Market Directive (Directive (EU) 2019/790) provides a new form of direct **liability** for online platforms (more specifically, online content-sharing service providers, such as YouTube) for their users' upload. To avoid this form of liability platforms have two possible options. The golden road traced by the Directive is to negotiate a license with the rightholder in order to make available the content uploaded by users. As an alternative, online platforms have to demonstrate, among other things, to proactively ensure the unavailability of the (infringing) content. This latter option has attracted the criticisms of copyright scholars and civil society representatives, being a provision which will lead to establish upload filters and limit freedoms on the Internet (Cerf et al., 2018; Kretschmer et al., 2019; Reda, 2019). The Directive establishes some guarantees: **users** rights – such as quotation, criticism, pastiche – shall be preserved (Article 17(7), the **proactive measures** cannot lead to any general monitoring obligation (Article

17(8)), and the platform must provide for adequate complaint and redress mechanism to allow users contesting the decisions about access denial and removal of content (Article 17(9)). However, several doubts remain as to the impact of these provisions on fundamental rights such as freedom of expression and data protection (Quintais et al., 2019; Quintais, 2020; Romero Moreno, 2020; Samuelson, 2020; Schmon, 2020).

References

Brkan, Maja, and Grégory Bonnet., ' Legal and Technical Feasibility of the GDPR's Quest for Explanation of Algorithmic Decisions: of Black Boxes, White Boxes and Fata Morganas.' [2020] European Journal of Risk Regulation 18, 50.

Brkan, Maja. 'Do algorithms rule the world? Algorithmic decision-making and data protection in the framework of the GDPR and beyond.' [2019] International journal of law and information technology 91, 121.

Bygrave, Lee. "Article 22", in Kuner, Christopher, Lee A. Bygrave, and Christopher Docksey. *Commentary on the EU General Data Protection Regulation (GDPR). A Commentary*. Oxford University Press, 2020, 522.

Cerf, Vint et al. 'Joint Letter to the European Parliament' [2018], Eletronic Frontier Foundation <https://www.eff.org/files/2018/06/13/article13letter.pdf.>

Cook, Cody, et al. 'The gender earnings gap in the gig economy: Evidence from over a million rideshare drivers.' [2018] No. w24732. National Bureau of Economic Research.

Edelman, Benjamin, Michael Luca, and Dan Svirsky. 'Racial discrimination in the sharing economy: Evidence from a field experiment.' [2017] American Economic Journal: Applied Economics 1, 22.

Edwards, Lilian, and Michael Veale. 'Slave to the algorithm: Why a right to an explanation is probably not the remedy you are looking for.' [2017] Duke L. & Tech. Rev. 18.

Finck, Michele. 'Smart Contracts as Automated Decision-Making under Article 22 GDPR.' [2019] International Data Privacy Law 1, 17.

Goodman, Bryce, and Seth Flaxman. 'EU regulations on algorithmic decision-making and a "right to explanation".' [2016] ICML workshop on human interpretability in machine learning (WHI 2016), New York, NY. http://arxiv. org/abs/1606.08813 v1.

High-Level Expert Group on AI, Policy and investment Recommendations for Trustworthy AI, [2019] <https://ec.europa.eu/digital-single-market/en/news/policy-and-investment-recommendations-trustworthy-artificial-intelligence.>

Kaminski, M. E., & Malgieri, G. 'Algorithmic Impact Assessments under the GDPR: Producing Multi-layered Explanations.' [2019] Available at SSRN 3456224.

Kretschmer, Martin et al. 'The Copyright Directive: Articles 11 and 13 must go Statement from European Academics in advance of the Plenary Vote on 26 March 2019' [2019], <https://www.create.ac.uk/wp-content/uploads/2019/03/Academic_Statement_Copyright_Directive_24_03_2019.pdf?x42614>

Malgieri, Gianclaudio, and Giovanni Comandé. 'Why a right to legibility of automated decision-making exists in the general data protection regulation.' [2017] International Data Privacy Law.

Malgieri, Gianclaudio. 'Automated decision-making in the EU Member States: The right to explanation and other "suitable safeguards" in the national legislations.' [2019] Computer Law & Security Review 35.5 (2019): 105327.

Mantelero, Alessandro. 'AI and Big Data: A blueprint for a human rights, social and ethical impact assessment.' [2018] Computer Law & Security Review 754, 772.

Mendoza, Isak, and Lee A. Bygrave. 'The right not to be subject to automated decisions based on profiling.' [2017] EU Internet Law. Springer, Cham 77, 98.

Quintais, João, et al. 'Safeguarding user freedoms in implementing article 17 of the copyright in the digital single market directive: recommendations from European Academics.' [2019] Available at SSRN 3484968.

Quintais, João. 'The New Copyright in the Digital Single Market Directive: A Critical Look.' [2020] European Intellectual Property Review.

Reda, Julia. 'EU copyright reform: Our fight was not in vain' [2019], <https://juliareda.eu/2019/04/not-in-vain/>

Romero Moreno, Felipe. ''Upload filters' and human rights: implementing Article 17 of the Directive on Copyright in the Digital Single Market.' [2020] International Review of Law, Computers & Technology 1, 30.

Samuelson, Pamela. 'Pushing Back on Stricter Copyright ISP Liability Rules.' [2020] Michigan Technology Law Review, Forthcoming.

Schmon, Christoph. 'Copyright Filters Are On a Collision Course With EU Data Privacy Rules' [2020] <https://www.eff.org/deeplinks/2020/02/upload-filters-are-odds-gdpr>

Taylor, Linnet, Luciano Floridi, and Bart Van der Sloot, eds. 'Group privacy: New challenges of data technologies.' [2016] Vol. 126. Springer.

Veale, Michael, and Lilian Edwards. ‘Clarity, surprises, and further questions in the Article 29 Working Party draft guidance on automated decision-making and profiling.’ [2018] Computer Law & Security Review 398, 404.

Wachter, Sandra, Brent Mittelstadt, and Chris Russell. ‘Counterfactual explanations without opening the black box: Automated decisions and the GDPR.’ [2017] Harv. JL & Tech. 31 (2017): 841.

Wachter, Sandra, Brent Mittelstadt, and Luciano Floridi. ‘Why a right to explanation of automated decision-making does not exist in the general data protection regulation.’ [2017] International Data Privacy Law 76, 99.

Wachter, Sandra. ‘Affinity profiling and discrimination by association in online behavioural advertising.’ [2020] *B*erkeley Technology Law Journal 35.2.

WP29, ‘Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679’, WP251rev.01, [3 October 2017- 6 February 2018], <https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612053>

Xenidis, Raphaële, and Linda Senden. ’EU non-discrimination law in the era of artificial intelligence: Mapping the challenges of algorithmic discrimination.’ [2019] Ulf Bernitz et al (eds), General Principles of EU law and the EU Digital Order (Kluwer Law International, 2020) 151, 182.

Zuiderveen Borgesius, Frederik. ‘Discrimination, artificial intelligence, and algorithmic decision-making.’ [2018] Study for the Council of Europe, <https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73>

9. Bot

The word “bot” is a tech slang, an abbreviation for “robot”, in this sense, they share the same conceptual core: a reprogrammable machine built to perform a variety of tasks (RIA, s.d.). In the specific case of online bots, they are labeled as automatic or semi-automatic computer programs that run over the Internet (Franklin & Graesser, 1996; Gorwa & Guilbeault, 2018).

One of a bot's main assets is its ability to perform simple and repetitive tasks faster than a human, and at scale, with some arguing that the most repetitive tasks in human jobs (and some jobs entirely) will be replaced by this increasingly software automation (Bort, 2014) . Bots’ activities online may have impressive proportions and, in this perspective, it is worth noting that 37.9% of total internet traffic in 2018 was carried out by bots, with 53.4% of them coming from the United States (Imperva, 2020).

Some experts and companies divide bots into two broad categories: benevolent and malicious bots (Jones, 2015; Cloudfare, s.d.). The friendly ones are subdivided according to the functions they perform: social bots simulates human behavior in automated interactions to manage social media accounts; commercial bots, usually used to increase online engagement in companies or as chatbots to autonomously conduct a conversation instead of employing people to communicate with consumers; web crawlers bots, also known as Google bots, which scan content on webpages all over the Internet and gather useful information; entertainment bots, that are designed to be appreciated aesthetically (art bots) or as characters to play against (game bots); helpful or informational bots, that surface helpful information and usually push notifications and breaking news stories.

The above mentioned examples are usually utilised to help and optimize human actions and tasks. However, several types of malicious bots exist, such as content-scraping bots, or bots that spread spam , or carry out credential stuffing attacks. Malicious bots can be: scrapers bots, that are designed to steal content or huge amounts of data; spam bots, designed to automatically circulate unrequested content around the web in order to drive traffic to the spammer’s website, fill out forms automatically, congest servers or just cause disturbance; scalper bots, also known as automated purchasing, that are designed to purchase sought-after products and services; and hacker bots, that exploit security vulnerabilities to distribute malware, deceive individual people, attack websites or entire networks, in this latter case, devices that are affected are known as “zombies” and infected networks are “botnets” (combination of “robot” and “network”) .

These botnets are programmed to perform mischievous tasks such as DDoS attacks, theft of confidential information, click fraud, cyber-sabotage, and cyber-warfare. As an instance, in September 2016, a botnet called Mirai was responsible for one of the biggest cyber-attack in history when launched a DDoS attack on the servers of Dyn, one of the main DNS providers, which resulted in a blackout for various internet services (Antonakakis et al. 2017).

More recently, a type of malicious bot has dominated digital policy debates: impersonators bots. This type of bot mimic human behavior predominantly in order to manipulate public opinion and exercise social control (Bessi & Ferrara, 2016; Howard et.al, 2018). Twitter and other social networks are the home of such political bots, and for this reason they have been highly criticized and have become subject to regulation around the world.

References

Antonakakis, Manos et al. Understanding the Mirai Botnet [2017]. Available at: <https://www.usenix.org/conference/usenixsecurity17/technical-sessions/presentation/antonakakis>

Bessi, Alessandro & Ferrara, Emilio. ‘Social bots distort the 2016 US Presidential election online discussion’. [2016] First Monday, 21(11-7). Available at: <https://firstmonday.org/ojs/index.php/fm/article/view/7090/5653>

Bort, Julie. ‘Bill Gates: People Don't Realize How Many Jobs Will Soon Be Replaced By Software Bots’, [2014] Available at: <https://www.businessinsider.com/bill-gates-bots-are-taking-away-jobs-2014-3>

Botnerds. ‘Types of Bots: An Overview’. Available at: <http://botnerds.com/types-of-bots/>

Cloudflare. ‘What is a bot?’ Available at: <https://www.cloudflare.com/learning/bots/what-is-a-bot/>

Delaney, Kevin J. ‘The robot that takes your job should pay taxes, says Bill Gates’. [2017] Available at: <https://qz.com/911968/bill-gates-the-robot-that-takes-your-job-should-pay-taxes/>

Franklin, Stan & Art, Graesser. “Is It an Agent, or Just a Program?: A Taxonomy for Autonomous Agents.” [1996] In Intelligent Agents III Agent Theories, Architectures, and Languages, 21–35. Lecture Notes in Computer Science. Springer, Berlin, Heidelberg. Available at: <https://www.researchgate.net/profile/Stan_Franklin/publication/221457111_Is_it_an_Agent_or_Just_a_Program_A_Taxonomy_for_Autonomous_Agents/links/0f317530ba440e7979000000/Is-it-an-Agent-or-Just-a-Program-A-Taxonomy-for-Autonomous-Agents.pdf>

Gorwa, Robert & Guilbeault, Douglas. ‘Unpacking the social media bot: A typology to guide research and policy.’ [2018] <https://arxiv.org/pdf/1801.06863.pdf>.

Howard, Philip N., Woolley, Samuel & Calo, Ryan. ‘Algorithms, bots, and political communication in the US 2016 election: The challenge of automated political communication for election law and administration’, [2018] Journal of Information Technology & Politics 81, 93 DOI: 10.1080/19331681.2018.1448735

Imperva. ‘Bad Bot Report 2020: Bad Bots Strike Back.’ [2020] Available at: <https://www.imperva.com/resources/resource-library/reports/2020-Bad-Bot-Report/>

Jones, Steve. ‘How I learned to stop worrying and love the bots’. [2015] Social Media+ Society, 1(1), 2056305115580344. Available at: <https://journals.sagepub.com/doi/pdf/10.1177/2056305115580344>

Robotic Industries Association (n.d.). ‘Defining The Industrial Robot Industry and All It Entails’ Available at: <https://www.robotics.org/robotics/industrial-robot-industry-and-all-it-entails>

10. Child Pornography / Child Sexual Abuse Material

NB: While the term “child pornography” has beenused traditionally, and continues to be used on occasion, it is increasingly understood to be inappropriate since it suggests a degree of complicity or consent on the part of the child. Instead, the term “child sexual abuse material” (CSAM) (or sometimes “child sexual abuse imagery” (CSAI)) is now considered to be a more appropriate describe the phenomenon and is the term used here.

This entry: (i) sets out the agreed international definitions of the term as found in relevant legal instruments and (ii) provides examples of existing regulatory responses to child sexual abuse material.

(i) Agreed international definitions

Child sexual abuse material is prohibited under a number of international legal instruments, two of which provide relatively clear definitions. The first is Article 2(c) of the Optional Protocol to the Convention on the Rights of the Child on the sale of children, child prostitution and child pornography (OP-SC-CRC), which defines the term “child pornography” as “any representation, by whatever means, of a child engaged in real or simulated explicit sexual activities or any representation of the sexual parts of a child for primarily sexual purposes”. This may be considered the minimum core of what constitutes child sexual abuse material.

The second, broader definition is provided by Article 9 of the Budapest Convention, where “child pornography” includes: “pornographic material that visually depicts (a) a minor engaged in sexually explicit conduct; (b) a person appearing to be a minor engaged in sexually explicit conduct; (c) realistic images representing a minor engaged in sexually explicit conduct”. While paragraph (a) broadly overlaps with the definition in OP-SC-CRC, paragraphs (b) and (c) go further by including persons appearing to be minors and realistic images representing minors. Under the Budapest Convention, states are, however, free not to apply those paragraphs, meaning that paragraph (a) is the core part of the definition.

Instruments vary in terms of the age at which a person is considered a “child” or a “minor”. While OP-SC-CRC does not define “child”, the term is defined in Article 1 of the Convention on the Rights of the Child itself as any “human being below the age of eighteen years unless under the law applicable to the child, majority is attained earlier”. The Budapest Convention is narrower in the discretion it offers, defining “minors” as “all persons under 18 years of age”; while it does allow state parties to set a lower age-limit, however this cannot be less than 16 years.

(ii) Existing regulatory responses

Most states have sought to comply with their obligations under international law to prohibit child sexual abuse material through criminalisation, creating specific criminal offences relating to child sexual abuse material. The International Centre for Missing and Exploited Children has published model legislation (and a global review of existing legislation) which include as a minimum definition, “the visual representation or depiction of a child engaged in a (real or simulated) sexual display, act, or performance” with “child” defined as “anyone under the age of 18” (ICMEC, 2018).

In addition to criminalisation, many governments take action, sometimes through regulation and sometimes informally, to prevent access to child sexual abuse imagery online. In many states, governments have encouraged ISPs, in particular, to use filters to block access to certain websites known to carry or have carried child sexual abuse imagery. Governments have also encouraged and supported other self-regulatory initiatives, such as the creation of hash databases of known child sexual abuse imagery by companies and non-governmental organisations, which are then shared so as to more easily block known images across many platforms. These include the Internet Watch Foundation (in the United Kingdom), the National Centre for Missing and Exploited Children (in the USA) and the Canadian Centre for Child Protection (in Canada).

References

International Centre for Missing and Exploited Children (ICMEC), ‘Child Sexual Abuse Material: Model Legislation & Global Review’, [2018] 9th Edition, Available at: https://www.icmec.org/wp-content/uploads/2018/12/CSAM-Model-Law-9th-Ed-FINAL-12-3-18.pdf.

11. Common Carrier

Common carriage is defined by the duties imposed on public networks in exchange for their right to use public property as a right of way, and other privileges:

Common carriers and public carriers are under duty to carry goods lawfully delivered to them for carriage. The duty to carry does not prevent carriers from refusing to transport goods that they do not purport to carry generally. Carriers may indeed restrict the commodities that they will carry. Further, everywhere, carriers may refuse to carry dangerous goods, improperly packed goods, and goods that they are unable to carry on account of size, legal prohibition, or lack of facilities (Longley (1967); Ridley, Jasper and Whitehead (1982); Encyclopædia Britannica (2009).

This definition offers several reasons not to common carry that can be extended to Internet Service Providers – spam and viruses for instance may be refused. In common law countries such as the United Kingdom and United States, carriers are liable for damage or loss of the goods that are in their possession as carriers, unless they prove that the damage or loss is attributable to certain excepted causes (‘Acts of God, acts of enemies of the Crown, fault of the shipper, inherent vices of the goods, and fraud of the shipper’, perils of the sea and particularly jettison). In the wonderfully descriptive language of the English common law (Longley 1967): ‘Fault of the shipper as an excepted cause is any negligent act or omission that has caused damage or loss— for example, faulty packing. Inherent vice is some default or defect latent in the thing itself, which, by its development, tends to the injury or destruction of the thing carried. Fraud of the shipper is an untrue statement as to the nature or value of the goods. And jettison in maritime transport is an intentional sacrifice of goods to preserve the safety of the ship and cargo.’

That provides several more reasons for loss – one thinks of the loss of undersea cables, or alleged foreign power Denial of Service (DoS) attacks, as we saw in Chapter 1. It might be stretching a definition to suggest that P2P streams can be ‘jettisoned’ in order to allow other traffic to progress during peaktime congestion.

It is worth stating what common carriage is not. It is not a flat rate for all packets. It is also not necessarily a flat rate for all packets of a certain size. It is, however, a mediaeval non-discrimination bargain between Sovereign and transport network or facility, in which an exchange is made: for the privileges of classification as a common carrier, those private actors will be granted the rights and benefits that an ordinary private carrier would not. As Barbara Cherry has written, common carriers are not a solution to a competition problem, they far predate competition law. They prevent discrimination between the same traffic type – if I offer you transport of your High Definition video stream of a certain protocol, then the next customer could demand the same subject to capacity, were the Internet to be subject to common carriage (it is not).

Citizens believe they have ancient rights of way and of service. The United Kingdom Carriers Act of 1830 was the first legislation for carriage of goods, codifying the common law. The Act applied to all common carriers by land (‘more effectual Protection of Mail Contractors, Stage Coach Proprietors, and other Common Carriers’ according to the Carriers Act 1830 CHAPTER 68 11_Geo_4_and_1_Will_4) , including road and railway carriage, then in its infancy for passengers but well-established for coal and other commodities. The United Kingdom Railways Act 1844 does include provisions for common carriage and ‘Parliamentary trains’ (low cost trains that stop at all stations, later known as ‘milk trains’ because they ran pre-dawn to avoid inconveniencing more expensive trains at peak hours). Common carriers in mediaeval times included farriers and public houses (every horse to be shoed and person to be allowed shelter without discrimination between travelers). As per Lane v. Cotton (1701) 1Ld.Raym. 646, 654 (per C.J. Holt)`‘If a man takes upon him a public employment, he is bound to serve the public as far as the employment extends; and for refusal an action lies, as against a farrier refusing to shoe a horse...Against an innkeeper refusing a guest when he has room...Against a carrier refusing to carry goods when he has convenience, his wagon not being full.’

Common carriage should not be confused with charging tolls for higher speed networks, though the Turnpike Riots of 18th Century England were associated with turning the King’s Highway into a private road, and UK opposition to road charging continues to this day.

Telecoms networks were established to be common carriers as they achieved maturity, following telegraphs, railways, canals and other networks. Noam explained in 1994 the practice:

Common carriage, after all, is of substantial social value. It extends free speech principles to privately-owned carriers. It is an arrangement that promotes interconnection, encourages competition, assists universal service, and reduces transaction costs. Ironically, it is not the failure of common carriage but rather its very success that undermines the institution. By making communications ubiquitous and essential, it spawned new types of carriers and delivery systems… the pressure on common carriers come from two other directions: private NGNs offered by systems integrators; and broadband services offered by cable television operators. Neither operates as a common carrier, nor is it likely to. Noam (1994, p. 435) explains that: ‘When historically they [infrastructure services] were provided in the past by private firms, English common law courts often imposed some quasi-public obligations, one of which one was common carriage. It mandated the provision of service of service to willing customers, bringing common carriage close to a service obligation to all once it was offered to some.’

He thus forewarned that net neutrality would have to be the argument employed by those arguing for non-discriminatory access, as well as accurately predicting the death of common carriage ten years later. Note under common carriage, discrimination is quite possible, but not between customers, only between identical loads: see National Association of Regulatory Utility Commissioners v. FCC, 525 F.2d 630, 642 (D.C.Cir. 1976).

In the United States, it was finally established that a public telegraph company (and more especially the largest) has a duty of non-discrimination towards the public- see Western Union Telegraph Co. v. Call Publishing Co., 181 U.S. 92, 98 (1901). The loss of common carriage is an epoch-breaking move towards deregulation, which means that attempts to ensure universal access to an unfettered Internet will require new regulation.

References

Carriers Act 1830 CHAPTER 68 11_Geo_4_and_1_Will_4 at http://www.opsi.gov.uk/RevisedStatutes/Acts/ukpga/1830/cukpga_18300068_en_1

Encyclopædia Britannica, ‘Common Carrier’ at http://www.britannica.com/EBchecked/topic/128177/common-carrier

Longley, Henry N. ‘Common Carriage of Cargo’, [1967] Matthew Bender & Co.: New York.

Noam, Eli M. ‘Beyond liberalization II: the impending doom of common carriage’, [1994] Telecommunications Policy 435, 452.

Ridley, Jasper and Geoffrey Whitehead. ‘The Law of the Carriage of Goods by Land, Sea and Air’, [1982] 6th ed., Shaw: Crayford, Kent

12. Content creator/influencer

During the past 10 years, peer-to-peer platforms have democratized and decentralized media services. It is currently possible for any individual around the world to make a social media channel or account (e.g. on YouTube, Instagram, or TikTok), and make content for a living. These developments are facilitated by increased opportunities, from a marketing and technological perspective, to monetize online presence (see also the entry for 'content/web monetization'). Within this framework, a new marketing phenomenon, known as 'influencer marketing', has spread online. It consists of a monetization model based on 'reviews and endorsements of products online, usually communicated through social networks' (Riefa & Clausen, 2019). Outside marketing aspects, the term 'content creator' is used to emphasize the fact that social media' users take the career path of making media content, especially since controversies surrounding the non-disclosure of advertising, or the low levels of diligence exercises by some social media personalities have attracted a negative connotation of the term 'influencer' (The Guardian, 2019).

'Content creators' might therefore be a better term to refer to influencers other than those who engage in influencer marketing as their primary business model. However, it is important to stress that so-called "influencers" are only a subset of content creators, which is a general term that may be utilised to identify any individual user creating content either for professional or for personal purposes.

Furthermore, it must be noted that defining influencers is no easy task. From a semantic perspective, the Cambridge Dictionary defines influencers as 'a person who is paid by a company to show and describe its products and services on social media, encouraging other people to buy them' (Cambridge Dictionary, 2020). In a study on social media advertising, the European Commission proposed a similar definition: 'a person who has a greater than average reach and impact through word of mouth in a relevant marketplace, and influencer marketing relies on promoting and selling products or services through these individuals' (European Commission, 2018). So far, the concept has been integrated in numerous self-regulatory measures around the world, such as the Dutch Advertising Code for Social Media & Influencer Marketing, where influencer marketing is understood to be a marketing activity involving an advertiser and its distributors, in relation to a (paid) communication about a product or brand for the benefit of the advertiser (Art 2(e) Advertising Code for Social Media & Influencer Marketing, 2019). The exercise of influence is a core component of these views. According to the Word of Mouth Marketing Association, influence is 'the ability to cause or contribute to another person taking action or changing opinion/behavior'. These definitions are built around three common considerations: 1) The existence of a transaction whereby a person is paid to promote something; 2) The person operates on social media; 3) The person has a sphere of influence on which it exercises commercial persuasion.

While these features are a reasonable representation of part of the influencer industry, they lead to an incomplete picture on three grounds. First, such features only characterize influencers stricto sensu, namely to indicate those social media users who engage in influencer marketing as a business model (see Figure 1 below).

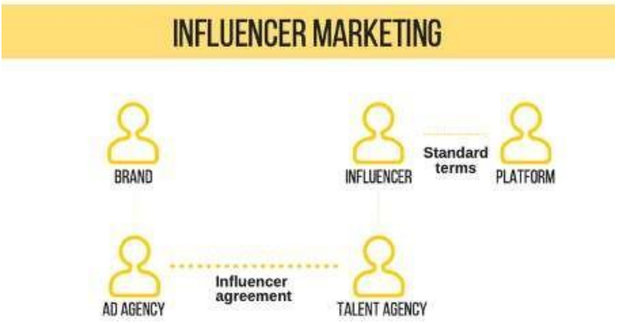

Figure 1 - Goanta & Wildhaber, 2020

However, the notion of influencers should be seen from a broader perspective. Influencers come in all sizes and species, and they can range from humans to pets, or even accounts of curated content (e.g. meme accounts such as those used by Michael Bloomberg in his 2020 Instagram campaign ads). At the same time, on the basis of the size of their following, there can be mega-influencers (most renowned creators in a given industry), micro-influencers (rising stars with less followers and popularity than mega-influencers), or nano-influencers (small-scale influencers focused on word-of-mouth in more granular communities). Secondly, influencer marketing regards a plethora of monetization models to build their revenue, such as crowdfunding on Patreon, direct selling of own merchandise (‘merch’), or ad revenue through programmes such as AdSense or Instagram TV (see also the entry for ‘content/web monetization’). Thirdly, these strategies do not only concern commercial content but also extend to political speech. Influencers do not exercise commercial persuasion connected to commercial transaction but also engage in communications of a different nature than commercial (e.g. political communication). Moreover, with the rise of social justice influencers, influence can also be exercised through e.g. the promotion of social messages and calls for action to support civil society organizations through donations, which is different than promoting goods/services, although the activity itself may be based on similar monetization models as commercial influencers (e.g. endorsement contracts; De Gregorio & Goanta, 2020).

On the basis of these insights, we propose a more all-encompassing definition of an influencer, as the person behind a social media account who creates monetized media content with the goal of exercising commercial or non-commercial persuasion, and that has an impact on a given follower base. From a policy perspective, addressing influencers marketing could affect the right to freedom of expression and, therefore, regulators should take into account the degree of interferences of potential regulatory interventions. Within this framework, consumer law can play a critical role in defining what the boundaries of unfair commercial practices are, and to what extent some practices are prone to manipulating consumer behaviour or political ideas for commercial gain.

References

European Commission, ‘Behavioural Study on Advertising and Marketing Practices in Online Social Media’ [2018]. Final Report, available at <https://ec.europa.eu/info/files/advertising-and-marketing-practices-online-so- cial-media-final-report-2018_en>

Riefa, C & Clausen, L. ‘Towards Fairness in Digital Influencers’ Marketing Practices’, [2019] Journal of European Consumer and Market Law 64.

'Advertising Code for Social Media & Influencer Marketing' (Reclamecode 2019) < https://www.reclamecode.nl/nrc/reclamecode-social-media-rsm/>

Word of Mouth Marketing Association. ‘The WOMMA Guide to Influencer Marketing’, [2017] available at <http://getgeeked.tv/wp-content/uploads/uploads/2018/03/WOMMA-The-WOMMA-Guide-to-Influencer-Marketing-2017.compressed.pdf>.

Stokel-Walker, C. ‘Instagram: beware of bad influencers…’, [2019] The Guardian, available at <https://www.theguardian.com/technology/2019/feb/03/instagram-beware-bad-influencers-product-twitter-snapchat-fyre-kendall-jenner-bella-hadid>.

Goanta, C. & Wildhaber, I. ‘Controlling Influencer Content Through Contracts: A Qualitative Empirical Study of the Swiss Influencer Market’, [2020] C. Goanta & S. Ranchordás (eds.), The Regulation of Social Media Influencers (Elgar).

De Gregorio, G. & Goanta, C. ‘The Influencer Republic’, [2020]

13. Content

This entry: (i) sets out the way that the term “content” is used in common parlance and (ii) provides examples of existing regulation which define the term (or synonyms of it).

(i) Use in common parlance

There is no universally agreed definition of the term “content”; it does not appear in any major international instruments. At its broadest, the term can be considered to refer to the visual and aural elements of the internet that users experience via websites and applications. This would include all of the text, images, videos, animations and sounds that a user can see, hear or otherwise access.

In recent years, the term “content” is also increasingly used to refer to a specific visual or aural element, with the term “piece of content” referring to such an element. This could be a particular post that a user has uploaded to a social media platform, or an image or video uploaded or shared.

(ii) Existing regulation

While “content” is the term used in common parlance, it is not the only term used - at least so far - in regulation which sets out rules relating to online content. New Zealand’s Harmful Digital Communications Act 2015 and the USA’s Communications Decency Act both use the term “content” (although neither defines it). Australian Criminal Code Amendment (Sharing of Abhorrent Violent Material) Act 2019 uses the term “material” instead, noting that “material” can be audio material, visual material or audio-visual material. The European Union’s E-Commerce Directive (2000) uses the term “information”, but without defining it.

14. Content/web monetization

‘Web monetization’ is a term of art used to identify two aspects. First, from a broad perspective of the Internet’s history, it may encompass the ways in which content on the Web may be monetized, namely how to turn website traffic into profit. Second, from a more narrow perspective, it may refer to the infrastructure needed to achieve that, and in particular to the recent browser API standard with the same name (‘Web Monetization’), developed by Coil in collaboration with Mozilla Foundation.

Nutshell history of web monetization

The Internet as we know it is built on the Transmission Control Protocol (TPC, later complemented with the Internet Protocol resulting in the TCP/IP standard). Focused on the transmission of data across networks, in other words, access to information, this protocol was not originally designed or tailored for commercial gain, as ‘there was no native payment system built into the web at the time’ (Melendez, 2019).

To capitalize on the traffic generated on the Internet, web – namely website - monetization entailed, from very early on, reliance on web advertising, namely the displaying of ads on (popular) websites. Web advertising consists of popular practices such as pay per click, pay per impression, paid subscriptions or donations. Pay per click entails a marketing strategy by which an advertiser pays a platform that displays its ads on advertising networks (e.g. Google AdSense). Each time a user clicks on a link displayed e.g. in a query made on a search engine (e.g. Google, but also Youtube), the advertiser has to pay for that click. In comparison, the pay per impression model entails that the advertiser needs to pay for every time an ad is shown to a user, and does not require the user to click it to that end. Additional models rely on other streams of income, such as paid subscriptions (e.g. newspapers that require viewers to pay for access to content), or donations (e.g. Wikipedia).

Social media

Apart from traffic generated on regular websites (e.g. Blogger, Medium, but also personal websites), a massive source of online presence for current Internet users is social media. With 3.81 billion active users on social media, 2.49 billion of which are Facebook and 2 billion on Youtube, social media has long been a fertile environment for the monetization of user attention. Social media content creators, also referred to as influencers (see entry for ‘Influencers/content creators’) bring views, likes and clicks to platforms that are keep on making new features to reward them. The business models behind content monetization are in constant fluctuation, and so far can be broadly divided in four different models (see Figure 1 below): influencer marketing, ad revenue, subscription/tokenization/crowdfunding and direct selling. Influencer marketing entails the payment for the endorsement of a good or service, made by an advertiser or brand to an influencer or their representative. Ad revenue is the monetization generated through the display of ads on the social media channel belonging to a creator (e.g. AdSense or Instagram TV). Subscription models entail paying a fee to access content made by a creator on a given platform (e.g. Youtube), and is similar to crowdfunding on platforms such as Patreon, where a subscription is made to support the creator across any platform they may use. In addition, tokenization allows followers to spend money on platform-specific tokens, which they can give to creators in specific moments during their enjoyment of the content (e.g. on Twitch there is even a combination of subscriptions and tokens, where a subscription offers so-called 'emotes', and subscribers can make gifts available to creators). Lastly, direct selling entails content creators selling own branded goods to their fan base.

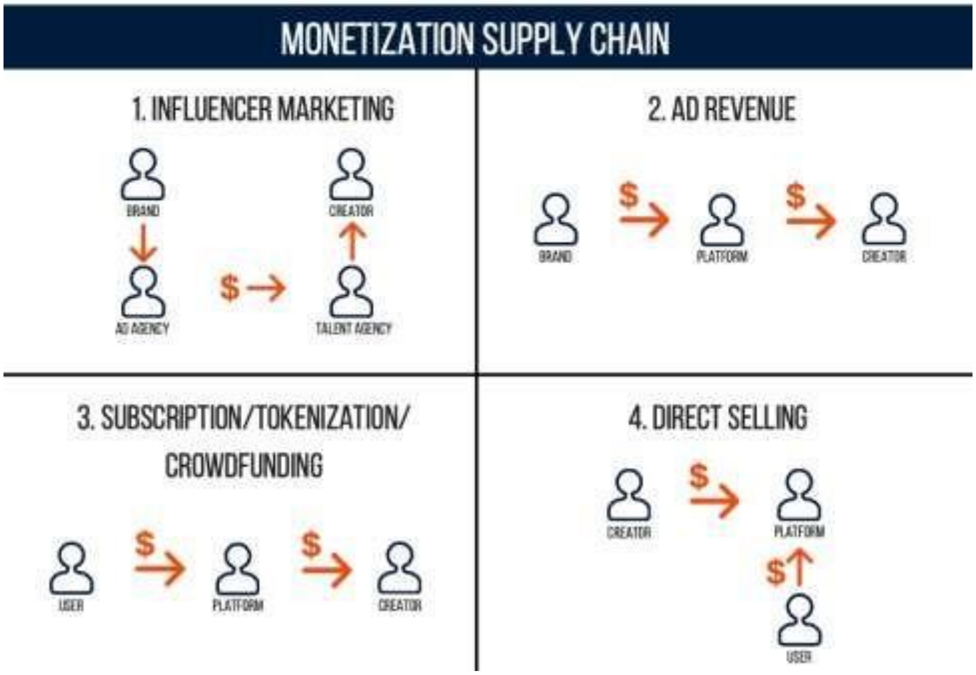

Figure 1 - Monetization business models

The Web Monetization protocol

In the past decades, information transfer protocols have been used to get as much information as possible on user behaviour. Under the guise of personalizing advertising to fit individual needs and preferences, behavioural advertising targets users across platforms (Centre for Data Ethics and Innovation). New data sharing architectures such as Solid aim to challenge the Internet's advertising-based business model and give more privacy and ownership to users with respect to their data (Solid, 2020). The Web Monetization protocol is a proposed browser API standard which supports the generation of a payment stream between the user and the website being viewed (Web Monetization, 2020). Payment streams are based on an open protocol suite called Interledger, used to send payments across different ledgers (Interledger, 2020).

References

Centre for Data Ethics and Innovation, ‘Review of online targeting: Final report and recommendations, [February 2020], <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/864167/CDEJ7836-Review-of-Online-Targeting-05022020.pdf>.

Ken Melendez, ‘The State of Web Monetization’, Medium, [13 December 2019], <https://coil.com/p/kenmelendez/The-State-of-Web-Monetization/KTVijO7ah>.

Statista, ‘Social media - Statistics & Facts’, [18 May 2020]<https://www.statista.com/topics/1164/social-networks/> Solid, <https://solid.mit.edu>.

Web Monetization, <https://webmonetization.org>.

Interledger, <https://interledger.org>.

15. Coordinated flagging

See "**flagging**". Coordinated flagging refers to a form of large-scale organized campaign where a group of individuals decide to simultaneously flag the social media content of a specific individual or specific group of individuals, marking such content as offensive, with the purpose of having the impacted individual(s) banned or suspended from the platform, or their content taken down. This is commonly understood to be a form of technology-facilitated abuse and harassment, where often the content is not offensive, and in fact may be content that itself calls attention to abuse that the author is experiencing, and then is further targeted for speaking out about. Coordinated flagging is one of several ways in which online abusers game or exploit content moderation features on platforms to target their victims, often members of historically marginalized or vulnerable communities, as a form of silencing with the intent or effect of driving such users away from online spaces. For example, "In 2012, accusations swirled around a conservative group called 'Truth4Time,' believed to be coordinating its prominent membership to flag pro-gay groups on Facebook." (Crawford & Gillespie, 2016).

Such campaigns may also include a central political purpose other than the politics involved in targeting historically oppressed groups online. Brittany Fiore-Silfvast has described coordinated flagging as a type of "user-generated warfare" (Fiore-Silfvast, 2012), and gives the following example of a coordinated flagging campaign known as "Operation YouTube Smackdown" (OYS), with the slogan, "Countering the Cyber-Jihad one video at a time":

OYS began out of a conversation among conservative bloggers who were inspired by the potential for private citizens to fight the war through the Internet. […] The blogger called on his blogger friends to join the effort by volunteering to scour YouTube for footage from the "enemy" and flag it for YouTube's corporate staff to review and remove. After one of the bloggers volunteered his blog to serve as the coordinating site of operations, a handful of other bloggers began to connect their blogs and direct their readership to OYS. It was there and then that the conservative bloggers and their readership began organizing themselves into a network army that would fight Internet terrorists on YouTube (Fiore-Silfvast, 2012 p. 1972-73).

Coordinated flagging is also known as "strategic flagging" or "organized flagging", and results in user flags playing a governing role "not expressing individual and spontaneous concern but as a social and coordinated proclamation of collective, political indignation—all through the tiny fulcrum that is the flag, which is asked to carry even more semantic weight." (Crawford & Gillespie, 2016 p. 421)

References

Kate Crawford, Tarleton Gillespie, 'What is a flag for? Social media reporting tools and the vocabulary of complaint' [2016] New Media & Society 410, 420.

Brittany Fiore-Silfvast, "User-Generated Warfare: A Case of Converging Wartime Information Networks and Coproductive Regulation on YouTube" [2012] International Journal of Communications 1965.

16. Coordinated Inauthentic Behavior

Coordinated inauthentic behaviour (CIB) is a term frequently associated with concepts such as disinformation, online misinformation, computational propaganda, and the mass leveraging of bots or fake accounts to carry out a particular set of actions or disseminate particular messages across social media platforms. The term originated with Facebook, which has defined CIB as “domestic, non-government campaigns that include groups of accounts and Pages seeking to mislead people about who they are and what they are doing while relying on fake accounts”, including “fake engagement, spam and artificial amplification” (Facebook, 2020). To be clear, CIB can also be engaged in or instigated by foreign actors and governmental actors; in these cases, such activity is still considered a form of CIB, but Facebook then categorizes it as “Foreign or Government Interference (FGI)”. The more general definition that describes the behaviour itself, regardless of the actor involved, may thus be more universally applied outside of Facebook’s specific internal categorization.

Despite the increasing popularization of the term, observers have noted that coordinated inauthentic behaviour still does not have a completely stable or clear definition. For example, platform regulation scholar Evelyn Douek questioned whether or not CIB would include a “tactical and relatively sophisticated” campaign where teenagers on TikTok and K-pop fans purposely reserved tickets to a campaign rally for the U.S. president in Tulsa, Oklahoma, in June 2020, in order to “artificially inflate expected attendance numbers and mess with the Trump campaign’s data collection” while displacing genuine supporters and ensuring large numbers of empty seats at the rally (Gleicher, 2018). In response, “Facebook’s head of security … explained that the teens’ stunt wouldn’t have met Facebook’s definition of CIB because it did not involve the use of fake accounts or coordinate to mislead users of the platform itself (as opposed to misleading people *off* the platform)” (Douek, 2020). As another example complicating definitional boundaries, Douek cites “how a network of 14 purportedly independent large Facebook pages drove traffic to the conservative site the Daily Wire, one of the most popular publications on Facebook, including by publishing the *same* articles at the *same* time with the *same* text” (Douek, 2020). Facebook’s explanation for *not* treating this activity as CIB was that “CIB is reserved for the most egregious violations and this didn’t meet the threshold because the accounts weren’t fake” (Douek, 2020).

Thus, the definition of “coordinated inauthentic behaviour” is still in flux, both in attempts to interpret and establish exactly what Facebook itself means by CIB, as well as in establishing what CIB means as a standalone term in the field of platform regulation generally, independent of what Facebook itself may consider to be CIB for its own internal purposes. As Douek points out, “Rare is the piece of online content that is truly authentic and not in some way trying to game the algorithms. Coordination and authenticity are not binary states but matters of degree, and this ambiguity will be exploited by actors of all stripes.” (Douek, 2020).

References

‘March 2020 Coordinated Inauthentic Behavior Report’ [2 April 2020], Facebook <https://about.fb.com/news/2020/04/march-cib-report/>.

Nathaniel Gleicher, ‘Coordinated Inauthentic Behavior Explained’ [6 December 2018], Facebook <https://about.fb.com/news/2018/12/inside-feed-coordinated-inauthentic-behavior/>.

Evelyn Douek, ‘What Does “Coordinated Inauthentic Behavior” Actually Mean?’ [2 July 2020] Slate, <https://slate.com/technology/2020/07/coordinated-inauthentic-behavior-facebook-twitter.html>.

17. Co-regulation

Self-regulation as developed by the concerned (market) participants can be strengthened by involving governmental agencies into the rule-developing and rule-implementing processes. Depending on the degree of involvement, the respective forms of co-operative rule-making are called (i) co-regulation, (ii) regulated self-regulation, (iii) directed self-regulation or (iv) audited self-regulation (Weber 2014: 23/24). The most common model balancing the interests of international organizations, States, businesses, and civil society is the co-regulation as described hereinafter. Such kind of multistakeholder approach can serve legitimate State purposes as well as efforts of the private sector.

Co-regulation as model (having been coined by Hoffmann-Riem [2000] in relation to the media markets) means that the government provides for a general framework which is then substantiated by the private sector; i.e. the State legislator sets the legal yardsticks and leaves the codification of the given principles by way of specific rules to private bodies. Thereby, regulation can remain flexible and innovation-friendly. Additionally, the government remains involved in the private rule-making activities at least in a monitoring function supervising the progress and the effectiveness of the initiatives in meeting the perceived objectives (Senn 2011: 43, 139-148, 230; Marsden, Meyer and Brown 2020: 9).

Co-regulation is a regulatory model leaving the actual “regulator” independent from the government as long as the rules remain within the legislative framework. Whether for example Codes of Conduct developed by the private sector need to get an approval from a public authority to become effective depends on the applicable legal provisions (being partly the case in financial markets). Such a requirement appears to be justified if private rule-making risks to implement not sufficiently adequate normative standards. Governments can assess the representativeness of self-regulatory standards and judge the appropriateness of best practices; interventions appear justified if a higher level of protection measures is desirable (Marsden, Meyer and Brown 2020: 9).

Many examples for co-regulatory mechanisms exist (Marsden, Meyer and Brown: 9/10), for example in the media markets and in the Internet regulatory ecosystem (i.e. Nominet, EURID). Social media platforms are often exposed to court proceedings (e.g. Delfi, MTE v. Hungary); the implementation of standards and best practices can help to limit the respective risks.

References

Hoffmann-Riem Wolfgang. ‘Regulierung der dualen Rundfunkverordnung’, [2000] Baden-Baden.

Marsden Chris, Trisha Meyer and Ian Brown. ‘Platform values and democratic elections: How can the law regulate digital disinformation?’, [2020] Computer Law & Security Review 36-105373 1, 18. Available at <https://www.journals.elsevier.com/computer-law-and-security-review>

Senn Myriam. ‘Non-State Regulatory Regimes’, Understanding Institutional Transformation’, [2011] Berlin

Weber Rolf H. ‘Realizing a New Global Cyberspace Framework’, [2014] Zürich

18. Content Curation

(I) In the context of online services, “content curation” refers to the selection of relevant content from a larger subset of available content. Thorson and Wells define curation as the “production, selection, filtering, annotation or framing of content” (Thorson & Wells, 2016). In the modern environment of information abundance, curation fulfils an essential function: “To curate is to select and organize, to filter abundance into a collection of manageable size, one that in its smaller shape fulfils an informational or strategic need more efficiently than the buzzing flow of all available options” (Thorson & Wells 2016).

Curation is performed by a variety of actors through a variety of methods. Many discussions focus on the role of dominant online platforms, who curate content primarily through algorithmic features such as search, ranking and recommendation (e.g. Van Couvering 2009, Helberger, Kleinen-Von Königslöw & Van Der Noll 2015). Broader understandings of curation also recognize the role of others including individual users, advertisers, content providers in shaping online information flows (e.g. Thorson & Wells 2017; Napoli 2019). For instance, users can interact with **recommender systems** through rating and sharing content, and also have their own means to disseminate content through other channels; whereas content providers source the pool of available content from which rankings and recommendations are surfaced.

Content curation is closely connected to, though distinct from, **content moderation**. They can be seen as two sides of the same coin: Moderation speaks to the combatting of undesired content, whereas curation speaks to the surfacing of desired content. Accordingly, moderation is more associated with the removal of content or the sanctioning of users, whereas curation is associated with policies related to the design of search, recommender and ranking systems. This being said, content moderation can also be effectuated through curation systems, e.g. by down-ranking content or speakers. In this light, the design of ranking algorithms implicates both content moderation and content curation. Indeed, the zero-sum nature of ranking, in which advantaging certain content necessarily disadvantages other content, makes it so that any act of content curation in recommender systems can also be seen as a form of content moderation, and vice-versa.

(II) Content curation is not a legal concept, and it has not yet made its way into legislation or case law. However, content curation has received increasing attention in internet policy debates, reflecting a growing recognition that platforms influence online ecosystems not merely by enforcing content prohibitions but more fundamentally by structuring content visibility. Influential reports on this topic have been issued by the World Wide Web Foundation and the UN Special Rapporteur on Freedom of Expression, amongst others. (World Wide Web Foundation 2019) New legal standards are also developing to regulate platform content recommender systems, as a particularly influential form of content curation (Cobbe & Singh 2019). Key examples include the EU’s Platform-To-Business Regulation and Germany’s pending *Medienstaatsvertrag*. Ranking systems are also subject to other regulations which constrain curation, such as delisting rights found in data protection law, and the abuse of a dominant position under competition law.

References

Ávila, Renata, Juan Ortiz Freuler and Craig Fagan. ‘The Invisible Curation of Content: Facebook’s News Feed and our Information Diets.’ [2018] World Wide Web Foundation. Available at: http://webfoundation.org/docs/2018/04/WF_InvisibleCurationContent_Screen_AW.pdf

Cobbe, Jennifer and Jatinder Singh. ‘Regulating Recommending: Motivations, Considerations and Principles”. [2019] European Journal of Law and Technology 10(3).

Helberger, Natali, Kleinen-Von Königslöw, Katharin & Van der Noll, Rob, ‘Regulating the new information intermediaries as gatekeepers of information diversity’, [2015]

Napoli, Philip. ‘Social Media and the Public Interest: Governance of News Platforms in the Realm of Individual and Algorithmic Gatekeepers’, [2015] Telecommunications Policy 39(1).

Napoli, Philip. ‘Social Media and the Public Interest: Media Regulation in the Disinformation Age’. [2019] New York: Columbia University Press.

Thorson, Kjerstin and Chris Wells. ‘Curated Flows: A Framework for Mapping Media Exposure in the Digital Age’. [2016] Communication Theory 26(3).

Van Couvering, Elizabeth. ‘Search engine bias: the structuration of traffic on the World-Wide Web’. [2010] PhD Thesis, The London School of Economics. Available at: http://etheses.lse.ac.uk/41/

19. Dark patterns

This entry discusses: (I) the notion of dark pattern, its history and evolution; (II) a taxonomy of dark patterns; (III) existing regulatory and consumer advocacy responses to dark patterns.

(I) A "dark pattern" is a user interface design choice that benefits an online service by coercing, steering, or deceiving users into making unintended and potentially harmful decisions (Mathur et al. 2019). The expression was coined in 2010 by online designer Harry Brignull, who created an online guide and repository of cases (darkpatterns.org, currently maintained by Alexandre Darlington) and referred to a "user interface that has been carefully crafted to trick users into doing things, such as buying or signing up for things". It should be noted that this early definition included the three central elements of deceptiveness, deliberateness and accomplishment of the deceptive purpose. Later definitions refined the concept following a less deterministic approach, which dispensed with the requirements of specific intent and of specific effects of deception on users: for instance, Mathur et al. (2019) refer more generally to "*benefiting an online service* by coercing, steering or deceiving" (emphasis added) and Luguri and Strahilevitz (2019) to "user interfaces whose designers knowingly confuse users, *make it difficult for users to express their actual preferences*, or manipulate users into taking certain actions" (emphasis added). The term is also closely linked to the literature on malicious interface design techniques, defined as "interfaces that manipulate, exploit or attack users" (Conti and Sobiesk 2010); and to the broader concept of nudging, defined as "influencing choice without limiting the choice set or making alternatives appreciably more costly in terms of time, trouble, social sanctions, and so forth" (Hausmann and Welch 2010). The distinctive element however, common to all the existing definitions, is the covert and insidious nature of dark patterns, which in certain cases may fall into legally actionable fraud, unfair commercial practices or other violations of consumer and data protection rules (see section III below).

(II) Existing literature has broken down dark patterns into different categories. The most complete taxonomy to date has been offered by Luguri and Strahilevitz (2019), who have reviewed existing taxonomies and identified 7 general categories, each divided into types or "variants", for a total of 17 types of dark pattern. The following is the list of categories and their corresponding types:

- Nagging, which includes only one type, and is constituted by "Repeated requests to do something the firm [as opposed to the user] prefers";
- Social Proof, including "Activity Message" (Informing the user about the activity on the website, e.g., purchases, views, visits), "Testimonials" (Testimonials on a product page whose origin is unclear);
- Obstruction, including "Roach Motel" (Asymmetry between signing up and canceling), "Price Comparison Prevention" (Frustrating comparison shopping), "Intermediate Currency" (Set purchases in virtual currency to obscure cost);
- Sneaking, including "Sneak into Basket"(Adding additional products to users' shop- ping carts without their consent), "Hidden Costs" (Revealing previously undisclosed charges to users right before they make a purchase) and "Hidden Subscription" (Charging users for

unanticipated / undesired automatic renewal), "Bate and Switch" (Customer sold something other than what's originally advertised);

- Interface interference, including "Hidden Information / Aesthetic Manipulation / False Hierarchy" (Visually obscuring important information), "Pre-selection" (Pre-selecting firm-friendly default), "Toying with Emotion" (Emotionally manipulative framing), "Trick Questions" (Intentional or obvious ambiguity), "Disguised Ad" (Inducing consumers to click on something that isn't apparent ad) and "Confirmshaming" (Framing choice in way that seems dishonest / stupid);
- Forced Action, including "Forced Registration" (Tricking consumers into thinking registration necessary);
- Urgency, including "Low stock / high-demand message" (Falsely informing consumers of limited quantities) and "Countdown Timer" (Giving a message that an opportunity ends soon with a blatant false visual cue).

Domain-specific dark patterns have also been identified, sometimes creating new categories or types. For instance, in the privacy field Bösch et al. (2016) added "Hidden Legalese Stipulations" (hiding malicious information in lengthy terms and conditions) and the French Data Protection Authority identified a range of actions interfering with privacy choices from the perspective of "pushing the individual to accept sharing more than what is strictly necessary", "influencing consent", "creating frictions with data protection actions" and "diverting the individual"(Commission nationale de l'informatique et des libertés, 2019); while in the context of users´spatial relationship with digital devices Greenberg et al. (2014) introduced "Captive Audience" (taking advantage of users' need to be in a particular location or do a particular activity to insert an unrelated interaction) and "Attention Grabber" (visual effects that compete for users' attention).

(III) As dark patterns may constitute a violation of existing legal rules, some specific guidance has been recently issued by regulators in the consumer protection (Authority for Consumers and Markets, 2020) and data protection field (CNIL, 2019). Furthermore, consumer organizations have published reports finding problematic use of dark patterns with regard to data collection (Norwegian Consumer Council, 2018; Transatlantic Consumer Dialogue and Heinrich Böll Stiftung, 2020) and academic studies have been conducted to demonstrate the influence of dark patterns on the compliance with GDPR requirements for a valid consent (Nowens et al., 2020). These guidance documents and reports highlight the possible liability arising from dark patterns in relation to misleading and aggressive commercial practices, the violation of privacy by design and the rules on free, informed and specific consent. They also note the insufficiency of self-regulation, which by contrast is a central feature of a legislative bill (the DETOUR Act) introduced into the US Senate in 2019 by Senator Mark Warren to prohibit large online platforms from using deceptive user interfaces, known as "dark patterns" to trick consumers into handing over their personal data. The bill would entrust an industry association with the formulation of guidelines, and even a safe harbor against enforcement by the Federal Trade Commission, for design practices of large online platforms.

References

Authority for Consumers and Markets (ACM), 'Guidelines on the Protection of the online consumer. Boundaries of online persuasion' [2020]. Available at <https://www.acm.nl/sites/default/files/documents/2020-02/acm-guidelines-on-the-protection-of-the-online-consumer.pdf>

Commission nationale de l'informatique et des libertés (CNIL).'Shaping Choices in the Digital World From dark patterns to data protection: the influence of ux/ui design on user empowerment'. [November 2019] IP Reports Innovation and Foresight N°06 Available at <https://linc.cnil.fr/sites/default/files/atoms/files/cnil_ip_report_06_shaping_choices_in_the_digital_world.pdf>

Bösch, Christoph, Benjamin Erb, Frank Kargl, Henning Kopp, and Stefan Pfattheicher. 'Tales from the dark side: Privacy dark strategies and privacy dark patterns'. [2016] Proceedings on Privacy Enhancing Technologies 237, 254. Available at <https://petsymposium.org/2016/files/papers/Tales_from_the_Dark_Side_Privacy_Dark_Strategies_and_Privacy_Dark_Patterns.pdf>

Conti, Gregory and Edward Sobiesk. 'Malicious interface design: exploiting the user', [2010] Proceedings of the 19th international conference on World wide web 271, 280. Available at <https://doi.org/10.1145/1772690.1772719>

Greenberg, Sault, Sebastian Boring, Jo Vermeulen, and Jakub Dostal. 'Dark Patterns in Proxemic Interactions: A Critical Perspective'. [2014] P*roceedings of the 2014 Conference on Designing Interactive Systems (DIS '14).* ACM, New York 523, 532. Available at https://doi.org/10.1145/2598510.2598541

Hausmann, Daniel and Brynn Welch, 'Debate: To Nudge or Not to Nudge', [2010] 18 Journal of Political Philosophy 123, 136. Available at https://doi.org/10.1111/j.1467-9760.2009.00351.x

Luguri, Jamie and Lior Strahilevitz, ´Shining a Light on Dark Patterns' [August 1, 2019]. *University of Chicago, Public Law Working Paper* No. 719; *University of Chicago Coase-Sandor Institute for Law & Economics Research Paper* No. 879. Available at SSRN: https://ssrn.com/abstract=3431205

Mathur, Arunesh, Gunes Acar, Michael J. Friedman, Elena Lucherini, Jonathan Mayer, Marshini Chetty, Arvnid Narayanan. 'Dark Patterns at Scale: Findings from a Crawl of 11K Shopping Websites'. [July 17, 2019 working paper]. Available at https://arxiv.org/abs/1907.07032

Norwegian Consumer Council. 'Deceived by Design: How tech companies use dark patterns to discourage us from exercising our rights to privacy´. [2018] Available at https://fil.forbrukerradet.no/wp-content/uploads/2018/06/2018-06-27-deceived-by-design-final.pdf

Midas, Nouwens, Iaria Liccardi, Michael Veale, David Karger, Lalana Kagal. ‘Dark Patterns after the GDPR: Scraping Consent Pop-ups and Demonstrating their Influence’, [2020] *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* 1, 13. Available at https://arxiv.org/abs/2001.02479

Transatlantic Consumer Dialogue and Heinrich Böll Stiftung. ´Privacy in the EU and US: Consumer experiences across three global platforms´ [2020]. Available at https://eu.boell.org/en/2019/12/11/privacy-eu-and-us-consumer-experiences-across-three-global-platforms

20. Data Portability

Data portability is defined as “the right [of a person] to receive the personal data concerning him or her, which he or she has provided to a controller, in a structured, commonly used and machine-readable format”. This definition is contained in Article 20 of the GDPR, which was the first legislative source establishing such right.

As per the Article 29 Working Party Guidelines, the right only covers data that were provided to the controller by the user, but also includes data acquired by the controller by observing the user’s behaviour, such as activity logs; it does not include further elaborations, such as inferred or derived data.

The Guidelines identify also negative conditions for the exercise of this right, in particular that (1) it does not concern processing necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller; (2) its exercise is of no prejudice to the right to erasure provided by article 17 GDPR; and (3) it does not adversely affect the rights and freedoms of others. Of these negative conditions, the third is the most open-ended and uncertain from a business perspective, particularly considering that personal data that is subject to the request may simultaneously involve personal data of third parties (“networked data”). The Article 29 Working Party gives the example of a directory of data subject’s contacts, suggesting that the data controller can only accept to process such requests to the extent that there is a valid legal basis, for example a legitimate interest, which could be met if the new data controller was to provide a service allowing the data subject to process his personal data for purely personal or household activities. This interpretation, which presumes that the original data controller obtains specific and sufficiently reassuring information about the subsequent use of the received data, seeks to protect the data protection of third parties, which could otherwise be seriously affected by a too broad interpretation of the right to data portability. At the same time, the reference to “private or household uses” is also a safeguard against possible effects on competition derived from a strategic use of the right to data portability in order to gather commercial value from third party data.

Aside from the specific instance of networked data, other concrete possibilities of conflict may arise between the right to data portability and the rights or freedom of third parties. The A29 WP Guidelines merely mention one of these possibilities, specifically the tension with intellectual property or trade secrets, recalling one of the Recitals of the GDPR according to which “the result of those considerations should not be a refusal to provide *all* information to the data subject”. This is certainly not an exhaustive indication of how such conflicts should be resolved, but provides a hint that one-sided solutions (e.g., absolute refusal in deference to trade secrets) would not be acceptable. It can thus be expected that a data controller takes reasonable measures to provide as much information requested as possible by decontextualizing personal data from proprietary algorithms or trade secrets. This arguably won’t be an issue as far as *provided* data is concerned, since such data does not reveal anything about the inner working of the systems used to store and analyze them. On the other hand, intellectual property and trade secrets may present some challenges when it comes to *observed* data, which can be difficult to disentangle from the categories designed by the controller to process the data inputs. Even in cases where de-contextualization is not feasible, however, the fact the data is transferred onto the user or a different data controller does not as such imply that the underlying intellectual property will necessarily be violated: the data subject and the second controller surely bear liability for any illegitimate processing of those data. This somewhat cynical understanding appears reflected in the statement by the Article 29 WP Guidelines that "a potential business risk cannot, in and of itself, serve as the basis for a refusal to answer the portability request". Yet it obviously raises a question of what is the threshold of substantiation of a risk, such that they entitle a data controller-right holder to prevent future infringements of IP rights in the context of data portability requests. This is a matter largely left open to future guidelines (by the EU Data Protection Board) and legislation, with Recital 73 of the GDPR offering examples and stressing the need for any restrictions to data portability to be in compliance with the EU Charter of Fundamental Rights and the European Convention of Human Rights.

Data portability is one of the rights that are meant to give individuals control over their data. Its purpose is also to allow the individual to switch to a different service provider without having to provide all their information again. Thus, Article 20 of the GDPR also foresees the right to transmit the data to another controller, automatically *"if technically feasible"*. This would enable more choice and more competition in digital service markets, similar to what happened when number portability was introduced in the mobile telephony market. However, the Internet industry has not enthusiastically embraced the concept, and the implementation of the data portability right mostly remains limited to exporting the user's data into a file, while *"technical solutions for standardised data exchange remain in their infancy" (CERRE, 2020).*

References

Regulation (EU) 2016/679 of the European Parliament and of the Council (EC) 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 [27 April 2016]. Available at: < https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32016R0679>

Working Party , 'Guidelines on the right to data portability' [2016]. Available at: < https://ec.europa.eu/newsroom/document.cfm?doc_id=44099>

Jan Krämer, Pierre Senellart & Alexandre de Streel, ' Making data portability more effective for the digital economy:' (Centre on Regulation in Europe 2020) < https://cerre.eu/wp-content/uploads/2020/07/cerre_making_data_portability_more_effective_for_the_digital_economy_june2020.pdf>

Gianclaudio Maglieri, 'Trade Secrets v Personal Data: a possible solution for balancing rights' 2016] International Data Privacy Law 6 (2) 102, 116.

Inge Graef, Martin Husovec and Nadezhda Purtova, 'Data Portability and Data Control: Lessons for an Emerging Concept in EU Law' [December 15, 2017]. TILEC Discussion Paper No. 2017-041; Tilburg Law School Research Paper No. 2017/22. Available at SSRN: <https://ssrn.com/abstract=3071875 or http://dx.doi.org/10.2139/ssrn.3071875> accessed 10 July 2018.

21. Defamation

This entry: (i) sets out guidance on how the term “defamation” is u nderstood and (ii) provides examples of existing regulatory responses to defamation.

(i) Guidance on understanding the term “defamation”

Defamation is prohibited in the majority of states around the world (see below), however there is no universally accepted definition of the term. Definitions largely coalesce around the communication of a statement (ordinarly false) about another person that unjustly harms their reputation. While it is beyond the scope of this glossary to seek to provide a definitive definition of “defamation”, there are two critical considerations for policymakers seeking to address online manifestations of defamation.

First, it is unlikely that a distinct and separate definition of defamation when it takes place online will be necessary. Instead, existing definitions of defamation should be reviewed to ensure that they apply to all forms of defamation, whether offline or online. There should not be different legal processes, sanctions or remedies relating to defamation depending on whether it took place offline or online.

Second, any definitions of defamation should be consistent with international human rights standards, particularly the right to freedom of expression, as set out in relevant guidance. The then UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Ambeyi Ligabo, provided particularly useful guidance in 2007, stating that:

“A statement can be considered as defamatory under certain specific conditions: it must be published, in a spoken, written, pictured or gestured form. Written and pictured statements, which include drawings, video clips, and movies and so on, are considered more serious offences as they last longer than mere verbal statements, which are generally defined as slander. The statement must be false, in the sense that its contents should be totally untrue; it has to be injurious - there is no defamation without injury - and finally, unprivileged, in the sense that certain categories of individuals cannot be sued while making statements, especially in their professional capacity. Last but not least, a statement can be considered as defamatory if done with actual malice, which means that there was a real willingness to harm the defamed person.” (Ligabo, 2007).

There is also a strong consensus that defamation should not be a criminal offence, but dealt with under civil law (Ligabo, 2007 and UN, 2011).

(ii) Existing regulatory responses

As noted above, defamation is prohibited in the majority of states around the world. Often this is done via civil law provisions which allow individuals to bring legal proceedings against those that have defamed them, and to seek damages or some other remedy for the harm caused. While considered to be inconsistent with international human rights law and standards, some states also have provisions in their criminal laws prohibiting defamation, thus enabling individuals to be prosecuted and punished for defamation.

While the prohibitions of defamation through civil and/or criminal law mean that legal persons who publish defamatory statements, such as newspapers, can be held liable, a small number of states also allow for online platforms to be held liable for defamatory statements posted or shared by third parties on those platforms. In some states, the liability exists at the point that the defamatory statement is posted; in others, it only arises once the platform has become aware of the defamatory statement and fails to remove it within a reasonable period of time.

In at least two European states, Estonia and Hungary, online platforms have brought cases to the European Court of Human Rights arguing that holding them liable for the defamatory statements of third parties constituted a violation of the right to freedom of expression. In one of the cases, Delf v. Estonia (Application no. 64569/09), the court held that there was no violation on the basis that the comments were clearly unlawful, that the platform professionally managed the comments section of its website where the comments were made, and that the platform took insufficient measures to remove the comments without delay. In the other case, Magyar Tartalomszolgáltatók Egyesülete and Index.hu Zrt v. Hungary (Application no. 22947/13), the court held that there had been a violation on the basis that the comments weren't clearly unlawful, the platform was not operated on a commercial basis, and the platform took general measures to prevent defamatory comments.

References

Ambeyi Ligabo, UN Human Rights Council, 'Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression', UN Doc. A/HRC/4/27, [2 January 2007], Paragraph 47.

ARTICLE 19, 'Defining Defamation: Principles on Freedom of Expression and Protection of Reputation', [2017], available at: https://www.article19.org/data/files/medialibrary/38641/Defamation-Principles-(online)-.pdf.

UN Human Rights Committee, 'General comment No. 34: Article 19: Freedoms of opinion and expression, UN Doc. CCPR/C/CG/34', [12 September 2011], Paragraph 47.

22. Deindexing

This term refers to the intentional removal and unintentional removal of results from search engines and/or indexing websites. A website can be deindexed using robot.txt, which can be implemented to prevent other sites from crawling pages or sites, or manually delisted. Search engines can implement themselves to remove or reduce the visibility of unwanted or low quality results, such as spam or “clickbait.” There are paid services that can be hired to remove unwanted results from search engines, such as reputation management companies. Governments have required the deindexing of specific categories of information. For example, through the “Right to be Forgotten” which allows individuals to request that their personal data be removed from search engine results, amounting to deindexing through the delisting of results.

A more technical understanding of “deindexing” refers to the de-indicization of a page from the search engine crawlers, which removes the need for a presentation of a “sanitized version” of the results in the first place. This can be obtained by websites voluntarily by using robot.txt files, which convey that specific message to search engines. However, it is more complex to accomplish when the information should only be removed from search engines in connection with a particular keyword, as that cannot be accomplished (at least for the time being) without human intervention.

23. Demonetization

Demonetization refers to a unilateral action in the form of a sanction that a platform (traditionally a social media platform) takes in order to remove a creator/influencer's access to the streams of revenue the platform controls. As indicated under two other relevant entries ('Content creators/influencers' and 'Content/web monetization'), one of the business models through which users can make money on platforms such as Youtube or Instagram (less so Facebook) is the monetization of their content through platform-specific programmes, allowing advertisers to display ads in various forms (e.g. banners, multimedia clips) throughout or over their content (Goanta & Ranchordás, 2020). Compared to deplatforming, which entails the removal of a user, demonetization is a less stringent sanction, as it only removes potential income for particular videos. Demonetization can take place in two ways: income can be completely removed, or it can be redirected. The latter situation can occur when a creator receives a copystrike, and the claimant of the copystrike has copyright over material used by the creator. This way, the claimaint of the copystrike can ask for the ad-generated income on the video of the creator infringing their copyright.

Demonetization is closely related to content moderation, because it is a way in which platforms control content. However, due to the inherent characteristics of private governance, such as the lack of transparent criteria for the interpretation of community guidelines in determining which content is in contravention to these rules, platform discretion is a serious problem in the content creator community (Caplan & Gillespie, 2020; Lobato, 2016).

Additional problems with demonetization concern the content creators´ rights to due process and to an effective remedy . For example, although an appeal process is available on Youtube, this does not compensate for the loss of revenue that occurs when consent creators are deprived of monetization in the first hours after publication, which are likely to be the most remunerative ones(Caplan & Gillespie, 2020), and sometimes even for months (Koi, 2020). Further, it has been noted that Youtube´s policy establishes that only those creators who have at least 1000 video views in a week or 10000 channel subscribers can request a re-evaluation of demonetization by a human being. This has been criticized as establishing a "tiered governance" system (Caplan & Gillespie, 2020), where the rules governing the relationship with creators are different depending on their monetization potential. While this may be economically sensible, it is in conflict with the universal and unwaivable nature of fundamental rights.

References

Catalina Goanta & Sofia Ranchordás, 'The Regulation of Social Media Influencers', [2020] Cheltengam: Edward Elgar.

Caplan, R., & Gillespie, T. 'Tiered Governance and Demonetization: The Shifting Terms of Labor and Compensation in the Platform Economy'. [2020] Social Media + Society.

Lobato, R. ‘The cultural logic of digital intermediaries: YouTube multichannel networks’. [2016] Convergence: The International Journal of Research into New Media Technologies, 22(4) 348, 360.

Koi, C. ‘Wrongfully demonetized, how many months without revenue on average? - YouTube Community (7/2/20)’, [2020] Available at https://support.google.com/youtube/thread/56781687?hl=en

24. Deplatforming

Deplatforming refers to the ejection of a user from a specific technology platform by closing their accounts, banning them, or blocking them from using the platform or its services. It is worth nothing that deplatforming may be permanent or temporary. Temporary suspensions and impossibility to access one's account can be considered as deplatforming.

Deplatforming is an extreme form of content moderation and a form of punishment for violations of acceptable behavior as determined by the platform's terms or service or community guidelines. Platforms justify the removal or banning of a user and/or their content based on violations of its terms of service, thereby denying the user access to the community or service that it offers. Deplatforming can and does occur across a range of platforms and can refer to:

- Social media companies, like Facebook, YouTub or Twitter;
- Commerce platforms such as Amazon of the Apple iStore;
- Payment platforms, like PayPal or Visa;
- Service platforms, like Spotify or Stitcher;
- Internet infrastructure services like Cloudflare or web hosting.

Deplatforming can be a form of content moderation by tech platforms that find certain content objectionable, or face public pressure to restrict a user's access to the platform, often as a result of the content of that person's speech or ideas, or in response to harassing behavior. It has also been deployed to reduce online harassment, hate speech, and coordinated inauthentic behaviour, such as propaganda campaigns. Deplatforming can also occur because of pressure from other platforms, at the same or in different levels of the stack.

Because deplatforming can refer to a range of platforms, and is often implemented as a form content moderation, this approach by platforms that do not host content, or which are further down the internet stack, raises concerns about the expansion of content-based censorship beyond content-hosting services or platforms. The review of accounts and content can be automated or the result of human review, of a combination of both.

The term is explicitly political because it often refers to banning a user from a platform because of the content of their speech and ideas. Deplatforming has been used as a response to hate speech, terrorist content, and disinformation/propaganda. For example, the major social media firms have removed hundreds of ISIS accounts since 2015, seeking to reduce the UN-designated terrorist group's reach online, which forced them onto less public and more closed platforms, reducing their visibility and public outreach, but also making it more difficult to monitor their activities. In 2018, Facebook and Istagram deplatformed (Facebook, 2018) the Myanmar (Facebook, 2018) military after it was involved in the genocide of Rohingya, closing hundreds of pages and accounts related to the military and banning several affiliated users and organizations from its services.

Deplatforming by dominant platforms has pushed extremists to less popular or less public platforms that offer an alternative set of rules or have not yet grappled with what their rules should be. Deplatforming has given rise to alternative platforms, such as the social media site Gab, the crowd-funding site Patreon and the messaging service Telegram.

Several platforms shut down and banned Alex Jones and Infowars from their platforms in mid-2018 in response to their support for white supremacy and involvement in disinformation campaigns, which helped politicize and publicize the concept of deplatforming.

De-platforming can reduce the ability to inject a message into public discourse and recruit followers, but it can also push supporters to obscure and opaque platforms where it is substantially more difficult for law enforcement to monitor their activities. One rationale for deplatforming controversial people or organizations is to prevent them from negatively influencing others. Critics argue this is an ineffective tactic because the affected person will just go to another platform, but a Georgia Tech study found (Chandrasekharan et al., 2017) that deplatforming was an effective moderation strategy that reduced the unwanted speech or behavior, and created a demonstration effect for other users that helped enforce norms. In response to takedowns on major platforms, extremists often migrate to lesser-known or protected online forums. Research has also shown that people who are deplatformed often fail to transfer audiences from major to minor platforms. Researchers refer to the "online extremists' dilemma" which describes (Clifford & Powell, 2019) how online extremists are forced to balance public outreach and operational security in choosing which digital tools to utilize.

References

Facebook, ' Removing Myanmar Military Officials From Facebook' (Facebook 28 Aug 2018) < https://about.fb.com/news/2018/08/removing-myanmar-officials/>

Facebook, 'Update on Myanmar' (Facebook 15 Aug 2018) <https://about.fb.com/news/2018/08/update-on-myanmar/>

Eshwar Chandrasekharan, Umashanthi Pavalanathan, Anirudh Srinivasan, Adam Glynn, Jacob Eisenstein, Eric Gilbert, ' You Can't Stay Here: The Efficacy of Reddit's 2015 Ban Examined Through Hate Speech' [2017] Proc. ACM Hum.-Comput. Interact., Vol. 1, No. 2, Article 31 31, 31:22

Bennett Clifford, Helen Christy Powell, ' De-platforming and the Online Extremist's Dilemma' (Lawfare 2019) < https://www.lawfareblog.com/de-platforming-and-online-extremists-dilemma>

25. Device Neutrality

Device neutrality ensures the users right of non-discrimination in the services and apps they use, based on platform control by hardware companies (Hermes Center, 2017; ARCEP, 2018). That means users can have the possibility to choose which operating system (software) or application they prefer to use, regardless of the brand of device they are using. In other words, device neutrality is instrumental to achieve equal access to applications, contents and services which is essential to achieve an open Internet. This is a fundamental civil rights issue as it ensures that the user has the right and possibility to use, for example, the information and communication security tools they prefer on their devices. It is usually framed as a consumer protection rather than a technical measure because it enables citizens to fight against aggressive or deceptive commercial practices that limit their use of applications and unfairly favour their own content or demote that of competitors.

The device neutrality argument defends that consumers have the right to uninstall softwares and to remove apps and content they are not interested in, which at the moment is impossible because companies don’t give the right to remove default applications. It also defends the possibility that all content and service developers can access the same device function. The French Telecommunications Regulator ARCEP (2018) also advocates for device neutrality, stressing the need for more transparency in app store rankings and easier access to applications offered by alternative apps stores.

This concept was first introduced in a legislative proposal in Italy in 2014 by MP Stefano Quintarelli, who proposed a bill called “*S.2484 Disposizioni in materia di fornitura dei servizi della rete internet per la tutela della concorrenza e della libertà di accesso degli utenti*”. The bill was approved at the Chamber of Deputy and it is still waiting to be voted in Senate, and it addresses Device Neutrality in Article 4 stating that that: “Users have the right to find online, in a format suitable to the desidered technology platform, and to use in fair and non-discriminatory ways software, proprietary or open source, contents and legitimate services of their choice”.

Concerning the correlation between the concept of Device Neutrality and Net Neutrality rights, the first one can be seen as a extension of the second, mostly because they both reinforce the principle of “innovation without permission”, which means that anyone, anywhere, can create and reach an audience without anyone standing in the way (Kak & Ben-Avie, 2018). In a similar way Device Neutrality defends that users have right to non discrimination of the services or apps in their devices regardless the hardware companies, Net Neutrality defends the right to non discrimination by Internet service providers, regardless of the content or applications utilised by Internet users, unless such discriminatory tretment is necessary and proportionate to the achievement of a legitimate aim (Belli & De Filippi, 2015; Belli, 2017) . One’s about equal access to applications and the other about equal access to the Internet.

Another set of rules that could possibly go in the opposite direction to that of device neutrality are the anti-circumvention laws, which provides penalties for those who wish to make changes on their devices and operational systems. At first, these rules were created with the objective of protecting intellectual works from copyright infringement, but have proved useless over the years, only harming competition, innovation, freedom of expression and scientific research (Doctorow, 2019; EFF, s.d.).

References

ARCEP. 'Devices, The Weak Link in Achieving an Open Internet, Report on their limitations and proposals for corrective measures'. [2018] Available at: <https://www.arcep.fr/uploads/tx_gspublication/rapport-terminaux-fev2018-ENG.pdf>

Luca Belli & Primavera De Filippi (Eds.) 'Net Neutrality Compendium. Human Rights, Free Competition and the Future of the Internet'. [2015] Springer Available at: ttps://www.ohchr.org/Documents/Issues/Expression/Telecommunications/LucaBelli.pdf

Luca Belli. 'Net Neutrality, Zero-rating and the Minitelisation of the Internet'. [2017] Journal of Cyber Policy. Routledge. Vol 2, nº 1. 96, 122

Borchert, Katharina. 'EU Copyright Law Undermines Innovation and Creativity on the Internet'. [2016] Available at: <(https://medium.com/mozilla-internet-citizen/eu-copyright-law-undermines-innovation-and-creativity-on-the-internet-6ef65147809d#:~:text=A%20key%20part%20of%20what,anyone%20standing%20in%20the%20way>

Doctorow, Cory. 'Bird Scooter tried to censor my Boing Boing post with a legal threat that's so stupid, it's a whole new kind of wrong'. [2019] Available at: <https://boingboing.net/2019/01/11/flipping-the-bird.html>

Electronic Frontier Foundation (EFF) (n.d.). 'Digital Millennium Copyright Act.' Available at: <https://www.eff.org/issues/dmca>

Hermes Center. 'After Net Neutrality, Device Neutrality', [2017] Available at: <https://www.hermescenter.org/net-neutrality-device-neutrality/>

Italian Parliament. 'S.2484 Disposizioni in materia di fornitura dei servizi della rete internet per la tutela della concorrenza e della libertà di accesso degli utenti', [2016] Available at: <https://parlamento17.openpolis.it/atto/documento/id/255634>

Kak, Amba U. & Ben-Avie, Jochai. 'ARCEP report: "Device neutrality" and the open internet', [2018] Available at: <https://blog.mozilla.org/netpolicy/2018/05/29/arcep-report-device-neutrality/>

26. Digital Rights

In a framework of "digital citizenship" (see generally eg Ribble 2011), "[b]eing a full member in a digital society means that each user is afforded certain rights, and these rights should be provided equally to all members." (id, 35). Digital rights are, in such a general sense, connected to "boundaries" [which] "may come in the form of legal rules or regulations, or as acceptable use policies" (id). Therefore, one of the key related terms is **responsibility**, which also points to the idea that "those who partake in the digital society would work together to determine an appropriate-use framework acceptable to all" (id).

The term "digital rights" is a concept that has gained recognition through an evolving interpretation of rights recognized by governments all over the world. It is worth noting there is a conspicuous lack of l definition of the term. In fact, it is not specifically referenced in the definitions provided by core documents of legal doctrine and policy relevant for the field of platform law and policy, particularly in treaty law, international regulations or national Constitutions. However, claims are progressively being brought in front of the courts or raised in political debates emphasizing the importance of the digital environment.

At the same time, any general search for the term outside these arenas of legal debate easily shows that the term gained major significance in recent times. As the term is widely used in common parlance as well as in all kinds of internet policy debates, advocacy, and legal practice, it is beyond the scope of our definition to cover all the rights that have been addressed in the above-mentioned law and policy context.

A common trait in its usage is the emphasis on the internet's impact on everyday life in our societies on a global scale, through the pervasiveness of online human interaction nowadays. When striving for a possible recognition by the institutions normally guaranteeing fundamental rights, digital rights would be said to act as corollaries of other rights that are available. For example, the UN Human Rights Council in several bi-annual resolutions (2012, 2014, 2016, 2018) has (re-)affirmed "that the same rights that people have offline must also be protected online, in particular freedom of expression".

Any progress towards realizing the idea of universal (**digital) access** would impact the idea of "digital rights". The ever-growing pervasiveness of digital technologies can also lead to changing the social contract between societal actors and delineating a new understanding around rights, values, responsibilities and entitlements for the benefit of all in a digital society (see generally, eg Vesnic-Alujevic et al 2019).

The current digital divide(s) are, therefore, a major policy issue relevant for the digital rights-campaigns. In this regard, it seems important to have multi-level provisions that explicitly entrench the new digital rights, such as **digital access**. Besides, existing fundamental rights are still key for the development of policies surrounding recommendations for platforms' **responsibility** towards rights-holders.

References

Mike Ribble. ‘Digital Citizenship in Schools, Second edition’, [2011] Washington, DC: International Society for Technology in Education.

United Nations General Assembly, Human Rights Council Twentieth Session, 20/L13… The Promotion, Protection and Enjoyment of Human Rights on the Internet, A/HRC/20/L.13 (June 29, 2012) (+ 2014: https://ap.ohchr.org/documents/dpage_e.aspx?si=A/HRC/RES/26/13 , 2016: https://ap.ohchr.org/documents/dpage_e.aspx?si=A/HRC/RES/32/13 & 2018: https://ap.ohchr.org/documents/alldocs.aspx?doc_id=29960).

Vesnic-Alujevic et al. ‘The Future of Government 2030+: A Citizen Centric Perspective on New Government Models’, [2019] Available at: <https://ec.europa.eu/jrc/en/publication/future-government-2030>.

27. Disinformation

Defining this phenomenon has shown to be far from simple. Scholars from different fields have provided definitions of this phenomenon (Tandoc Jr et al. 2018). The information disorder has been defined as the mix of 'misinformation', 'disinformation' and 'malinformation' which respectively reflect increasing levels of harm and involve different content (Wardle and Derakhshan, 2017). False information would include information disseminated as intentionally false and impossible to verify to mislead the public (Allcott and Gentzkow, 2017). Adopting this definition would imply that only news disseminated with the intention to mislead readers would fall into the field of disinformation. Therefore, other (false) information outside the framework of intent could be considered free expressions of each one's thoughts. This could cover for instance information shared due to mistakes or satire as well as investigative journalism which does not base its findings on entirely truthful facts but on reconstructions of truth. According to the European Commission's High-Level Group on Fake News and Online Disinformation ('HLEG'), disinformation is "false, inaccurate, or misleading information designed, presented and promoted to intentionally cause public harm or for profit. The risk of harm includes threats to democratic political processes and values, which can specifically target a variety of sectors, such as health, science, education, finance and more" (HLEG, 2018). The expression "disinformation" has been considered a more adequate way to describe the spread of false content. Precisely, this situation is not just connected to (fake) news but also false or misleading content like fake accounts, videos and other fabricated media (Chesney and Citron, 2019). Moreover, the HLEG distinguishes the notion of disinformation from that of 'misinformation', i.e. 'misleading or inaccurate information shared by people who do not recognise it as such', and underlinesunderline that disinformation does not include illegal speech (e.g. hate speech).

Disinformation is not a phenomenon of the digital age. Its digital dissemination is the novelty. The digital dimension entails the worldwide reach of online content beyond territorial boundaries and media environment (Sunstein, 2017). The spread of false information online during the 2016 Brexit referendum and the US presidential election can be considered two paradigmatic examples of how disinformation influences internal politics, interfering with electoral processes and undermining trust in public institutions and the media. Besides, the pandemic has shown how disinformation can affect information quality and health on a global scale (Majó Vazquez et al., 2020).

The spread of false content can be understood by looking at the new media framework (Martens et al., 2018). The characteristics of media manipulation highlight how the media sector tend to gravitate toward sensationalism, the need for constant novelty, and the aim of achieving profits instead of professional ethical standards and civic responsibility (Marwick and Lewis, 2017). Digital spaces are perfect spaces for disseminating information at minimal or no cost. Within this framework, the role of online platforms, including social media, becomes critical to understand how false and misleading information spread online. The monetization of these expressions capturing users' attention and becoming viral highly depends on the algorithmic system pushing certain messages to the top and promoting further engagement. Unlike traditional media outlets, social media usually perform content moderation activities implementing automated systems which can decide in a heartbeat how information is organised online. Beyond media strategy to disseminate disinformation, scholars have emphasised the role of the political context. Indeed, the role of technology platforms, bots and foreign spies has tended to be overemphasised (Benkler, 2018). Political parties and, in particular, populism movements have relied on strategies of disinformation to support their political ideas (Bayer et al., 2019).

This framework shows why, before focusing on the challenges in addressing disinformation, it is worth defining the boundaries of false content considering the digital environment as its primary context. These definitions allow us to understand the multifaceted character of disinformation requiring public actors to face the complexities relating to the regulation of freedom of expression online to tackle this phenomenon. This is why dealing with disinformation means addressing the boundaries of the right to free speech, thus, involving democratic values (Pitruzzella and Pollicino, 2020). Indeed, tackling disinformation requires public actors to decide to what extent speech is protected and balanced with other constitutional rights and liberties, as well as how to pursue other (legitimate) interests. Nonetheless, the regulation of speech does not involve any longer just the States and the speaker, but also multiple players outside the control of the State, such as social media companies. In the information society, freedom of expression is like a triangle (Balkin, 2018). Therefore, due to the role of online platforms in this field, regulation should also take into account the effects of regulatory choices over the role and responsibilities of these actors.

From a policy perspective, different regulatory solutions have been adopted worldwide (Robinson et al. 2020; De Gregorio and Perotti, 2019). While the US has not proposed a precise strategy to deal with this phenomenon, the Union focused on soft law commitments by platforms, precisely the code of practice on disinformation, and *ad hoc* measures targeting the context of the European elections or (Pollicino et al., 2020). Domestic legislation of European states provides a highly fragmented regulatory picture (e.g. Germany, France). From a global perspective, other regulatory experiences have shown a tendency towards the criminalisation of disinformation (e.g. Singapore, Russia). This fragmentation does not only challenge the protection of the right to freedom of expression online but also undermines the principle of the rule of law rather than promote a clear regulatory framework to address this global phenomenon. For instance, vagueness about definitions and threshold of harm or illegality would negatively impact on the right to freedom of expression. Within this framework, the importance of judicial scrutiny of these regulatory measures could ensure a fair assessment of the case and mediation from an independent authority. The role of judicial authority to scrutinise measures to remove false content could contribute to safeguarding the right to freedom of expression against discretional decisions taken by platforms or non-independent public bodies. Besides, the promotion of fact-checking activities, supporting professional media outlets and investing resources for digital literacy campaigns could play a critical role (Ireton and Posetti 2018). Relying on these measures would entail a lower impact on freedom of expression while building the instruments to fight disinformation on a global scale.

References

Allcott, Hunt and Gentzkow, Matthew, ‘Social Media and Fake News in the 2016 Election’ [2017] 31(2) Journal of Economic Perspectives 211

Balkin, Jack M., ‘Free Speech is a Triangle’ [2018] 118 Columbia Law Review 2011

Bayer, Judith et al., ‘Disinformation and Propaganda—Impact on the Functioning of the Rule of Law in the EU and Its Member States’ [2019] Study for the LIBE Committee;

Benkler, Yochai et al., Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics (Oxford University Press 2018)

Chesney, Robert and Citron, Danielle, ‘Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security’ [2019] 107 California Law Review 1753

De Gregorio, Giovanni & Perotti, Elena, ‘Tackling Disinformation around the World’ [2019] <https://www.wan-ifra.org/reports/2019/05/03/public-affairs-media-policy-briefing-tackling-disinformation-around-the-world>.

Ireton, Cherilyn and Posetti, Julie (eds), ‘Journalism, ‘Fake News’ & Disinformation’ [2018] UNESCO <https://unesdoc.unesco.org/ark:/48223/pf0000265552>

Pollicino, Oreste et al., ‘The Regulatory Conundrum to Face the Raise and Amplification of False Content in Internet’ [2020] Global Community Yearbook of International Law and Jurisprudence, forthcoming

European Commission, ‘A Multi-Dimentional Approach to Disinformation. Final report of the High-Level Expert Group on Fake News and Online Disinformation’ [2018]

Majó-Vázquez, Silvia et al., ‘Volume and Patterns of Toxicity in Social Media Conversations during the COVID-19 Pandemic’ Reuters Institute [9 July 2020] <https://reutersinstitute.politics.ox.ac.uk/ volume-and-patterns-toxicity-social-media-conversations-during-covid-19-pandemic>

Martens, Bertins et al., ‘The Digital Transformation of News Media and the Rise of Disinformation and Fake News’ [2018] JRC Digital Economy Working Paper no. 2;

Marwick, Alice and Lewis, Rebecca, ‘Media Manipulation and Disinformation Online’ Data & Society (15 May 2017);

Robinson, Olga et al., ‘A Report on Antidisinformation Initiative’ [2019] <https://comprop.oii.ox.ac.uk/ wp-content/uploads/sites/93/2019/08/A-Report-of-Anti-Disinformation-Initiatives>;

Pitruzzella, Giovanni and Pollicino, Oreste, ‘Disinformaiton and Hate Speech. A European Constitutional Perspective’ (Bocconi University Press 2020);

Sunstein, Cass R., ‘#Republic: Divided Democracy in the Age of Social Media’ (Princeton University Press 2017);

Tandoc Jr, Edson C. et al., ‘Defining “Fake News”: A Typology of Scholarly Definitions’ [2018] 6(2) Digital Journalism 137;

Wardle C., and Derakhshan H., ‘Information Disorder: Towards an Interdisciplinary Framework for Research and Policy Making’ Council of Europe report DGI(2017)09 <https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-researc/168076277c>

28. Dispute resolution (online)

Online dispute resolution (ODR) is an umbrella term referring to out-of-court dispute resolution techniques which are offered using technological infrastructures, particularly on the internet (Thomson & Sherr, 2012). As a sub-category of alternative dispute resolution (ADR), ODR is a 'private independent but binding justice system' (Mistelis, 2006), by which parties to a dispute have access to procedures through which they can settle their differences (Hörnle, 2009).

Some authors argue that ODR systems should have essential properties such as simplicity (user-friendliness), adaptability (processes automatically adjusted to the needs of the parties) and interoperability (ensuring connectivity of stakeholders regardless of data architecture differences), in order to be properly integrated in Internet-based industries such as e-commerce (Kaufmann-Kohler & Schultz, 2004).

As a system of private justice, ODR has been historically focused on facilitating the solving of disputes between sellers and consumers, which is why one of the most cited case studies of successful ODR is eBay's Resolution Center (Del Duca, L.F. et al., 2014)). However, this success from early commercial activity on the Internet did not transfer smoothly to other categories of intermediation business models which occurred at later stages. For instance, originally, **sharing economy** platforms such as AirBnB or Uber would catalogue disputes between hosts and guests or drivers and passengers as customer care problems, often solved through the use of FAQs or automated forms which would provide certain forms of immediate or reviewed relief (e.g. the return of a deposit fee; the return of the charged ride rate). Especially in these cases, the limited platform support in even reaching the other contracting party before or after the completion of the transaction has amplified the pitfalls of the limited liability regimes information society services have been traditionally benefitting from. The same can be said for social media platforms, which provide the intermediation of media services as well as peer content, in an environment where toxicity and abusive language is abundant. This leads to various harms, some of which can be easily labelled legally (e.g. insult, defamation, incitement to hatred), and some of which are more difficult to pinpoint from the perspective of existing legal frameworks (e.g. swarm bullying or cancel culture). All in all, the rationale behind dispute resolution is that users are in search for justice (Citron & Jurecic, 2018; Katsh & Rabinovich-Einy, 2017), and when justice deals with the removal of content, it goes into the realm of content moderation, absent a framework for the resolution of disputes between peers, beyond systems for the reporting of abusive content, which, ironically, are often themselves abused (e.g. cancel culture). Some platforms may have content and reporting management centers (e.g. Youtube) for areas of activity where platform liability may arise in the absence of additional measures (e.g. copyright; Google, 2020). However, given the existing disconnect between real courts and Internet harms, as well as the lack of successful, scalable models for ODR across the various services offered by information society services, it can generally be said that ODR is a field still in its infancy.

References

Thomson, S. & Sherr, A, ' Definitions of Online Dispute Resolution' in Martin Gramatikov (eds), Costs and Quality of Online Dispute Resolution: A Handbook for Measuring the Costs and Quality of ODR (1st, Maklu, 2012).

Mistelis, L. ‘ADR in England and Wales: a successful case of public private partnership’, [2003] 6(3) ADR Bulletin.

Hörnle J, *Cross-Border Internet Dispute Resolution* (Cambridge University Press 2009)

Loebl, Z. *Designing Online Courts: The Future of Justice Is Open to All* (Kluwer 2019).

Del Duca, L.F. et al. ‘*eBay's De Facto Low Value High Volume Resolution Process: Lessons and Best Practices for ODR Systems Designers’*, 6 Yearbook of Arbitration and Mediation 204 (2014).

Tworek et al. ‘Dispute Resolution and Content Moderation: Fair, Accountable, Independent, Transparent, and Effective’, [2020] Transatlantic Working Group, available at < https://www.ivir.nl/publicaties/download/Dispute_Resolution_Content_Moderation_Final.pdf>.

Citron, D. & Jurecic, Q. ‘Platform Justice: Content Moderation at an Inflection Point’, [2018] Aegis Series Paper No. 1811, available at < https://www.lawfareblog.com/platform-justice-content-moderation-inflection-point>.

Katsh, E. & Rabinovich-Einy, O. *Digital Justice* (Oxford University Press 2017).

Google. ‘What is a Content ID claim?’, [2020] available at < https://support.google.com/youtube/answer/6013276?hl=en>.

29. Due diligence

In several legal fields, both in international law and in domestic legislation, the concept of due diligence corresponds to what a responsible entity – be it a state or a business enterprise – ought to do under normal conditions in a situation with its best practicable and available means, with a view to behave responsibly and fulfil its obligations. (Dupuy 1977:13) In this perspective, due diligence refers to a level of judgement, care, prudence and determination that an entity is reasonably expected to undertake under specific circumstances.

In some contexts, due diligence refers also to the process by which an entity interested in a specific activity entailing potential risks, such as a purchase or an investment, identifies, analyses and define how to manage such risks before entering in an agreement or transaction.

Hence, due diligence entails a range of analyses and considerations before performing given activities and/or during the performance, in order to prevent and mitigate risks that could determine harm.

In the field of Environmental Law, for example, due diligence signifies the conduct to be expected from a responsible stakeholder, in order to effectively protect other stakeholders and the global environment. (Dupuy 1977:3) Failure to exercise due diligence, therefore, means incapacity to fulfil the standard of conduct expected from a responsible stakeholder in the specific situation.

The International Law Commission considers due diligence as a primary environmental obligation of States. In Articles 3-7 of the Convention on the Prevention of Transboundary Harm from Hazardous Activities, for instance, four features of due diligence can be distinguished and applied, by analogy, to other fields:

- taking all appropriate measures to prevent and minimise the risk
- cooperating with other stakeholders
- implementing obligations through all necessary regulatory actions, including monitoring mechanisms
- a prior assessment of the possible external harm should be done before giving authorisation for an activity or a major change

The Recommendations on Terms of Service and Human Rights developed by the IGF Coalition on Platform Responsibility attempt to define "due diligence" standards for online platforms with regard to **three essential components: privacy, freedom of expression and due process**. The existence of a responsibility of private sector actors to respect human rights was affirmed in the UN Guiding Principles on Business and Human Rights, from which the Recommendations derive their inspiration and the core elements applied to the platform responsibility domain. The Recommendations aim to provide a benchmark for respect of human rights, both in the relation of a platform's own conduct as well as with regard to the scrutiny of governmental requests that they receive. As part of their responsibility, platforms should:

- 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; and
- have in place processes to enable the remediation of any adverse human rights impacts they cause or to which they contribute

References

Dupuy, Pierre, ‘Due diligence in the International Law of Liability’ (Legal Aspects of Transfrontier Pollution, OECD, 1977), Paris, France

30. Duty of care

In legal dictionaries, the term "duty of care" has been put either as a (general) principle of "prudence, meticulousness, care" or an obligation thereto (Le Docte Legal Dictionary in Four Languages, 2011). It is generally defined as "having regard to interests" (id.) While it could refer to a specific and closed type of obligation, which is put only on the government as a duty to protect its members (thus, officials), the concept as defined here also has relevance for relations amongst or between individuals in both the governmental and private sphere.

As used in normal parlance, the duty would require the taking of certain measure(s), and in this context also companies or other non-governmental parties are being targeted. Thus, in terms of "platform responsibility", it means to **responsibly** deal with the negative externalities of (commercial) practices by the internet firms and other market parties involved.

As follows from the academic literature on internet law (see eg Van Eijk et al 2010, Tjong Tjin Tai et al 2015; see also De Streel, A. et al. (2020)), which is commonly based on comparative law methods, the concept "duty of care" is prone to having contested contours. While the term is formally embedded in public laws, for example as part of tort law to support a mechanism for defining negligence in private relationships, it seldom has precise definitions of its own. Nonetheless, it could be said that all (self-, co-, and the more formal) regulatory responses to issues that are identified as relevant in the law and policy making concerning online platforms aim to dictate a "duty-of-care" across both public and private entities based on public order needs. For example, based on other concepts such as trust, these duties of care can be imposed on non-governmental actors when public policy measures identify them as **information fiduciaries**. Similarly, in the **safe harbor** regimes that exist in global internet law, duties of care are established within the dynamics of interpreting the exemptions of **liability** provided for by the regime, such as the European Union's e-Commerce Directive leaving the possibility for Member States to impose reasonable duties of care on service providers 'in order to detect and prevent certain types of illegal activities' (EU Directive on electronic commerce, 2000). An exceptional new legal measure is put forward by the proposal in the UK announced in its "Online Harms White Paper" (2019) aiming to oblige companies to protect users against certain harmful **content** and an online regulator to deal with **internet safety/security** issues, which would put a regulatory framework in place with a "mandatory duty of care". This White Paper and the proposals have been met with several criticisms especially concerning the vagueness of the term "duty of care" that would confront users and platforms alike. It is said that such a general use of the term would bring undue uncertainty if implemented as a statutory response to online **harm**.

References

EU 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 (EU Directive on electronic commerce), OJ L 178/1, 17.7.2000. Available at <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32000L0031&from=en>.

UK Department for Digital, Culture, Media & Sport and Home Office, ' Online Harms White Paper' (UK Government 2019) < https://www.gov.uk/government/consultations/online-harms-white-paper/online-harms-white-paper>

UK Government , ' Press release ' (UK Government 2019) < https://www.gov.uk/government/news/uk-to-introduce-world-first-online-safety-laws>

‘Symposium: Online Harms White Paper’, [2019] Issue of the Journal of Media Law, Vol 11, issue 1, at <https://www.tandfonline.com/toc/rjml20/11/1?nav=tocList&>).

Hans-Werner Zehnhoff AM, Hugues Timmermans, Erika Schmatz, Yvonne Salmon, Le Docte: Legal Dictionary in Four Languages. (1st, Intersentia, Antwerpen 2011)

N.A.N.M. van Eijk, T.M. van Engers, C. Wiersma, C. Jasserand & W. Abel, ' Moving Towards Balance: A Study into Duties of Care on the Internet' [2010] Institute for Information Law Research Paper No. 2012-16 , Available at <https://www.ivir.nl/publicaties/download/Moving_Towards_Balance.pdf>.

De Streel, A. et al. , Online Platforms\' Moderation of Illegal Content Online, Study for the committee on Internal Market and Consumer Protection, (1st, Policy Department for Economic, Scientific and Quality of Life Policies, European Parliament,, Luxembourg 2020) Available at <https://www.europarl.europa.eu/RegData/etudes/STUD/2020/652718/IPOL_STU(2020)652718_EN.pdf>.

Tjong Tjin Tai, T.F.E.; Koops, E.J.; Op Heij, D.J.B.; E Silva, K.K.; Skorvánek, I., Duties of care and diligence against cybercrime (1st, Tilburg University., 2015) Available at <https://pure.uvt.nl/ws/portalfiles/portal/5733322/Tjong_Tjin_Tai_cs_Duties_of_Care_and_Cybercrime_2015.pdf>.

31. End to End encryption

End-to-end encryption (E2EE) is the use of cryptography implemented through the Transport Layer Security protocol (TLS) for the protection of a message so that it can only be read by the communicating users (Electronic Frontier Foundation, 2020; W3C, 2015), and not by third parties acting as intermediaries in the transfer of this data. This is made possible through the use of asymmetric cryptography (also known as public-key cryptography), which generates two keys (large numbers with mathematical properties) for the decryption of the message: a private key for encryption and a public key for decryption (Electronic Frontier Foundation, 2018), unlike symmetric cryptography, where the same key is used for both encryption and decryption (Goanta & Hopman, 2020).

Recent definitional issues around E2EE have shown that some companies may indicate that they implement E2EE when in fact that is not the case (Schneier, 2020; Lee & Grauer, 2020). Given the harms which may arise out of not abiding by security standards a company may refer to in order to appease its user base, the assessment of these implementations can become crucial for consumer protection, contract law and misleading marketing.

More specifically, proper E2EE would require that the sending and receiving user manage their encryption keys and procedures directly, and that third party communication software only ever deals with the encrypted content. As this is inconvenient for the average user, almost all current implementations that claim to offer “end-to-end encryption” (e.g. in instant messaging and videoconferencing tools) offer in fact “managed app-to-app encryption”, in which the application also takes care of creating and managing the user’s keys and of encrypting and decrypting the messages. As a consequence, the application also has access to the unencrypted content and could examine it or make it available to its maker or to other parties.

References

Yan Zhu, ' End-to-End Encryption and the Web' (W3C Technical Architecture Group (TAG) 2015) < https://www.w3.org/2001/tag/doc/encryption-finding/>

SSD.EFF.ORG, ' A Deep Dive on End-to-End Encryption: How Do Public Key Encryption Systems Work?' (SURVEILLANCE SELF-DEFENSE 2018) < https://ssd.eff.org/en/module/deep-dive-end-end-encryption-how-do-public-key-encryption-systems-work>

SSD.EFF.ORG, ' End-to-end encryption' (SURVEILLANCE SELF-DEFENSE 2020) < https://ssd.eff.org/en/glossary/end-end-encryption>

Goanta, C. & Hopman, M. , ' Crypto communities as legal orders.' [2020] Internet Policy Review, 9(2). , DOI: 10.14763/2020.2.1486.

Bruce Schneier, ' Security and Privacy Implications of Zoom' (Schneier on Security 2020) < https://www.schneier.com/blog/archives/2020/04/security_and_pr_1.html>

Micah Lee, Yael Grauer, ' Zoom Meetings Aren't End-to-end Encrypted, Despite Misleading Marketing' (The Intercept 2020) < https://theintercept.com/2020/03/31/zoom-meeting-encryption/>

32. Federated Service

Federated services are those provided by many organizations, in a coordinated manner. Each organization provides the service to its users, but these services interoperate with each other. This means that a user from one organization can exchange data and services with another user served by another organization. E-mail is a good example of a federated service: you can send and receive email messages from any email user, just with their address information. It is not important that it be served by the same organization as you. Thus, a message sent from Gmail or RiseUp reaches users of any other organization perfectly.

Using the idea of federation for other internet services, users could utilize their favorite service with all the benefits of security or usability, but with the additional capability of being able to communicate with each other across servers and organisations (like editing the same document using Google Docs and OneDrive or sharing the same playlist using Deezer and Spotify). Collaboration has always been key in the network services environment, and this federation feature allows maintaining the independence of companies while offering a unified vision to users.

33. Fact-checking

Fact-checking has gained prominence as a concept, in light of the global proportions gained by the divulgation of fabricated and misleading content, frequently categorised as “fake news” or misinformation/disinformation.

The term “fact-checking” however can refer to two different types of activities depending on whether it is performed as parts of editorial responsibility, before the publication of specific content or, as verification of the veracity of suspicious sensationalist content – that may be entirely fabricated in bad faith with the aim of misleading the public opinion – and that is already circulating.

In journalism, fact-checkers traditionally proofread and verify factual claims ex ante, to make sure that articles drafted by reporters correctly represent the facts, thus evaluating the solidity of the reporting, before publication, to avoid responsibility for false claims.

Ex post fact-checking seeks to verify claims and content, thus avoiding that public opinion is deceived while making public figures – typically politicians – accountable for the truthfulness of their statements. As such, fact-checkers in this line of work seek primary and reputable sources that can confirm or negate claims made to the public. (Mantzarlis 2018:82)

Considering the proven vulnerability and permeability to misinformation and disinformation of social networking platforms, this latter fact-checking activity has gained prominence and become particularly relevant to debunk so-called “fake news”, over the past years.

As reported by the UNESCO (Mantzarlis 2018:84), fact-checking is typically composed of three phases:

- Finding fact-checkable claims by scouring through legislative records, media outlets and social media. This process includes determining which major public claims (a) can be fact-checked and (b) ought to be fact-checked.
- Finding the facts by looking for the best available evidence regarding the claim at hand.
- Correcting the record by evaluating the claim in light of the evidence, usually on a scale of truthfulness.

References

Alexios Mantzarlis, ' Fact-Checking 101' in Ireton, Cherilyn, Posetti, Julie (eds), *Journalism, \'Fake News\' and Disinformation: A Handbook for Journalism Education and Training.* (1st, UNESCO, 2018).

34. Flagging

Flagging is a common “mechanism for reporting offensive content to a social media platform” (Crawford and Gillespie, 2016) or other digital platforms, and refers to the act itself of clicking or otherwise demarcating that a specific social media post, link, video, or other content should be removed or reviewed by the platform. According to Crawford and Gillespie, the flagging feature “is found on nearly all sites that host user-generated content, including Facebook, Twitter, Vine, Flickr, YouTube, Instagram, and Foursquare, as well as in the comments sections on most blogs and news sites.” (Crawford and Gillespie, 2016). Flagging processes can involve varying degrees of sophistication depending on the options offered by the platform. For example, some platforms may allow a user to flag content by simply reporting it as offensive but with no further detail or explanation. Other platforms may provide a drop-down menu or open field form upon a piece of content being flagged, which permits the user to write or select a pre-filled reason that they flagged the content (e.g., noting whether it is harassment, contains violence, or contains nudity), or categorizing what they consider the relevant infraction to be (e.g., image-based abuse, copyright infringement, or violating one or more community standards). While the above description is what flagging is widely understood to be at a basic level, Crawford and Gillespie discuss in their paper, “What is a flag for? Social media reporting tools and the vocabulary of complaint” the broader and more complex sociological role and influence that user flags hold or can be interpreted to have across digital platforms, as “a little understood yet significant marker of interactions between users, platforms, humans, and algorithms, as well as broader political and regulatory forces.” (Crawford and Gillespie, 2016).

References

Kate Crawford and Tarleton Gillespie, ' What is a flag for? Social media reporting tools and the vocabulary of complaint' [2016] New Media & Society 18:3 410, 412

35. Filter

This entry (i) introduces the concept and classification of Internet filters, (ii) provides a more detailed (but very general) analysis of network filters, and (iii) provides a similar analysis of endpoint filters.

(i) An introduction to Internet filters

An Internet [content] filter is a piece of software (and, sometimes, of specialized hardware) that selectively blocks content being transmitted in an Internet communication.

Multiple classifications of Internet filters are possible, depending on different factors. *Endpoint filters* operate at the end-points of a connection, i.e. on the server or on the user's device; *network filters* operate in the middle, somewhere on the connection path between the user and the server. *Upload filters* operate when the content is first posted onto the Internet, while *access filters* operate when a user tries to access existing content.

Independently from where and when the content is filtered, the filtering may happen on behalf of different parties. Filters can be voluntarily activated on request of the end-user, to provide services such as parental control for families or productivity control for companies. Network administrators and Internet access providers can deploy filters to prevent connection to harmful services, such as botnet command and control centres or phishing websites. Service and content providers can deploy filters to reject unacceptable content or to prevent access from specific jurisdictions. Governments and courts can mandate the deployment of filters according to applicable regulation, to enforce licensing requirements (e.g. gambling, pharmacies), to prevent access to illegal content (e.g. copyright infringements, child sexual abuse material, hate speech) or to censor their political opponents.

Internet filters are widely used as an alternative to content takedown for cases in which the content cannot or should not be taken down, as they allow to make content inaccessible even without any cooperation by the entity hosting it or by the country where it is located. These cases include:

- content that is legal, but objectionable to the end-user or the network administrator;
- content that is illegal in the country from which the Internet is being accessed, but legal in the country where it is hosted;
- content that is illegal in the country where it is hosted, but cannot be taken down easily and promptly enough for technical, practical or legal reasons.

There is ample debate in legal, moral, policy and technical terms on whether Internet filters are desirable and under which conditions.

(ii) Network filters

Network filters are usually applied by telecommunications providers, and specifically Internet access providers, since the most effective point where to apply them is on the local loop connection between the home network and the ISP's backbone. They are very common for implementing filters on behalf of any of the three stakeholders (the user, the State and the ISP).

Firewalls block connections according to the destination IP address and service. They are effective but suffer both from overblocking - as often a single IP address hosts hundreds of independent websites and services - and from easy circumvention - as the service operator can simply move the service to a different IP address.

Thus, content-level filtering methods have been developed. *Transparent proxies* silently intermediate connections at the application protocol level (HTTP, for example) and examine the actual content to decide whether to allow access to it. *Deep packet inspection (DPI) appliances* look at the content within network packets, to the same effect. Both methods access the actual content, including any personal information included in it, and thus infringe the user's privacy. They have become increasingly ineffective due to the widespread adoption of encrypted protocols (e.g. HTTPS); they would then require breaking the encryption or having a backdoor into it.

As an alternative, *rendez-vous filters* do not examine content, but only act on service connections necessary to obtain metadata for the actual retrieval of content. The most common type is *DNS filters*, which are applied at the endpoint of the connection with the ISP's DNS resolver, where the IP address for the desired hostname is retrieved. These filters do not look at the content and do not require breaking the encryption, but require that the user adopts the filtering resolver.

In policy terms, network filters can break network neutrality and so their use is often restricted by regulation. In the European Union, the Open Internet Regulation (Art. 3) (EU Regulation, 2015) only allows network filters if mandated by law or if necessary for network security and management. At the same time, many European countries have laws or court rulings that mandate the filtering of certain types of content or of specific websites, requiring ISPs to implement such blocks.

The Internet's technical and business community is divided over network filters. Internet platforms, application developers and the IETF prefer the use of endpoint filters whenever possible (Barnes et al., 2016), and have embraced a policy of encrypting connections as much as possible also to circumvent network filters. On the other hand, network operators and Internet service providers, often backed by their governments, find network filters desirable and useful for a variety of purposes.

This disagreement also has implications for competition, as the disruption of network filters by application makers can have the effect of drawing users away from services provided by ISPs and into services provided over-the-top by the platforms (Borgolte et al., 2019), where the filtered content (even if ruled illegal in the user's country) is immediately available.

(iii) Endpoint filters

Endpoint filters are applied at either edge of a network connection.

When applied on the user’s device, they generally are voluntary filters to prevent some users (for example, children) from accessing some content. In this regard, endpoint filter providers directly compete with network filter providers supplying similar services with different technologies.

On the other hand, endpoint filters on the server side are often used to prevent access or distribution of harmful or illegal content, similarly to network filters.

For example, *upload filters* are applied by platforms that distribute user-generated content to verify whether such content may be objectionable or illegal. In the European Union, Article 17 of the recent Copyright Directive (EU Regulation, 2019) de facto mandates the deployment of such filters to prevent the upload of copyright-infringing content.

Search filters are customarily deployed by search engines and other indexing services to hide pointers to content which is deemed illegal or inappropriate. Search engines, mostly based in the United States, customarily remove pointers to some results in response to complaints filed under the Digital Millennium Copyright Act (Google 2020).

Geoblocking is a type of server-side endpoint filtering in which content is made available or not depending on the estimated country of origin of the connection. In the European Union, geoblocking is seen as a potential distortion of the single market and thus has been regulated with the Geo-Blocking Regulation (EU Regulation, 2018).

As endpoint filters are generally implemented by entities other than telecommunication providers, they are usually not regulated against. They do not raise network neutrality concerns, yet platforms could also use them in non-neutral ways to influence which content and services users can access.

References

European Parliament (EC) Regulation (EU) 2015/2120 of the European Parliament and of the Council of 25 November 2015 laying down measures concerning open internet access and amending Directive 2002/22/EC on universal service and users’ rights relating to electronic communications networks and services and Regulation (EU) No 531/2012 on roaming on public mobile communications networks within the Union (Text with EEA relevance) [2015]

R. Barnes, A. Cooper, D. Thaler, E. Nordmark, ' Technical Considerations for Internet Service Blocking and Filtering' (IETF 2016) < https://tools.ietf.org/html/rfc7754>

Borgolte, Kevin and Chattopadhyay, Tithi and Feamster, Nick and Kshirsagar, Mihir and Holland, Jordan and Hounsel, Austin and Schmitt, Paul, , ' How DNS over HTTPS is Reshaping Privacy, Performance, and Policy in the Internet Ecosystem' [2019] Princeton University and The University of Chicago. Availble at: <https://ssrn.com/abstract=3427563>

European Parliament (EC) 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 (Text with EEA relevance.) [2019] 0J L Document 32019L0790

Google Transparency Report, 'Content delistings due to copyright' (Google 2020) < https://transparencyreport.google.com/copyright/overview?hl=en>

European Parliament and of the Council (EC) Regulation (EU) 2018/302 of the European Parliament and of the Council of 28 February 2018 on addressing unjustified geo-blocking and other forms of discrimination based on customers\' nationality, place of residence or place of establishment within the internal market and amending Regulations (EC) No 2006/2004 and (EU) 2017/2394 and Directive 2009/22/EC [2018] 0J L Document 32018R0302

36. (Digital) Gatekeeper

Digital gatekeepers are the equivalent of the guardians of critical infrastructure (e.g. highway, railway, utilities and telecommunications) in the digital world. Thus, the focal element of this definition is the critical nature of the services they provide, in enabling both the enjoyment of services that are considered essential for digital citizenship, and the provision of such services by third parties.

There are, however, competing views of the essential elements of this term. For instance, one of the first users of the term in the legal domain relies on the four criteria identified by an economist (Reiner Kraakman) who focuses on requirements that regulators should meet before designing an entity of gatekeeper, specifically "(1) serious misconduct that practicable penalties cannot deter; (2) missing or inadequate private gatekeeping incentives; (3) gatekeepers who can and will prevent misconduct reliably, regardless of the preferences and market alternatives of wrongdoers; and (4) gatekeepers whom legal rules can induce to detect misconduct at reasonable cost.

Another pioneer in the area, Emily Laidlaw, defines gatekeeper power as a function of the impact on participation in democratic culture, which in turn depends on: (1) when the information has democratic significance; and (2) when the com- munication occurs in an environment more closely akin to a public sphere. As a result of this, she identifies two different categories> Internet gatekeepers, which are those gatekeepers that control the flow of information; and internet information gatekeepers, which as a result of this control, impact participation and deliberation in democratic culture.

More recent scholarship seems to accentuate, rather than reconciliate these divergences. For instance, Thomas Kadri uses "digital gatekeepers" in a less metaphorical sense, referring to the property owners that may permit and restrict access to their websites much like landowners may do with private land in the real world. In this sense, he discusses how cyber-trespass law empowers them with *legal* rights of inclusion and exclusion over information on website.

By contrast, Rory Van Loo calls platforms the "New Gatekeepers" to describe how administrative agencies increasingly conscript them to "perform the duties of public regulator" and police other businesses. Similarly, Daniel Citron defines a special role of "digital gatekeepers" on preventing online hate, referring to entities that "have substantial freedom to decide whether and when to tackle" harms like cyber-harassment by deciding what content appears on their websites.

Adopting a more media-focused approach Eli Pariser has invoked the gatekeeper language to describe how platforms exercise editorial control over the news and information we consume, replacing the "old gatekeepers" that ran traditional broadcast and print media. Helberger and other reinforce this understanding, focusing on the control of critical resources, rather than on access to and supply of information, as a measure of heir ability to affect user choices and diversity of exposure.

More recently, we have a seen a resurgence of the concept of gatekeeper especially within the realm of competition law, as a threshold that triggers a more stringent scrutiny for intervention. Recent reports on proposed changes to the exiting framework of competition law for the digital age use similar terms, such as:

- bottleneck power, where consumers primarily single-home and rely upon a single service provider, which makes obtaining access to those consumers for the relevant activity by other service providers prohibitively costly (EU Special Advisser´s Report)

- intermediation power, linked to having “unavoidable trading partner” status (Competiton 4.0 report)

- strategic market status or competitive gateway, i.e. in a position to exercise market power over a gateway or bottleneck in a digital market, where they control others’ market access. The Furman report focuses on three main variables: i) the power to control access to certain goods and services and charge high access fees; (ii) the power to manipulate rankings or the prominence of a given good/service; and (iii) the power to control reputations.418 It also stresses how the concept of “Significant Market Power” that exists in telecom markets can provide some references on how to think about strategic market status in digital markets. Finally, the CMA in its Digital Markets Report complemented that for platforms funded by digital advertising some of the criteria should include measures of shares of supply in consumer-facing markets, reach across consumers, share of digital advertising revenues, control over the rules or standards which apply in the market and the ability to obtain and control unique datasets

References

Thomas E. Kadri, ' Digital Gatekeepers' [2020] Kadri, Thomas, Digital Gatekeepers (July 31, 2020). Texas Law Review, Vol. 99, Forthcoming,

Danielle Citron, 'Hate Crimes in Cyberspace ' [2014] Harvard University Press

Rory van Loo, ' The New Gatekeepers: Private Firms as Public Enforcers ' [2020] 106 Virginia Law Review 467

Emily Laidlaw, ' A framework for identifying Internet information gatekeepers' [2010] International Review of Law, Computers & Technology, Vol. 24, No. 3, 263, 276

Helberger, Natali, Kleinen-von Königslöw, Katharina, and van der Noll, Rob, ' Regulating the New Information Intermediaries as Gatekeepers of Information Diversity' [2015] Info, VOL. 17 NO. 6, 50, 71

Reinier H. Kraakman, ' Gatekeepers: The Anatomy of a Third-Party Enforcement Strategy' [1986] 2 J.L. ECON. & ORG. 53, 53 & n.1

Jonathan Zittrain, Harvard Journal of Law & Technology Volume 19, Number 2 Spring 2006

Jacques Crémer, Yves-Alexandre de Montjoye and Heike Schweitzer, *Competition Policy for the Digital Era* (1st, European Comission, Luxembourg 2019)

Stigler Center News, ' Stigler Committee on Digital Platforms: Final Report' (Chicago Booth 2019) < https://www.chicagobooth.edu/research/stigler/news-and-media/committee-on-digital-platforms-final-report>

Competition and Markets Authority, ' Online platforms and digital advertising market study' (UK Government 2020) < https://www.gov.uk/cma-cases/online-platforms-and-digital-advertising-market-study>

Commission Competition Law 4.0, ' A New Competition Framework for the Digital Economy' (BMWi 2020)

37. Governance

The concept of governance offers a 'decentred' perspective on regulation, which does not emanate solely from the state but instead emerges from (complex, interactive) constellations of public and private stakeholders. In the words of Julia Black, "'[g]overnance' is a much debated term, but most definitions revolve around the observation that both public and private actors are involved in activities of steering or guiding 'the governed' in ways that may or may not be interrelated." (Black, 2008) Narrower conceptions of 'governance' do exist, in which it remains the sole purview of the state, as do broad conceptions of 'regulation' which also recognize the role of private actors (Gorwa 2019). But despite these various shadings and permutations, most authors tend to see 'governance' as broadly synonymous with **regulation**, though connoting an emphasis on the role of private actors in processes of rule-making and enforcement.

In the context of internet governance, influential accounts including those of Van Eeten & Mueller (2013) and Hoffman, Katzenbach and Gollatz (2016) have argued that governance can and should be distinguished from 'regulation'. In the internet context, they argue, governance often consists of "rules and institutions that emerge as side effects of actors pursuing non-regulatory goals", often the result of "complex coordination processes" (Hoffman, Katzenbach and Gollatz (2016). This broader, non-regulatory conception of governance ecompasses all forms of coordination and interaction which lead to the creation of rules and principles, such as, for instance, standard-setting by Internet Service Providers or online platforms. However, to prevent "governance" from expanding to cover any and all forms of online interaction and coordination, they introduce a requirement of *reflexivity*: an act of coordination becomes an act of governance "when ordinary interactions break down or become problematic [...] and we see ourselves forced to discuss and negotiate the underlying norms, expectations and assumptions that guide our actions".

Whether it is understood as reflexive coordination, or simply as a form of regulation, "governance" has proven to be a highly relevant and widely-used concept in the context of platforms, these being private entities that play an influential role in governing online ecosystems (e.g. Van Dijck, Poel en De Waal 2018). See platform governance.

References

Black, Julia. 2008. "Constructing and Contesting Legitimacy and Accountability in Policycentric Regulatory Regimes". *Regulation & Governance* 2(2). Available at: https://onlinelibrary.wiley.com/doi/full/10.1111/j.1748-5991.2008.00034.x

Gorwa, Robert. 2019. "What is platform governance?". *Information, Communication and society* 22(6). Available at: https://gorwa.co.uk/files/platformgovernance.pdf

Hoffman, Jeannette, Christian Katzenbach and Kisten Gollatz (2016). Between Coordination and Regulation: Finding the Governance in Internet Governance. *New Media + Society* 19(9).

Van Dijck, Jose, Thomas Poell and Martijn de Waal. 2018. *The Platform Society: Public Values In A Connective World*. New York: Oxford University Press.

Van Eeten, Michel and Milton Mueller (2013). Where is the governance in Internet governance? *New Media & Society* 15(5).

38. Harassment

This entry discusses: (I) a brief history of the concept; (II) the notion of harassment and its variations – harassment in the working place, in the streets and online, etc.; (III) the concept of online harassment or cyberharassment; (IV) the possible regulatory frameworks.

(I) As MacKinnon (1986) has pointed out, harassment itself, as an act, is not a novelty in the lives of women. The real novelty was the initiative of feminist lawyers in the 1980s for its typification as a sex discrimination act under Title IV of the Civil Rights Acts of 1964, which made possible for women to bring cases of abuse and harassment in the workplace to the courts. At the time, it was seen as a "feminist invention" because those types of male conducts were deeply naturalized in society. Currently, its deconstruction is still an issue. Nevertheless, women have achieved their legal recognition (Honneth, 1996) and are able to demand reparation and have other claims fulfilled.

(II) Naming this experience of suffering and turning it into a matter of litigation has not only given legitimacy to the victims' demands but has also opened doors for further echoes of abuses that minorities face in their everyday lives. Such is also the case of street harassment, which has seen a marked increase in global movements against in the 2010s, thanks to transnational feminist awareness-raising campaigns that use the internet as its main channel (Kearl, 2015). Harassment in public spaces has a fundamental characteristic which differentiates it from harassment in workplaces: the perpetrators are almost always anonymous to the victims, which makes it difficult for punitive institutions to address it or even for the law to typify it properly, in most cases. As a common characteristic, harassment in the workplace and in the street are seen by the literature (MacKinnon, 2018) as a form of power expression by privileged sectors against minorities, taking advantage of situations of vulnerability which exist because of several inequalities.

(III) As a means of communication—currently with more centrality than ever—the Internet is not free of inequalities that permeate society and generate abuse. Online harassment or cyberharassment, as pointed out by Mary-Anne Franks (2011) “has existed as long as the internet itself has existed” (p. 678). While some people “trolled for lulz”, more insidious types target Internet users with violent threats, doxxing and — as we saw all-too-clearly in the 2016 U.S. election and in the 2018 Brazilian election — state-sponsored mass manipulation.

Franks (2011) differentiates cyberharassment from mere insults or juvenile behavior by pointing out that it targets, in most cases, individuals that belong to subordinated groups and affects their life deeply. As a study (entitled “O Reino Sagrado da Desinformação”, or “The Sacred Kingdom of Disinformation” [2019] in free translation) developed by the Brazilian independent news organization “Gênero e Número” has pointed out, the conservative agenda has been led by far-right-wing sectors on Twitter, aiming to manipulate public opinion with disinformation. Anti-feminist campaigns are often led by articulated networks of masculinists, conservatives and radical christian sects – Caplan and Marwick (2018) named it the "manosphere".

Online harassment occurs through several techniques, as described by Jhaver et al. (2018, p. 15), and targets victims who think differently from the status quo (Fladmoe; Nadim, 2017); women and particularly women of color and those from marginalized groups are more frequently and more intensely targeted with harassment. Online harassment is an umbrella term that refers to a set of specific, damaging behaviors and tactics. Tactics include, but aren't limited to, coordinated attacks, cyberstalking, dogpiling, dog-whistling, doxxing, vile and hateful comments, mob harassment and more. Perpetrators of harassment often target people on multiple platforms using multiple tactics at any given time.

(IV) Technology facilitates abuse on many levels. Kadri et al. (2020, p. 1084) showed that Facebook's "People You May Know" feature made it possible for a man to track and harass a woman that he went on a date with one year after it. Although in most cases online harassment is perpetrated by users that take advantage of internet infrastructure in several ways to cause harm to victims (Jhaver, 2018), differently from street harassment, it may be possible to address and combat it with the help of tech design. This means that since "law is code" (as Lessig, 2006, pointed out), digital abuse can be tackled by technology (Kadri, 2020).

In MacKinnon's (2017) and Franks (2011) terms a possible path for tackling online harassment would involve (1) typifying it, which would make possible for the victims to give a name to their suffering experiences and acknowledge them as violence; (2) producing a denaturalization movement in society; or, in Kadri et al.'s terms (2020, p. 1085) (3) confronting it would involve the development of "empathetic networks", through the diversification of the staff of big tech companies throughout its departments, and through consultations with experts and victims before and during the development of a tech product.

References

Fladmoe, A., & Nadim, M. (2017). *Silenced by hate? Hate speech as a social boundary to free speech.* Boundary Struggles. Contestations of Free Speech in the Public Sphere. Oslo: Cappelen Damm Akademisk, 45-76.

Franks, Mary Anne. (2011) *Sexual Harassment 2.0.* Md. L. Rev. v. 71, pp. 655.

Gênero e Número. (2019). *O Reino Sagrado da Desinformação.* Available at: http://www.reinodadesinformacao.com.br

Honneth, A. (1996). The struggle for recognition: The moral grammar of social conflicts. Mit Press.

Jhaver, S., Ghoshal, S., Bruckman, A., & Gilbert, E. (2018). *Online harassment and content moderation: The case of blocklists.* ACM Transactions on Computer-Human Interaction (TOCHI), v. 25, n. 2, pp. 1-33. Available at:

https://dl.acm.org/doi/abs/10.1145/3185593?casa_token=4L_hW6bPP48AAAAA:gojOpiq_6mGIpU_vN2Xg5HrXRLPX-vysyNHwmmKFYck4yUFv_w-qcTMILy5OnmqsYgFh3OKOHKep_w

Kadri, T. *Networks of Empathy.* Utah Law Review, (2020), Available at SSRN: https://ssrn.com/abstract=3638394

Kearl, H. (2015). Stop global street harassment: growing activism around the world: growing activism around the world. ABC-CLIO.

Lessig, L. (2006). *Code: Version 2.0*. Aufl. New York.

MacKinnon, C. (1986)*. Sexual harassment.* New York: Petrocelli.

MacKinnon, C. A. (2017). *Butterfly politics.* Harvard University Press.

Marwick, A. E., & Caplan, R. (2018). *Drinking male tears: Language, the manosphere, and networked harassment.* Feminist Media Studies, v. 18, n. 4, pp. 543-559.

Waldman, Ari E. *Images of Harassment: Copyright Law and Revenge Porn* (December 3, 2015). v. 23. Federal Bar Council Quarterly 15 (Sept./Oct./Nov. 2015), Available at SSRN: https://ssrn.com/abstract=2698720

39. Harm (online harm)

Harm is the result of words, actions (or even *inactions*) that cause physical, emotional or psychological damage to someone, including violence, defamation, or economic loss. This extends to potentially non-tangible damage, including causing a person or group of people to fear for their physical, emotional or psychological safety, experience anxiousness, limit their speech, feel intimidated in their personal or professional life, or worry for their personal or professional reputation.

Harm can also scale from the personal to societal, cultural and political realms.

A UK government white paper (UK Government, 2020) released in 2019 and updated in early 2020 describes “harmful content or activities” categorized into harms with a “clear” definition (e.g. child sexual exploitation and abuse, terrorist content), “less clear” definitions (e.g. cyberbullying, coercive behavior, intimidation, disinformation) or those harmful if underage children are exposed to said content or activities (e.g. children accessing pornography). However, this leaves ample room for interpretation about what “harm” actually means and who these “clear” or “less clear” content or activities harm. That leaves the content or activity to stand on its own, with the reader to interpret however they choose.

Platforms have also used the word “harm” as an outcome of content and activity on their platform although, again, harm isn’t always defined clearly and can be widely debated among users. For example, Twitter’s Trust and Safety team uses the term (“offline harm”, “type of potential harm”) when announcing new policies (i.e. a recent announcement to remove QAnon content, Twitter 2020), but harm is then contested. Some users see harm to “freedom of speech” as outweighing potential “offline harms,” while others may appreciate the recognition that particular content or activity causes harm.

References

Joint Ministerial foreword, ' Online Harms White Paper' (UK Government 2020) < https://www.gov.uk/government/consultations/online-harms-white-paper/online-harms-white-paper#the-harms-in-scope>

Twitter, ' Thread on Twitter Safety ' (Twitter 2020) < https://twitter.com/TwitterSafety/status/1285726277719199746?s=20>

40. Hash/Hash database

A hash is a function that can be used to generate a unique identifier or value that can then be converted to another value and decoded via a hash table, and is used for several purposes. With respect to content moderation and platform governance, a hash is akin to a digital fingerprint that is added to multimedia (photos, video, etc) which provides a unique identifier and enables that content to be identified across the internet and for the search for, and removal of, the content associated with the hash to be automated.

Hash databases enable the sharing of these unique identifiers, or hashes, across platforms without having to share the content itself. Hashing enables coordinated action, such as content takedown, and allows companies to share information about content deemed unacceptable for a given platform across different service. Hashing technology such as PhotoDNA (Microsoft, n.d) has been used to combat the spread of child pornography, terrorist content, and other unwanted or illegal content, such as extremist content.

In 2009, Microsoft and Dartmouth University launched (Gregoire, 2015) PhotoDNA to help combat the trafficking and sexual exploitation of children, and in 2018 (Langston, 2018) expanded its use for video. The hash database is provided for free to law enforcement and civil society partners, and overseen (Microsoft, n.d) by the National Center for Missing & Exploited Children (NCMEC) in the United States.

In 2016, Facebook, together with Google and Microsoft, created a hash database of ISIS videos to coordinate the removal of terrorist content. This collaboration formed the basis for the creation of the Global Internet Forum for Terrorist Content, which grew up around the hash database to include dozens of companies that coordinate around content removal, and spun off into a stand-along organization in mid-2020. Critics have raised concerns about the opaque nature of this collaboration and the failure of the companies involved to maintain a database or other form of access to affected content that researchers and independent auditors could review and study. Although founding companies said the hash database would only include Al Qaeda and ISIS-related propaganda, in the wake of the 2019 Christchurch massacre of Muslims in New Zealand there was pressure to expand the remit of the database to include other forms of extremism. As of 2018 there were more than 200,000 pieces of content in the database, according to the GICT transparency report (GIFCT, 2020).

Critics of the GIFCT and the approach to coordinated content takedown via hash databases express concern about the potential for the technology and approach to be co-opted toeradicate other types of content, such as hate speech or misinformation. It is also not entirely clear under data protection law how content associated with such hash databases ought to be saved, categorized, and made available for independent, third party oversight and research.

References

Microsoft, ' Photo DNA' (Microsoft) < https://www.microsoft.com/en-us/photodna>

Courtney Gregoire , First Microsoft PhotoDNA update adds Linux and OS X support, detections up to 20 times faster' (Microsoft 2015). Available at: https://blogs.microsoft.com/on-the-issues/2015/12/18/first-microsoft-photodna-update-adds-linux-and-os-x-support-detections-up-to-20-times-faster/#:~:text=This%20week%20marks%20the%20sixth,companies%2C%20organizations%2C%20and%20developers.

Jennifer Langston, ' How PhotoDNA for Video is being used to fight online child exploitation' (Microsoft 2018) <https://news.microsoft.com/on-the-issues/2018/09/12/how-photodna-for-video-is-being-used-to-fight-online-child-exploitation/>

GIFCT, ' GIFCT Transparency Report' (Global Internet Forum to Counter Terrorism 2020) < https://www.gifct.org/transparency/>

41. Hate Speech

This entry aims to (I) present the definition of hate speech both according to the human rights law and specialized literature; (II) draw a distinction between hate speech and harm; (III) and highlight the possible solutions to address this issue, in a multistakeholder approach.

Hate speech isn't a new phenomenon in society, and neither are the attempts for addressing and combating it. As the International Covenant on Civil and Politics Act (ICCP, United Nations, 1966) already sustained in the mid-1960s, right after the protection of free speech (art. 19) comes the call for the prohibition of hatred speech (art. 20), naming it as "[the] *advocacy* of *national, racial or religious* hatred that constitutes incitement to discrimination, hostility or violence". Subsequently, many other legal instruments tried to encompass "new" forms of discrimination, such as the Committee of Ministers of the Council of Europe Recommendation No R 97(20) 30.10.1997 on “hate speech”, which includes other vulnerable groups into the definition. Also, this document makes liable not only the individuals who advocate in favor of hatred speech, but also the ones that act to "*spread, incite, promote* or *justify*" any content related with "*racial hatred*, *xenophobia, anti-Semitism or other forms of hatred based on intolerance,* including intolerance expressed by *aggressive nationalism and ethnocentrism, discrimination and hostility towards minorities, migrants and people of immigrant origin*” (Council of Europe, 1997). These efforts show the commitment of several relevant institutions for cultural changes, being incorporated by domestic legislations worldwide.

In spite of the advantages of the existence of general definitions capable of encompassing all of what is considered to be hate manifestations, in a universalist approach, the absence of objectivity leaves it a huge space for the acting of judges and punitive institutions to apply the legislation. This can lead to several issues for law enforcement. In a very similar way, when it comes to online hate speech, there is a huge space for the acting of social network companies to define what they consider to be hate speech and apply content moderation on the online environment.

Considering the online forums as the new public sphere, platforms are being called out to assure the equal participation of users, to combat online violence and to enforce content moderation. Moreover, legislators are drawing domestic laws (such as the Germain NetzDG, and the Brazilian Fake News Draft Bill), so that platforms commit to the maintenance of a safe digital sphere and the protection of democractic values. On one side of the discussion, some groups are advocating for the prevalence of free speech, as framed in the US' 1st Amendment, above other principles. These groups tend to call platforms' initiatives to prohibit hate speech as censorship.

On the other side of the discussion, academics such as Mary-Anne Franks (2019) and Danielle Citron, say that the debate around hate speech should be centered on how some groups in society are being historically silenced and are powerless against several violations – such as women, non-white men and other minorities (Citron, 2014). In this sense, returning to the ICCP document, in their view institutions should make efforts to guarantee both the protection of free speech and the combatment of hate speech.

In spite of problematic publications shared on the online environment being often in a "grey zone", some initiatives could be adopted in a multistakeholder approach in order to tackle online hate speech:

(i) From the companies' point of view:

(1) The platforms should be opened to the community, in a sense that the users could be able to make their own contributions – either by engaging in conversations with the offenders, or by flagging the content that is understood as offensive, or even by reporting it for ex-post removal by moderators;

(2) Social media companies should establish specific guidelines and community standards and perform content moderation with human and automated techniques, following the basic principles of due process and transparency in the removal procedures.

(ii) From the institutions point of view:

(3) The Legislative branch should develop domestic legislations and ordinances that follow international norms and principles about addressing gender-based or race discrimination, centered on the victims' perspectives;

(4) The Executive branch should make and support policies and campaigns that aim to eradicate all forms of prejudice (in schools, in public places, in work spaces, online and in the State institutions);

(5) The Judiciary should establish jurisprudence centered on the victims' perspectives in order to define what type of behaviors are considered hate speech.

(6) International law standards – such as human rights conventions – regarding the definition of hate speech, and also the design of procedural rules to assure due process in content moderation, may significantly help to give more transparency to the removal process of problematic content. Improving public awareness and consciousness raising through normative standards can help to engender a safer and healthier online environment.

References

Citron, Danielle Keats. ‘Hate crimes in cyberspace’, [2014] Harvard University Press.

Franks, Mary Anne. ‘The Cult of the Constitution**.**’ [2019] Stanford University Press.

Recommendation No. R (97) 20 of the Committee of Ministers to Member States on “Hate Speech”, 30 October 1997.

OHCHR UN, ‘International Covenant on Civil and Political Act’. Available at: https://www.ohchr.org/en/professionalinterest/pages/ccpr.aspx

42. Human exploitation

There is no agreed definition on what human exploitation is, however international law lists the types of illegal activities related to this practice. Elaborated in 2000, the Palermo Protocol, conceived within the United Nations, is the international legal instrument that deals with human trafficking (Allain, 2013).

The Palermo Protocol (UNGA, 2008) defines human trafficking by a series of actions (recruitment, transport, transfer, accommodation or reception) that may be carried out by different means (threat, use of force, other forms of coercion, abduction, fraud, deception, abuse of authority, taking advantage of the situation of vulnerability of others, delivery or acceptance of benefits - pecuniary or not - for obtaining the consent of others over whom one has authority) for the purpose of exploitation, whatever it may be, of a person.

Despite classifying some practices, the list presented is not exhaustive, and other forms of exploitation can and should also be recognized for the purpose of trafficking.

That said, it can be understand that human trafficking and exploitation doesn’t have a singular meaning, yet it covers a number of forms of human exploitation that can appears in the form of **sexual exploitation**, when someone is deceived, coerced or forced to take part in sexual activity; **labor exploitation**, when people are coerced to work for little or no remuneration, often under threat of punishment; **domestic servitude**, when there are restrictions on the domestic worker’s movement and they are forced to work long hours for little pay; **forced marriage**, when a person is threatened with physical or sexual violence or placed under emotional or psychological distress to be forced married; **forced criminality**, when somebody is forced to carry out criminal activity through coercion or deception; **child soldiers**, when children are used for combats and are made to commit acts of violence or within auxiliary roles such as informants or kitchen hands; and **organ harvesting**, when an organ is removed with or without consent to be selled often as an illegal trade.

Among the major online platforms, Facebook has the extensive exclusive section dedicated to human exploitation in its Community Standards, where, in addition to the activities mentioned above, it also officially condemns content that are related to the activities of sale of children for illegal adoption; orphanage trafficking and orphanage voluntourism. Platforms usually prohibit content geared towards the recruitment of potential victims, to the facilitation of human exploitation; and to the promoting, depicting, or advocating these criminal activities.

According to the United Nations Office on Drugs and Crime (2018), millions of women, men and children are forced to work in inhumane conditions on farms, in clothing warehouses, on board fishing boats, in the sex industry or in private homes, generating billions of dollars a year. Civil society organizations have been warning that social networks are increasingly constituting a recruitment platform for human exploitation, being a tool to identify and contact potential victims.

Other internet-based services are also useful for abusers, such as anonymous online payments, and encrypted messaging. On the other hand, technologies have also been used to combat trafficking, standing out the text analysis tools to identify a writing pattern in sexual ads, and facial recognition mechanisms (Mzezewa, 2017).

Finally, it is important to notice that lately there has been a discussion regarding whether social media content moderators suffer a form of human exploitation in jobs where everyday they have to see violent content and decide whether it should remain online or not. Even thought that has been little formal study on the impact of this kind of routine, where some people spend eight to nine hours reviewing a series of suicide, harmful and sexual content in order to keep social media platforms safe from it, it has been proved that a number of workers had developed post-traumatic stress syndrome as a result of this activity, and this situation is becoming more and more commnon and it shouldn't go unnoticed (Cardoso, 2019).

References

Allain, Jean. (2013). Slavery in International Law: Of Human Exploitation and Trafficking - The Introduction. Available at: <https://www.researchgate.net/publication/286456166_Slavery_in_International_Law_Of_Human_Exploitation_and_Trafficking_--_The_Introduction>

Cardoso, Paula (2019). Precariado algorítmico: o trabalho humano fantasma nas maquinarias da inteligência artificial. Media Lab UFRJ. Available at: <http://medialabufrj.net/blog/2019/09/dobras-38-precariado-algoritmico-o-trabalho-humano-fantasma-nas-maquinarias-da-inteligencia-artificial/>

Mzezewa, Tariro (2017). 'Hacks That Help: Using Tech to Fight Child Exploitation'. <https://www.nytimes.com/2017/11/24/style/sex-trafficking-hackathon.html>.

UN General Assembly (2000). Protocol to Prevent, Suppress and Punish Trafficking in Persons, Especially Women and Children, Supplementing the United Nations Convention against Transnational Organized Crime, 15 November 2000, available at: <https://www.refworld.org/docid/4720706c0.html>

UNODOC (2018). 'Global Report on Trafficking Persons.' Available at: <https://www.unodc.org/documents/data-and-analysis/glotip/2018/GLOTiP_2018_BOOK_web_small.pdf>

43. Human Review

Human review is a term specific to the field of content moderation, or how digital platforms manage, curate, regulate, police, or otherwise make and enforce decisions regarding what kind of content is permitted to remain on the platform, and with what degree of reach or prominence, and what content must be removed, taken down, filtered, banned, blocked, or otherwise suppressed. Human review refers to the part of the content moderation process at whichcontent that users have flagged (see "flagging" and "coordinated flagging") as offensive is reviewed by human eyes, as opposed to assessed by an algorithmic detection or takedown tool, or other form of automated content moderation. These human reviewers may be the platform company's in-house staff, more frequently the case at smaller platform companies; the platform's own users who have voluntarily stepped into a moderator role, such as is the case with Reddit's subreddit moderators and Wikipedia's editors; or low-paid third-party, external contractors that number up to the thousands or tens of thousands, operating under poor working conditions in content moderation "factories", as relied on by larger platforms such as Facebook and Google's YouTube (Caplan, 2018). (These three roles of human reviewers correspond to Robyn Caplan's categorization of content moderation models, namely, the artisanal, community-reliant, and industrial approaches, respectively) (Caplan, 2018).

Human review is also used to check lower-level moderators' decisions as well as to check that automated or algorithmic content moderation tools are making the correct decisions (Keller, 2018), such as to "correct for the limitations of filtering technology" (Keller, 2018). As Daphne Keller points out, one danger of combining human review with algorithmic filters, though also a reason to ensure the continued involvement of human review in content moderation, is that "once human errors feed into a filter's algorithm, they will be amplified, turning a one-time mistake into an every-time mistake and making it literally impossible for users to share certain images or words." (Keller, 2018).

References

Robyn Caplan, 'Content or Context Moderation? Artisanal, Community- Reliant, and Industrial Approaches' [14 November 2018], Data & Society. Available at: <datasociety.net/wpcontent/ uploads/2018/11 /DS_Content_or_Context_Moderation.pdf>.

Daphne Keller, 'Internet Platforms: Observations on Speech, Danger, and Money' [2018], Hoover Institution. Available at: <https://www.hoover.org/sites/default/files/research/docs/keller_webreadypdf_final.pdf>.

44. Incitement to violence

Traditional incitement to violence, enshrined in Article 20 of the ICCPR and criminalised in many domestic jurisdictions, has taken on new significance due to the proliferation of online platforms and the growth of global terrorism (Bayefsky and Blank 2018). Social media and online platforms are a way in which to 'amplify' the harm of incitement to violence (Avni 2018, 30–31). While social media intermediaries provide a mechanism by which those inciting violence can access a broader and more diverse audience, the prohibition on incitement to violence is not meaningfully enforced (Matas 2018, 150). This can be explained by the difficulties democratic states face in balancing the problem of virtual hate speech with foundational principles of free speech (Guiora 2018, 142). Incitement to violence is the most 'severe' form of online **hate speech**, in that it 'threaten[s] with violence, incite[s] violent acts, and intend[s] to make the target fear for their safety' (Alexandra Olteanu et al. 2018, 5).

Various factors can lead to incitement to violence over digital platforms, including an absence of or unclear legislation on the issue, negative or stereotyped portrayal of minority groups in the media, structural inequalities in access to social media platforms, and the changing media landscape (Izsák 2015, 51–79). A modern example includes the spread of **hate speech** and incitement to violence via the internet in Myanmar, which played a 'significant' role in the Rohingya genocide (OHCHR 2014; Human Rights Council 2018, para. 74). Online service providers are beginning to acknowledge the role they play in the dissemination of material which incites violence; Facebook's Community Standards claims to 'remove language that incites or facilitates serious violence' (Facebook 2020, pt. 1), Google's User Content and Conduct Policy prohibits 'Hate Speech' which it defined to be 'content that promotes or condones violence' against an individual or group 'on the basis of their race or ethnic origin, religion, disability, age, nationality, veteran status, sexual orientation, gender, gender identity, or any other characteristic that is associated with systemic discrimination or marginalisation' (Google 2020, para. 3), and Twitter's violent threats policy prohibits 'statements of an intent to kill or inflict serious physical harm on a specific person or group of people' (Twitter 2019).

In the online age, the purpose of incitement to violence has shifted; what used to be 'justification for action and recruitment to prepare for action' has now become a simple 'call to action', and service providers must adequately monitor and protect against a risk of harm that they themselves have facilitated (Matas 2018, 163). As incitement to violence is an inchoate offense, in that harm does not need to be actioned but merely called for, the freedom which social media platforms provide to express opinions inherently inflate the *possibility* of violations. However, given the new and larger audiences accessible to those promoting violence, these platforms also increase the *likelihood* of incitement causing violence. Additionally, the growing prevalence of cyberbullying, virtual sexual harassment, and online stalking make clear that the act of violence itself in the age of online platforms is 'still developing and not univocal'(Dubravka Šimonović 2018, 5).

References

Alexandra Olteanu, Carlos Castillo, Jeremy Boy, and Kush R. Varshney. 2018. 'The Effect of Extremist Violence on Hateful Speech Online'. *International AAAI Conference on Web and Social Media; Twelfth International AAAI Conference on Web and Social Media*. https://www.aaai.org/ocs/index.php/ICWSM/ICWSM18/paper/view/17908/17013.

Avni, Micah. 2018. 'Incitement to Terror and Freedom of Speech'. *Incitement to Terrorism*, 30–36. https://doi.org/10.1163/9789004359826_005.

Bayefsky, Anne F., and Laurie R. Blank, eds. 2018. *Incitement to Terrorism*. *Incitement to Terrorism*. Brill Nijhoff. https://brill-com.wwwproxy1.library.unsw.edu.au/view/title/36109.

Dubravka Šimonović. 2018. 'Report of the Special Rapporteur on Violence against Women, Its Causes and Consequences on Online Violence against Women and Girls from a Human Rights Perspective'. A/HRC/38/47. Human Rights Council. https://www.ohchr.org/EN/HRBodies/HRC/RegularSessions/Session38/_layouts/15/WopiFrame.aspx?sourcedoc=/EN/HRBodies/HRC/RegularSessions/Session38/Documents/A_HRC_38_47_EN.docx&action=default&DefaultItemOpen=1.

Facebook. 2020. 'Community Standards'. 2020. https://www.facebook.com/communitystandards/violence_criminal_behavior.

Google. 2020. 'Terms and Policies - Currents Help'. 2020. https://support.google.com/googlecurrents/answer/9680387?hl=en.

Guiora, Amos N. 2018. 'Inciting Terrorism on the Internet: The Limits of Tolerating Intolerance'. *Incitement to Terrorism*, March, 135–49. https://doi.org/10.1163/9789004359826_016.

Human Rights Council. 2018. 'Report of the Independent International Fact-Finding Mission on Myanmar'. 39th Session. https://documents-dds-ny.un.org/doc/UNDOC/GEN/G18/274/54/PDF/G1827454.pdf?OpenElement.

Izsák, Rita. 2015. 'Report of the Special Rapporteur on Minority Issues'. A/HRC/28/64. https://undocs.org/pdf?symbol=en/A/HRC/28/64.

Matas, David. 2018. 'Combating Incitement to Violence on the Internet through Service Provider Action'. *Incitement to Terrorism*, March, 150–64. https://doi.org/10.1163/9789004359826_017.

OHCHR. 2014. 'Myanmar: UN Expert Warns against Possible Backtracking, Calls for More Public Freedoms', 2014. https://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=14910&LangID=E.

Twitter. 2019. 'Violent Threats Policy'. 2019. https://help.twitter.com/en/rules-and-policies/violent-threats-glorification.

45. Inclusive Journalism

In increasingly diverse societies, being labelled as „developed democracies" referred to as countries in transition to democracy, a need for fair, accurate and responsible journalism stands at the top of the requests for rebuilding media for democratic future. The process of reforming the media system after a conflict or a long period of absence of democratic institutions in different countries restate this need acknowledging that the reform would be impossible without reforming the journalistic sector of media. Journalism is a vehicle to public conversation and civic action and strengthening journalism training and education contributes to strengthening its value for the society.

The UNESCO Media Development Indicators, tailored to identify how media reflect diversity of society in order to fulfill its democratic potential, underline the importance of the presence of minority groups in mainstream media. The significance of diversity has been recognized by other key freedom of expression organizations which believe that freedom of expression should be enjoyed by *all* citizens regardless of their race, ethnicity, faith, religion, language, gender, social status, (dis)abilities or sexual orientation. In 2007 the UN Special Rapporteur on Freedom of Opinion and Expression, the OSCE Representative on Freedom of the Media, the Organisation of African States Special Rapporteur on Freedom of Expression and the ACHPR (African Commission on Human and Peoples" Rights) Special Rapporteur on Freedom of Expression and Access to Information, made a joint Declaration on Promoting Diversity in the Broadcast Media. The declaration stressed "the fundamental importance of diversity in the media to the free flow of information and ideas in society, in terms both of giving voice to and satisfying the information needs and other interests of all, as protected by international guarantees of the right to freedom of expression". The United Nation"s notion of inclusive society, „society for all" over-rides differences of race, gender, class, generation, and geography, and ensures inclusion, equality of opportunity as well as capability of all members of the society. Among the prerequisites for inclusive society - respects for all human rights, freedoms, a rule of law, the existence of strong civic society - equal access to public information and tolerance for cultural diversity education plays a critical role. It provides opportunities to learn the history and culture of one's own and other societies, which cultivates the understanding and appreciation of other societies, cultures and religions. The „inclusive society" implies a radical set of changes through which society restructures itself to embrace all of its members.

Journalism that is able to relate to the idea of inclusive society can be called „inclusive journalism". Inclusive journalism challenges the status quo in order to prevent the media from intentionally or unintentionally spreading prejudice, intolerance and hatred. The idea of inclusive journalism is rooted in and un-separable from the political notion of inclusive democracy.

Used interchangeably, inclusive democracy and inclusive society, indicate a type of political system that goes beyond recognizing formal equality of all individuals and involves taking actions and special measures to compensate for inequalities of unjust social structures. Young (2002: 53) says that "democratic norms mandate inclusion as a criterion of the political legitimacy of outcomes" and distinguishes two forms of social exclusion: „external‟ where groups and individuals are openly excluded for the decision making process and „internal‟ where "the terms of discourse make assumptions some do not share, the interaction privileges specific styles of expression, the participation of some people is dismissed as out of order" (ibid).

The objective of inclusive journalism, and educating and training journalists in increasingly diverse society, is to develop inclusive communicative competence. This ability involves reflective thinking, experience of social, political and cultural pluralism, recognition of otherness and critical stand towards the process of constructing identities. Inclusive journalism acts as a catalyst for society to get informed knowledge of its diverse „self‟, as well as an understanding of the relationship between the individual and society. Most university journalism programs are so focused on the questions of developing academic discipline by integrating theory and practice that they dislocate journalism from its natural embeddedness into community. Mensing notes that most university journalism programs preserve the structure of education based on industrial model of journalism and argues that "moving the focus of attention from the industry to community networks could reconnect journalism with its democratic roots and take advantage of new forms of news creation, production, editing, and distribution" (Mensing 2011, p.16).

In transition countries and post-conflict societies this dislocation could have profound consequences.

The essential curriculum for inclusive journalism, based on MDI's work is comprised of the following modules:

- *Developing Sensitivity to Diversity* – type of a module that aims to foster students understanding of the experiences of minorities;
- *How Diversity Is Reported* – traditional academic module based on using standard techniques of news story content analysis enabling students to reach an understanding of how their society‟s media cover diversity issues;
- *Reporting Diversity* – practice based module for students to gain experience of the issues involved in covering minority affairs; and
- *Social Diversity and the Media* -standard teaching (lectures/essays) module using elements taken from a number of academic disciplines . sociology, social psychology and political science that deal with the issue of social diversity and that offer useful insights to journalism students studying theories of media power and social function to provide students

In all modules developed through the Media Diversity Institute‟s inclusive journalism program, the question of assessment aroused as a key tool for enabling students to engage critically with social diversity and to acquire the skills necessary to conceptualize and produce a competent piece of journalism. Module assessment that combines academic and journalistic work has proved to be the model that the majority of journalism educators listed as the best way to evaluate different qualities in journalistic take on diversity issues. This “holistic and highly contextualized assessment” (Biggs 1999, p.152) clearly requires an active demonstration of knowledge of contemporary journalism. It deals with functional knowledge by setting up tasks that are an interesting and challenging learning experience for students rather than taking it as a judgemental instrument in academic analysis of media. The assessment that includes both formative and summative procedures might take different forms, as outlined in some of the modules developed within the MDI‟s inclusive journalism curriculum framework.

References:

Rupar, V., & Pesic, M. (2012). Inclusive journalism and rebuilding democracy. In N. Sakr, & H. Basyouni (Eds.), *Rebuilding Egyptian media for democratic future* (pp. 135-153). Cairo, Egypt: Aalam Al Kotob Publisher.

Young, Iris Marion, 1949- Title: Inclusion and democracy [electronic resource] / Iris Marion Young. Notes: Electronic reproduction. Oxford: Oxford University Press, 2004. (Oxford scholarship online). Publisher: Oxford: Oxford University Press, 2002 Description: 314 p. Internet Link: http://abc.cardiff.ac.uk/login?url=http://www.oxfordscholarship.com/oso/public/content/politicalscience/0198297556/toc.html

Mensing, D. (2011). “Realigning journalism education” In Franklin, B. and Mensing, D. (2011) *Journalism Education, Training and Employment.* New York: Routledge, pp.15-32.

Biggs, J. (1999). *Teaching for quality learning at university*.Great Britain: The Society for Research into Higher Education and Open University Press.

46. Information Fiduciary

Information fiduciaries are entities entrusted with the management of the personal information of third parties. The concept, first proposed by Balkin and Zittrain (2016), evokes an analogy with professional figures assigned with fiduciary duties due to the relationship with their clients, which leads to situations of asymmetrical power and. For instance, doctors, lawyers and accountants are all in a fiduciary relationship with their clients due to their superior knowledge and skills, which requires the establishment of a relationship of trust. As a result, they are bound by duties of care, loyalty and confidentiality.

Accordingly, the online platforms identified as information fiduciaries would owe their customers a duty of loyalty, that is, to act in the best interests of their customers, without regard to the interests of their own business. They would also owe a duty of care, that is, to act competently and diligently to avoid harm to their customers. This means, for example, that they would not be allowed to use data for different purposes from those stated at the time of collection, and they would be required to take reasonable steps to secure any information entrusted to them. Finally, they would be required to

The proposal has had some traction in Internet governance circles, leading even to the introduction in the US Senate of a bill (The “Data Care Act”) that would further specify the duties, including: the obligation to notify data breaches concerning an individual; the duty not to use data in a way that is unexpected and highly offensive to a reasonable end user; the duty not to disclose or sell personal data to third parties that do not have the same level of fiduciary duties, and to take reasonable measures to ensure that such duties are fulfilled.

At the same time, the proposal provoked criticism, for one because it does not contain limits on the collection of personal data, but especially because fiduciary obligations to customers are fundamentally incompatible with the nature of publicly listed corporations (where managers are under a fiduciary duty to maximize shareholder value) and the predominant business model on the Internet (where personal data are regularly used for advertising purposes).

References

Jack M. Balkin, Jonathan Zittrain, ‘A Grand Bargain to Make Tech Companies Trustworthy’ [3 October 2016] at https://www.theatlantic.com/technology/archive/2016/10/information-fiduciary/502346/

Khan, Lina and Pozen, David E., ‘A Skeptical View of Information Fiduciaries’ [2019]. Harvard Law Review, Vol. 133, pp. 497-541

47. Infrastructure

Infrastructure, according to the Cambridge Dictionary's definition, is “the basic structure of an organization or system which is necessary for its operation”. More specifically, the Internet Infrastructure Coalition (2018) defines Internet infrastructure as follows:

“Internet infrastructure is the physical hardware, transmission media, and software used to interconnect computers and users on the Internet. Internet infrastructure is responsible for hosting, storing, processing, and serving the information that makes up websites, applications, and content.”

Namely, the physical infrastructure comprises all the equipment that transmits data through the network, such as submarine and terrestrial cables, backbones, routers, satellites, antenna towers, and even smartphones (Constantinides et. al, 2018); and all the equipment that stores internet data, such as data centers and database servers. As for the virtual infrastructure, we have another important set of foundations, for instance, open standards (eg. IEEE 802.11s), the Internet protocol suit (TCP/IP), the Domain Name System (DNS), and the Hypertext Transfer Protocol (HTTP).

A relevant trend related to infrastructure is the growing investment of online platforms in physical infrastructure, such as submarine cables. This attention to infrastructure is due to the fact that the current foundations are not being sufficient to support the traffic generated by the big platforms like Google, Facebook and Microsoft (Burgess, 2018). In addition to the control that such platforms already exercise in the content layer, additional control in the infrastructure layer can exacerbate problems related to competition, privacy and net neutrality.

Another important problem concerns attacks on critical infrastructure, targeting end users, devices, network services and web servers. Computer emergency response teams have done a great job to address this issue and combat these attacks over the last decades (Bada et.al, 2014). However, one type of attack that these teams do not address is physical incidents, caused by both humans and animals (Arthur, 2013; Moss, 2020). The growing threat of a physical attack should not be underestimated, as it can cause huge damage to economies and national security (Starosielski, 2019).

References

Arthur, Charles (2013). ‘Undersea internet cables off Egypt disrupted as navy arrests three’ Available at: <https://www.theguardian.com/technology/2013/mar/28/egypt-undersea-cable-arrests >

Bada, M., Creese, S., Goldsmith, M., Mitchell, C., & Phillips, E. (2014). Computer Security Incident Response Teams (CSIRTs): An Overview. Global Cyber Security Capacity Centre, 1-23. Available at: <http://www.elizabethphillips.co.uk/Research/CSIRTs.pdf>

Burgess, Matt (2018). 'Google and Facebook are gobbling up the internet's subsea cables' Available at: <https://www.wired.co.uk/article/subsea-cables-google-facebook>

Cambridge Dictionary. 'Infrastructure meaning' Available at: <https://dictionary.cambridge.org/dictionary/english/infrastructure>

Constantinides, P., Henfridsson, O., & Parker, G. G. (2018). 'Platforms and infrastructures in the digital age.' Available at: <http://ide.mit.edu/sites/default/files/publications/ISR%202018%20Constantinides%20Henfridsson%20Parker%20Editorial.pdf>

Internet Infrastructure Coalition (2019). 'What is the Internet's Infrastructure?' Available at: <https://www.i2coalition.com/what-is-the-internets-infrastructure-video/>

Moss, Sebastian (2020). 'How cows caused a small Google network outage' Available at: <https://www.datacenterdynamics.com/en/news/how-cows-caused-small-google-network-outage/>

Starosielski, Nicole (2019). 'Strangling the Internet' Available at: <https://limn.it/articles/strangling-the-internet/>

48. Intermediary Liability

This entry defines in advance what an intermediary is, to later clarify its liability.

There is a broad spectrum of actors labeled as internet intermediaries, such as internet service providers (ISPs), web hosting providers, social networks, cloud service providers, domain name registrars and search engines. Certainly, this is a non-exhaustive list, since there several types of internet related services and different organizations and national laws have their own definitions and categorizations of internet intermediaries (OECD, 2010; OAS, 2011; Article 19, 2013).

"Intermediary liability" refers to the legal responsibility of intermediaries regarding both the actions taken and the content generated by users of their services (MacKinnon et al., 2015). That is, this type of liability does not concern the legal responsibility related to the platform own content or other ancillary issues (eg. tax payment, labor obligations).

Despite not having a binding character, an important document on intermediary liability called Manila Principles (EFF, 2015) is still used today by renowned academics who study this topic and by countries as a model to implement fair and democratic digital policies.

There are different levels of calibration for intermediary liability (Bankston et. al, 2012). First, we have the strict liability model under which intermediaries are held responsible for user-generated content. Next in order, the safe harbour model is a little bit more flexible, where the intermediary is exempt from liability in relation to content published by third parties, but still needs to comply with certain requirements. Lastly, there is the broad immunity model which protects intermediaries from liability for a great variety of content posted by users.

China has the most notorious strict liability model, where intermediaries are in a situation where they must brutally monitor all content posted by their users, otherwise they may be penalized. Brazil has a unique example of the safe harbour model, its *Marco Civil da Internet* defines that intermediaries will only be held responsible if they do not remove content after receiving a court order, except in matters regarding copyright and revenge porn (Zingales, 2015). Regarding the broad immunity model, the Section 230 of the Communications Decency Act (CDA), a U.S. law, is the most cited example of expansive protections against liability, where intermediaries are not responsible for third-party content.

References

Article 19 (2013). ‘Internet intermediaries: Dilemma of Liability’ Available at: <https://www.article19.org/wp-content/uploads/2018/02/Intermediaries_ENGLISH.pdf>

Bankston, Kevin, Sohn, David, & McDiarmid, Andrew. (2012). Shielding the Messengers–Protecting Platforms for Expression and Innovation. *Center for Democracy and Technology*, *12*. Available at: <https://cdt.org/wp-content/uploads/pdfs/CDT-Intermediary-Liability-2012.pdf>

Electronic Frontier Foundation (EFF) (2015). 'The Manila Principles on Intermediary Liability Background Paper' Available at: <https://www.eff.org/files/2015/07/08/manila_principles_background_paper.pdf>

MacKinnon, Rebecca; Hickok, Elonnai; Bar, Allon; and Lim, Hai-in. (2015). Fostering Freedom Online: The Role of Internet Intermediaries. Other Publications from the Center for Global Communication Studies. Available at: <http://unesdoc.unesco.org/images/0023/002311/231162e.pdf>,

OAS (2011). 'Special Rapporteurship for Freedom of Expression' Available at: <https://www.oas.org/en/iachr/expression/showarticle.asp?artID=848>

Organisation for Economic Co-operation and Development (OCDE) (2010). OECD. March 2010. The Economic and Social Role of Internet Intermediaries. Available at: <www.oecd.org/internet/ieconomy/44949023.pdf>

Zingales, Nicolo (2015). 'The Brazilian approach to internet intermediary liability: blueprint for a global regime?'. Internet Policy Review, 4(4). DOI: 10.14763/2015.4.395 Available at: <https://policyreview.info/articles/analysis/brazilian-approach-internet-intermediary-liability-blueprint-global-regime>

49. Internet Safety/Security

The terms internet safety and internet security are closely connected. In the words of the European Commission (2020, 1), "Security is not only the basis for personal safety, it also provides the foundation for confidence and dynamism in our economy, our society and our democracy." Thus, internet safety/security are perceived as general policy issues. These are general policy concerns about worldwide challenges such as crime, health and safety on an individual level.

However, when the internet **infrastructure** is perceived as a critical environment, specific security-challenges are dealt with by the internet governance measures that are tailored to bring about guarantees surrounding the internet's technical functioning. Importantly, platforms and other service providers who are tasked with the provision of **access** to its users play a general role in this sense.

As introduced above, the concept of "internet safety/security" evokes other terms commonly understood as threats in the digital environment. For this reason, many national legislations on regulating misconduct are relevant for this topic. This is reflected in the approach taken in the *Convention on Cybercrime* (Council of Europe, 2001), which aims to have its members maintain, update or introduce substantive criminal law measures to deal with the problem of cybercrime. Regarded as the first international treaty on this topic, this Convention is widely used as a reference for developing law and policy (see for example the site on EU Law on Cybercrime). In pursuance of developing these solutions, in recent years several soft-law measures have been taken, such as the "EU Code of conduct on countering illegal hate speech online" (European Commission 2016). The EU took the initiative for seeking more proactive responses and accountability from major private internet-companies. As pointed out by the European Commission (2020, 13): "The latest evaluation shows that companies assess 90% of flagged content within 24 hours and remove 71% of the content deemed to be illegal hate speech. However, the platforms need to improve further transparency and feedback to users and to ensure consistent evaluation of flagged content."

With threats being often described as “hybrid” in form, in recent years, many elements of responsiveness to the issues surrounding internet safety and security are leading to the regular renewal of or the adoption of new strategies by different governments all over the world. As an example, the latest “National Cyber Strategy” (2018) in the US presented the intention to increase the imposition of “costs” on all kinds of different players in the internet environment, in order to “to deter malicious cyber actors and prevent further escalation” (id, p. 2). Examples of the implementation of this strategy by the US government are the executive orders by president Trump seeking to restrict the possibilities of users to access several (social) media and mobile applications (such as WeChat and TikTok; see the orders of The White House, 6 August 2020). The US government’s orders against “WeChat” and “TikTok” were motivated as being based on national cybersecurity-concerns. A detailed plan to install specific prohibitions was announced in order to specifically limit the offer of the targeted apps in US stores. However, both of these executive orders have led to further administrative actions and judicial (counter-)measures limiting their execution and the reopening of the access to the US market. Another approach that was recently initiated by the Russian Government takes the form of establishing a network that can operate alongside the WWW in case of an attack. This so-called RuNet is envisioned as an obligation for internet providers in Russia by implementing new rules established by the "the Federal law No. 90-FZ on Amendments to Certain Legislative Acts of the Russian Federation (in terms of ensuring the safe and sustainable functioning of the Internet in the territory of the Russian Federation").

It is likely that more of such new technical safety measures or even new protocols for the technical functioning of the internet will be brought up as policy choices in response to safety and security threats, also on the international level such as ITU. For example, such topics are high on the list for discussion at the World Telecommunication Standardization Assembly – 2020 (such as a "New IP" protocol system persistently promoted by Huawei and the Chinese Government).

Major policy areas that are related to this term can be found in connection to all the national legal interventions that are responsive to fundamental (human) rights (such as children's rights) as well as the general issues of cybercrime already pointed to above (such as racism, xenophobia, hate crime, theft, etc). These areas get regular attention in terms of proposals for common legal and regulatory responsibilities, which would be applicable across the internet. In this regard, a notable, legislative initiative is the UK's recently proposed "online safety laws" as put forward in the "Online Harms White Paper" (2019), which proposed a broad (statutory) "**duty of care**".

References

Huawei and the Chinese Government, 'New IP protocol system' [2019] Available at: https://www.itu.int/md/T17-TSAG-C-0083

World Telecommunication Standardization Assembly, [2020]. Available at: <https://www.internetsociety.org/resources/doc/2020/itu-wtsa-2020-background-paper/>

Russian Federation, 'Federal law No. 90-FZ on Amendments to Certain Legislative Acts of the Russian Federation' [2019] Available at: http://publication.pravo.gov.ru/Document/View/0001201905010025?index=0&rangeSize=1;

ICNL, ' Russia' (ICNL 2020) <https://www.icnl.org/resources/civic-freedom-monitor/russia>

European Commission, 'EU Law on Cybercrime'. Available at: https://ec.europa.eu/home-affairs/what-we-do/policies/cybercrime_en

White House, 'National Cyber Strategy', [2018]. Available at: https://www.whitehouse.gov/wp-content/uploads/2018/09/National-Cyber-Strategy.pdf

Convention on Cybercrime adopted 23 November 2001, entered into force 1 July 2004, ETS No.185, available at <www.coe.int/en/web/conventions/full-list/-/conventions/treaty/005 >

European Commission. 2016. "The EU Code of conduct on countering illegal hate speech online." Availabe at < https://ec.europa.eu/info/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/eu-code-conduct-countering-illegal-hate-speech-online_en>

European Commission. 2020. "Communication on the EU Security Union Strategy." Available at <https://ec.europa.eu/info/sites/info/files/communication-eu-security-union-strategy.pdf>.

UK Department for Digital, Culture, Media & Sport and Home Office. 'Online Harms White Paper', [2020] Available at < https://www.gov.uk/government/consultations/online-harms-white-paper/online-harms-white-paper>

UK Government, 'Press release', [2019] <https://www.gov.uk/government/news/uk-to-introduce-world-first-online-safety-laws>

'Symposium: Online Harms White Paper', [2019] Journal of Media Law, Vol 11, issue 1; Available at: https://www.tandfonline.com/toc/rjml20/11/1?nav=tocList&)

The White House/US President Trump, Executive Order on Addressing the Threat Posed by WeChat, 6 August 2020. Available at <https://www.whitehouse.gov/presidential-actions/executive-order-addressing-threat-posed-wechat/>.

The White House/US President Trump, Executive Order on Addressing the Threat Posed by TikTok, 6 August 2020. Available at <https://www.whitehouse.gov/presidential-actions/executive-order-addressing-threat-posed-wechat/>.

50. Interoperability

Interoperability is the ability to transfer and render useful data and other information across systems, applications, or components. The combination of transmission and analysis involves several layers of the so-called Open Systems Interconnection model (OSI model), requiring the achievement of various levels of interoperability. At a minimum, one should distinguish the lower and the upper layer, pointing to a division between infrastructural interoperability and data interoperability.

At the infrastructure (lower) layer, interoperability is achieved through the use of common protocols for the conversion, identification and logical addressing of data to be transmitted over a network. The most common standards in this layer are Ethernet and TCP/IP. Protocols are also used for communication between computer programmes over telecommunications equipment, through common languages such as HTTP for web content, and SMTP, IMAP and POP3 for emails.

At the application (upper) layer, interoperability is attained by reading and reproducing specific parts of computer programs, called interfaces, which contain the information necessary to "run" programs in a compatible format. However, different interfaces are needed depending on who actually "runs" the program1550: if it is from the perspective of the user/consumer of the computer program, user interfaces are relevant to the ex- tent that they enable him or her to visualize and deploy a specific set of commands or modes of interaction with the program, that can potentially be replicated into another (different) application. Importantly, although this kind of interoperability can increase a program's utility to the user, it is not required for the purpose of its technical functioning. Most choices for user interfaces are indeed dictated not so much by functional elements of the program, as by the pursuit of the goals of user friendliness, aesthetical appeal and promotion of brand-specific features.

In a data-driven economy, the importance of open technical standards can hardly be overstated: common technical and legal protocols for interconnection and data processing enable communication and portability, thereby stimulating innovation and promoting competition of services within a given technological paradigm.

The degree to which such standards are truly open is likely to be a significant point of contention among different types of businesses. Granting automatic access to technology implementers can affect a technology provider's ability to appropriate the value of its innovation in downstream markets; this in turn may lead important players in the industry to not only abstain from standard-setting efforts, but also implement strategies aimed at foreclosing interoperability with competitors' technologies (horizontal interoperability) and preventing third parties from building on top of their technology (vertical interoperability).

From the perspective of the developer of a computer program, the relevant interfaces for interoperability are the Application Programming Interfaces, i.e. any well-defined software interfaces which define the service that one component, module or application provides to other software elements. However, interoperable APIs do not necessarily imply the ability of either users or developers to meaningfully relate the outputs of interoperable computer programs, unless they are expressed in the same language (most commonly, JPEG for images, HTML for webpages, PDF for documents and MP3 for music). This can be achieved through the so called “data interfaces”, which are responsible for restoring and retrieving data in a specific format.

References

A. Van Rooijen, The Software Interface between Copyright and Competition Law: A Legal Analysis of Interoperability in Computer Programs (Kluwer Law, Alphen aan den Rijn 2010)

C. R. B. de Souza et al, “Sometimes You Need to See Through Walls- A Field Study of Application Programming Interfaces”, in Computer supported cooperative work, ACM Press 2004, p. 63-71

IEEE Guide to the POSIX Open System Environment (OSE), in IEEE Std 1003.0-1995 , vol., no., pp.0_3-, 1995, doi: 10.1109/IEEESTD.1995.81544.

51. Liability

Liability refers to a legally enforceable responsibility for a harmful event. Liability can be civil or criminal, which are fundamentally different concepts in their origin and nature: the former implies a responsibility from a financial perspective, which can be explicitly foreseen by a statute but also be the result of contractual arrangements; whereas the latter implies the commission of a criminal offence, and thus necessarily depends on the existence of a primary norm in the legal system establishing a prohibited conduct (either active or passive). The primary goal of these two liabilities is also different, as the former is aimed to ensure compensation, while the latter aims at deterrence.

In the context of platforms and more generally of intermediaries, an important distinction should be made between primary and secondary liability: while the former requires the violation of a specific rule of conduct directed to the intermediary, the second arises from duties that are triggered by the conduct of third parties. However, there is some confusion in the use of these terms across jurisdictions, as the dividing line between primary (also known as “direct”) and secondary (also known as “indirect”) liability is not always clear-cut: several statutes attribute primary liability on intermediaries for the failure to prevent, or the implicit authorization of, third party conduct (for instance, the doctrine of “authorization” in Australian and UK copyright law).

Terminological clarifications aside, two main justifications are used to impose secondary liability: participation and relationship. The latter is the one that can be most easily circumscribed, as it requires the existence of a specific relationship between the primary and secondary infringer, where the latter benefits from the harm and is sufficiently close in relationship to the primary wrongdoer. The best example of this is employment relationships (based on the principle of *respondeat superior*), but the same rationale has been extended under the doctrine of vicarious liability to a range of scenarios where the secondary infringer had the right and ability to control the conduct of the primary infringer, and it is deriving financial benefit. It should be noted that the liability exemptions for hosts in the US Digital Millennium Copyright Act and in the European E-Commerce Directive specifically leave out circumstances where an intermediary had authority or control over a third party activity, but the former also includes the additional requirement of deriving no financial advantage from such activity.

Participatory liability depends on the existence of a requisite degree of participation, which can range from mere facilitation to purposeful combination. In the UK, for instance, three types of participatory liability have been recognized: combination, authorization and inducement liability. Combination is the most intuitive scenario, where two or more parties have a common design or enterprise, and the infringing acts are in pursuance or furtherance of that. Initially, this was interpreted strictly at common law to require an identity of concerted action to a common end; however, more recent cases adopted a more liberal approach requiring a combination (even tacit) to secure the doing of acts which eventually prove to be infringements. The doctrine has been used in *Dramatico Entertainment Ltd v British Sky Broadcasting*, [2012] EWHC 268 (Ch), [2012] 3 CLMR 14 to find that the Pirate Bay website facilitates its users’ infringement of copyright, on grounds that there were hardly any lawful uses of the site. Most recently, three limits to its expansion were identified in *Fish & Fish Ltd v Sea Shepherd UK* [2013] EWCA Civ 544, [2013] 3 All ER 867. First, common design will not be assumed simply because a person sells a product to another knowing that it is going to be used to commit a tort; second, there needs to be something more than just a close relationship between the parties; and third, approval of a person's plan will not be sufficient in itself to give rise to common design.

Authorization liability implies a different form of participation consisting of (tacit or explicit) permission, or possibly an order, from a person having (or purporting to have) authority over the "immediate" wrongdoer. It requires sufficient knowledge of the relevant circumstances and the acts committed (or to be committed) by the primary infringer. In the UK, a court established in *Newzbin (No. 1)* [2010] FSR 21 [90] that its application depends on a number of factors, such as the nature of the relationship between the alleged authoriser and the primary infringer, whether the equipment or other material supplied constitutes the means used to infringe, whether it is inevitable it will be used to infringe, the degree of control which the supplier retains and whether he has taken any steps to prevent infringement.

Finally, inducement liability requires a further degree of participation, including acts such as persuasion and encouragement, for an infringing purpose. For instance, it was recently the doctrine on the basis of which a bookmaker was imputed of secondary copyright infringement for providing its customers with a link where they could find a database of infringing information concerning live football matches. In US law, this is recognized in the field of patents and copyright, for those who distribute a device with the object of promoting an infringing use; however, such intent must be shown by clear expression or other affirmative steps taken to foster infringement, which is not always easy for a plaintiff. Famously, in *Metro-Goldwyn-Mayer Studios Inc. v. Grokster* 545 U.S. 913 (2005), the US Supreme Court found inducement liability for peer to peer software offered by Grokster on the basis of three key factors: (1) efforts to satisfy a known demand for infringing content; (2) an absence of design efforts to diminish infringement; and (3) Grokster's financial benefit from the activity.

Both in criminal and in civil liability cases, a defendant can be subjected to an injunction, i.e. a court order that imposes a given conduct – be it an action or an inaction. This category of liability should be distinguished because, although it may arise in connection with the existence of intermediary liability, the cause of action is independent: the liability is attached to the failure of complying with the judicial order, rather than the responsibility for a third-party conduct. It has thus been suggested that the appropriate term is one of **accountability**, as discussed in that definition.

References

See Paul Davies, *Accessory Liability* (Hart Publishing, 2015); Hazel Carty, 'Joint Tortfeasance and Assistance Liability', (1999) 19 Legal Studies 489.

Richard Arnold and Paul S Davies, 'Accessory liability for intellectual property infringement: the case of authorization', Law Quarterly Review 442 (2017)

Graeme Dinwoodie, ‘A Comparative Analysis of Secondary Liability of Online Service Providers’ in Graeme Dinwoodie (ed.), *Secondary Liability of Internet Service Providers* (Springer 2017).

Jaani Riordan, A Taxonomy of Intermediary Liability’, in G. Frosio (ed.), The Oxford Handbook on Intermediary Liability (OUP, 2020).

Christina Angelopoulos, Harmonizing Intermediary Copyright Liability in the EU: A Summary, in G. Frosio (ed.), The Oxford Handbook on Intermediary Liability (OUP, 2020).

Husovec, Martin, Accountable, Not Liable: Injunctions Against Intermediaries (May 2, 2016). TILEC Discussion Paper No. 2016-012, Available at SSRN: https://ssrn.com/abstract=2773768 or http://dx.doi.org/10.2139/ssrn.2773768

52. Marketplace

A marketplace is a web-based service enabling the sale or goods or the provision of services by third party vendors. While the marketplace operator can also provide goods (of which it has full ownership) or services (through its own employees), this activity amounts merely to at distance/online sales. The marketplace operators process themselves or have built-on tools to process the payment of the good or service by users in favor of the third-party vendors. The marketplace *may* or *may not* collect a fee for its intermediation service. Vendors can be businesses or consumers. The target users can similarly be both coming from business or retail backgrounds.

Within marketplaces, sharing economy platform operators facilitate transactions between providers and users. The transaction can relate to the temporary use of/access to a good intermediation services, or any service between providers acting outside their professional activity and users. The range of this notion is largely challenged in the literature and should only encompass the most narrow sense (else, it would equal to the notion of "marketplace").

Marketplaces act as "points-of-control". Their influence on the underlying supply of goods or services is thus questioned under vertical restraints theories in competition law. Because they are points of control, policy makers can also use them to ensure the compliance with certain economic policies, especially in tax matters (e.g. by imposing reporting duties).

The third-parties providing goods or listing offers for services on these marketplaces can either be business or individuals. It widens therefore the opportunities for individuals to offer goods and services on a frequent basis, without having to enter within the scope of consumer protections legislations. Indeed, consumer protection legislation often require the service provider or the seller to be a business. Two situations qualified as unfair may thus arise because the marketplace hides the identity of the third parties as well as the frequency of the transactions occurring within that seller/provider. On the one hand, businesses will try to pass off as individuals selling goods or offering services without consumer protection warranties. On the other hand, individuals may grow an activity as large as a business and evade legislations applicable to that business (in terms of consumer protection but also in terms of licencing) - creating thus a situation of unfair competition. This is the crux of many judicial challenges to ensure parity between "brick-and- mortar" and "click-and-mortar" businesses (especially in the so-called sharing economy).

Because goods and services are offered by third-parties on marketplaces, the issue of the liability of the platform is often raised, notably in terms of intellectual property rights where the platforms have due diligence duties to ensure that trademark and copyrights protections are not infringed (see **notice-and-takedown**). Additionally, policymakers, incumbent marketplaces who feel unduly harmed by the unfair competition of click-and-mortar businesses also seek to find the platforms liable for other forms of illegal goods, services or activities (e.g. with regards to licence requirements). The success of this assertion varies largely from country to country.

References

COM 356 – ‘A European agenda for the collaborative economy’, [2016].

OECD. "The Role of Internet Intermediaries in Advancing Public Policy Objectives." In OECD Publishing, [2011].

Calo, Ryan, and Alex Rosenblat, ‘The Taking Economy: Uber, Information, and Power’, [2017] Columbia Law Review 117.

Edelman, Benjamin G, and Abbey Stemler. ‘From the Digital to the Physical: Federal Limitations on Regulating Online Marketplaces.’ [2019] Harvard Journal on Legislation, Forthcoming, 18-063.

Balfour, Abigail W. ‘Where One Marketplace Closes, (Hopefully) Another Won't Open: In Defense of Fosta.’ [2019] Boston College Law Review 60, no. 8, 2475.

Hoppner, Thomas, and Philipp Westerhoff. ‘The Eu’s Competition Investigation into Amazon’s Marketplace.’ [2018] Hausfeld Bulletin, no. Fall (2018).

Stemler, Abbey. ‘Feedback Loop Failure: Implications for the Self-Regulation of the Sharing Economy’, [2017] Minn. J. L. Sc. Tech. 18, 674, 712.

Lobel, Orly. ‘The Law of the Platform.’ [2016] Minn. L.R. 101, 87, 166.

Goldman, Eric. ‘An Overview of the United States’ Section 230 Internet Immunity.’ [2020] In Oxford Handbook of Online Intermediary Liability, edited by Giancarlo Frosio: Oxford, UK: Oxford University Press.

Goldman, Eric. "The Complicated Story of Fosta and Section 230." [In eng]. First Amendment Law Review 17 (2018-2019 2018): 279-93.

Chander, Anupam. "How Law Made Silicon Valley." Emory Law Journal 63 (2014): 639-94.

Hatzopoulos, V., & Roma, S. (2017). Caring for sharing? The collaborative economy under EU law. *Common Market Law Review*, *54*(1)

Hatzopoulos, V. (2019). After Uber Spain: the EU’s approach on the sharing economy in need of review. *Case Comment. European Law Review*, *44*.

53. Media Pluralism

At its core, media pluralism references to the kind of diversity of media sources and opinions available to any given audience. More specifically, Reporters Without Borders (RSF, 2016) stresses that media pluralism can either refer to "a plurality of voices, of analyses, of expressed opinions and issues (internal pluralism), or a plurality of media outlets, of types of media (print, radio, TV, or digital), and coexistence of privately owned media and public-service media (external pluralism)." Media pluralism is imperative to a healthy, functioning democracy, as it fosters an information ecosystem that enables citizens to access a range of opinions, confront ideas, make informed choices, and conduct their life freely. Yet, consumption habits, changing economic models, and technical systems are threatening media pluralism around the world. Media consolidation and concentration (Wikipedia) are also a key threat. As fewer individuals or organisations control increasing shares of mass media producers, editorial independence, narrative diversity, and public-interest reporting are much more limited and controlled.

In the age of digital and technological convergence, both internal pluralism and external pluralism are relevant to Internet governance discussions. When taken together, they reflect myriad digital policy areas – specifically access to information media sustainability. Diversity – ranging from gender perspectives, to the voices of minorities and marginalised groups – is a crucial component of internal pluralism, for instance. Internal media pluralism is also inextricably linked to bridging the digital divide(s) as well as encouraging skill development via digital media literacy and local capacity development. New technologies pose a threat to internal plurality as well, specifically the phenomenon of artificial intelligence (AI) applications being used to replace editors and content curators. Digital platforms have an important responsibility to promote and ultimately preserve internal media pluralism in their role as a primary gatekeeper (Helberger et al., 2015) to information diversity. Key recommendations (Global Forum For Media Development, 2020) to safeguarding this role include remodeling platform algorithms and moderation practices, as well as reversing commercial incentives that discriminate against journalism and news media.

On the other hand, external pluralism is intrinsically tied to discussions around digital markets and media market failure (Pickard, 2019), competition and innovation, media funding, and zero rating. Dominant Internet business models continue to place strain (Chicago Booth, 2019) on both legacy and new/digital media outlets, which in turn, makes local and regional media ecosystems more fragile, more prone to closures and the creation of "news deserts," (UNC) and more susceptible to media capture (Center for International Media Assistance) – a form of governance failure that occurs when the news media advance the commercial or political concerns of state and/or non-state special interest groups controlling the media industry instead of holding those groups accountable and reporting in the public interest. Looking ahead, media plurality and platform governance go hand-in-hand. Recognising how vital media plurality is and ultimately working to safeguard it is a critical endeavour going forward.

References

Reporters Sans Frontieres, ' Contribution to the EU public consultation on media pluralism and democracy' (European Commission 2016) < https://ec.europa.eu/information_society/newsroom/image/document/2016-44/reporterssansfrontiers_18792.pdf>

Wikipedia, 'Concentration of media ownership' (Wikipedia) < https://en.wikipedia.org/wiki/Concentration_of_media_ownership>

Natali Helberger, Katharina Kleinen-von Königslöw and Rob van der Noll, ' Regulating the new information intermediaries as gatekeepers of information diversity' [2015] VOL. 17 NO. 6, © Emerald Group Publishing Limited, ISSN 1463-6697 50, 71

Global Forum For Media Development, ' Joint Emergency Appeal For Journalism And Media Support' (Global Forum For Media Development 2020) < https://gfmd.info/emergency-appeal-for-journalism-and-media-support-2/>

Victor Pickard, ' Public Investments for Global News' (Center for International Governance Innovation 2019) < https://www.cigionline.org/articles/public-investments-global-news>

Stigler Center News, ' Stigler Committee on Digital Platforms: Final Report' (Chicago Booth 2019) < https://www.chicagobooth.edu/research/stigler/news-and-media/committee-on-digital-platforms-final-report>

UNC, ' Do You Live In A News Desert?' (UNC) < https://www.usnewsdeserts.com/>

Center for International Media Assistance, ' What is Media Capture?' (Center for International Media Assistance) < https://www.cima.ned.org/resources/media-capture/>

Article 19, ' Media Freedom' (Article 19) < https://www.article19.org/issue/media-freedom/>

Center for International Media Assistance, < https://www.cima.ned.org/>

Centre for Media Pluralism and Freedom (CMFP)

CMPF, ' Media Pluralism Monitor' (CMPF) < https://cmpf.eui.eu/media-pluralism-monitor/>

Council of Europe, ‘Recommendation CM/Rec(2018)11 of the Committee of Ministers to Member States on Media Pluralism and Transparency of Media Ownership’. Available at: https://rm.coe.int/1680790e13

IGF, ' Dynamic Coalition on the Sustainability of Journalism and News Media ' (IGF 2019) < https://groups.io/g/dc-sustainability>

GFMD, ' Internet Governance' (GFMD). Available at: https://gfmd.info/internet-governance/

Media Diversity Institute. Availble at: https://www.media-diversity.org/

Government of the Netherlands , ' The concept of pluralism: media diversity' (Media Monitor) < https://www.mediamonitor.nl/english/the-concept-of-pluralism-media-diversity/>

UNESCO, ' Media Pluralism and Diversity' (UNESCO) < https://en.unesco.org/themes/media-pluralism-and-diversity>

UNESCO, *World trends in freedom of expression and media development: global report 2017/2018* (1st, UNESCO, 2018). Available at: https://unesdoc.unesco.org/ark:/48223/pf0000261065

54. Microtargeting

(I) Targeting is a practice whereby online content, typically advertising content, is distributed towards particular audiences based on their personal data. In the words of William Gorton, microtargeting involves ‘creating finely honed messages targeted at narrow categories of voters’ based on data analysis ‘garnered from individuals’ demographic characteristics and consumer and lifestyle’ (Gorton, 2016). Targeting is often closely associated with personalization, and the terms are often used interchangeably. The prefix *micro-* is used to indicate that a highly specific audience is being targeted, although the precise criteria for this designation are rarely made explicit.

The most popular and influential microtargeting services are those offered by major online platforms such as Google and Facebook, but it is not limited to these services. Indeed, microtargeting can also be done *offline*; many offline campaign activities, such as door-to-door canvassing, pamphleteering and telephone banking can be targeted with the help of personal data, much in the same way as online advertising.

When microtargeting relates to political advertisements, it is referred to as political microtargeting, which Ira Rubinstein describes as form of ‘direct marketing in which political actors target personalized messages to individual voters by applying predictive modelling techniques to massive troves of voter data’ (Rubinstein, 2014). It is worth noting, however, that the concept of ‘political’ advertising is also ambiguous and continues to be contested, with some focusing on a narrower category of election campaign ads and others extending the term to cover all political ‘issues’ - which is itself a highly amorphous category (Leerssen et al, 2019). This same ambiguity about the boundaries of the political also arises in the context of microtargeting.

The threshold where targeting becomes microtargeting is not always clear. One way to distinguish micro-targeting is by reference to the size of the audience targeted. Another is to focus on the granularity of the personal data involved. Along these lines, Tom Dobber, Ronan O’Fathaigh and Frederik Zuiderveen Borgesius propose, in the context of political advertising, that ‘micro-targeting differs from regular targeting not necessarily in the size of the target audience, but rather in the level of homogeneity, perceived by the political advertiser’ (Dobber, Fahy and Zuiderveen Borgesius, 2019). In this reading, targeting an entire neighbourhood with a single message constitutes *regular* targeting, whereas tailoring different messages to different users within the neighbourhood, based on their personal data profiles, constitutes *micro*targeting. Overall, the available literature suggests a sliding scale between general and micro-targeting, rather than a strict binary.

(II) ‘Microtargeting’ is a novel concept from communications science without any clearly defined legal meaning. Although various laws affect the practice of targeted advertising, including data protection laws and campaign finance laws (Dobber, O’Fathaigh and Zuiderveen Borgesius, 2019), these have not historically relied explicitly on the concept of 'targeting' or 'microtargeting' in doing so. Only recently has the concept made its first appearance in official policymaking.

In its June 2020 resolution on competition policy, the European Parliament proposed a ban on micro-targeting performed by online platforms. The report "calls on the Commission to ban platforms from displaying micro-targeted advertisements and to increase transparency for users" (European Parliament 2020). A further operationalization of this concept has not (yet) been proposed, although accompanying statements from the amendment's author, Paul Tang MEP, appear to use microtargeting interchangeably with 'personalization' – suggesting a relatively low threshold that could potentially cover most if not all targeting practices involving personal data ("European Parliament wants to forbid personalised advertisements" 2020).

Occasionally, self-regulatory efforts by platforms and other online services also reference the concept of microtargeting. For instance, Google claimed that it had prohibited 'microtargeting' for political advertisements, by virtue of having restricted targeting options for these ads to a more limited selection: age, gender, and general location (Google 2019).

In the context of political advertising and campaigning laws, microtargeting practices are now under intense scrutiny in multiple jurisdictions, with reforms recently completed or ongoing in, *inter alia*, Germany, France, the United Kingdom, the Netherlands, Sweden, Ireland, the United States, and Canada (For an overview, see Van Hoboken et al 2019). The European Commission has also singled out microtargeting as a point of attention in the ongoing Digital Services Act reforms. Until now, the majority of these laws and proposals have focused on issues such as campaign finance and transparency, and have not (yet) tackled the legality of microtargeting as such.

References

Dobber, Tom, Ronan Ó Fathaigh and Frederik Zuiderveen Borgesius. 2019). "The regulation of online political micro-targeting in Europe". *Internet Policy Review* 8(4). Available at: https://policyreview.info/articles/analysis/regulation-online-political-micro-targeting-europe

European Parliament. 2020. Resolution of 18 June 2020 on competition policy – annual report 2019. 2019/2131(INI). Available at:https://www.europarl.europa.eu/doceo/document/TA-9-2020-0158_EN.html

"European Parliament wants to forbid personalised advertisements". 19 June 2020. *Paultang.nl*. Available at: https://paultang.nl/en/forbid-personalised-ads/

Google. 2019. "An Update on our political ads policy". *Google Official Blog*. Available at: https://blog.google/technology/ads/update-our-political-ads-policy

Gorton, William. 2016. "Manipulating Citizens: How Political Campaigns' Use Of Behavioral Science Harms Democracy". *New Political Science* 38(1).

Rubinstein, Ira. 2014. “Voter Privacy in the Age of Big Data” *Wisconsin Law Review* 5, pp. 861-936.

Leerssen et al, 2019. “Platform Ad Archives: Promises and Pitfalls”. *Internet Policy Review* 8(4). Available at: https://policyreview.info/articles/analysis/platform-ad-archives-promises-and-pitfalls

Van Hoboken, Joris, Naomi Appelman, Ronan Ó Fathaigh, Paddy Leerssen, Tarlach McGonagle, Nico van Eijk and Natali Helberger. 2019. The legal framework on the dissemination of disinformation through Internet services and the regulation of political advertising: A report for the Ministry of the Interior and Kingdom Relations. Institute for Information Law (IViR). Available at: https://www.ivir.nl/publicaties/download/Report_Disinformation_Dec2019-1.pdf

55. Moderation

Content moderation can be described as the result of editorial decisions made by the subject who govern the space where information is published. Moderation has also been defined as "the governance mechanisms that structure participation in a community to facilitate cooperation and prevent abuse." (Grimmelman, 2015).

Content moderation is not a novelty in the media sector. As content providers, traditional media outlets like televisions and newspapers have always selected the information to broadcast or disclose. This activity has also extended to the digital environment. Since the first online fora, we have seen how communities have moderated digital spaces to decide which content reflects the values or interest of the group without commercial purposes. In the last years, the commercial side of content moderation has evolved with online platforms, precisely social media, which have built a bureaucracy to moderate content (Klonick, 2019). This activity has been defined as "the screening, evaluation, categorization, approval or removal/hiding of online content according to relevant communications and publishing policies […] to support and enforce positive communications behaviour online, and to minimize aggression and anti-social behaviour." (Flew et al., 2019). This amount of content flowing on social media' spaces is not free but subject to a wide range of practices applied by platforms to manage content posted by their users (Niva Elkin-Koren & Maayan Perel, 2019a). Just in the case of Facebook, the amount of post moderated in different areas of the world is on a scale of billions each week.

While some practices intend to optimize the matching of content with the users who view it and would potentially engage with it, other practices intend to ensure that content complies with appropriate norms (Niva Elkin-Koren & Maayan Perel, 2019b). Social media decide how to organize users' news feed or set their recommendation system to target certain categories of users (i.e. soft moderation). Together with such activities, social media make editorial decisions which can also lead to the removal of online content to ensure respect and enforce community's rules (i.e. hard moderation). Content moderation decisions can be entirely automated, made by humans or a mix of them (Gorwa et al. 2020). While the activities of pre-moderation like prioritisation, delisting and geo-blocking are usually automated, post-moderation is usually the result of automated and human moderation. The massive amount of content to moderate explains why content moderation is usually performed by a mix of machines and human moderators which decide whether to maintain or delete the vast amount of content flowing every day on social media (Roberts, 2019).

Within this framework, social media platforms facilitate the global exchange of content generated by users, at gigantic scale while governing information flow online (Kaye, 2019). However, these characteristics are just one part of the jigsaw explaining the ability and reasons for platforms to discretionary establish how to carry out content moderation. Content moderation is the constitutional activity of social media (Gillespie 2019). The moderation of online content is an almost obligatory step for social media not only to manage removal requests but also 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. In other words, the activity of content moderation is performed to attract revenues by ensuring a healthy online community, protect the corporate image and show commitments with ethical values. Within this business framework, users' data are the central product of online platforms under a logic of accumulation (Zuboff, 2019).

In this scenario, content moderation produces positive effects for freedom of expression and democratic values. The organization, filtering and removal of content increases the possibilities for users to experience a safe digital environment without the interference of objectionable or harmful content. At the same, content moderation negatively impacts on the right to freedom of expression. Since social media can select which information deserves to be maintained and deleted according to standards based on the interest to avoid any monetary penalty or reputational damage. Such a situation is usually defined as collateral censorship (Balkin, 2014). Scholars have observed that 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 (Pasquale, 2016). As underlined, immunity allows Internet intermediaries 'to have their free speech and everyone else's too' (Tushnet, 2008). Moreover, an extensive activity of content moderation influences even the right to privacy and data protection. Indeed, users could fear to be subject under a regime of private surveillance over their information and data. It is worth observing that, in the last case, even the right to free speech is involved due to the users' concern to be monitored through the information they publish.

More broadly, content moderation challenges also democratic values, such as the principle of the rule of law, since social media autonomously determine how freedom of expression online is protected on a global scale without any public safeguard (Suzor, 2020). 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. Since online platforms are private businesses, they would naturally tend to focus on minimising economic risks rather than ensuring a fair balance between fundamental rights when moderating content (De Gregorio, 2019b). The international relevance of content moderation can be understood even by looking at how this activity has led to escalating violent conflict in countries like Myanmar or Sri Lanka, so that some States decided to shut down social media as increasingly happening in African countries.

Addressing the challenges of content moderation without undermining its social relevance for the digital environment is one of the primary points from a policy perspective. In making decisions on online content, social media platforms apply a complex system of norms, driven by consumption, commercial interests, social norms, liability rules and regulatory duties, where each set of norms may interact with others (Belli and Zingales, 2017). Scholars have mostly proposed to protect the system of immunity (Keller, 2018) or reinterpret its characteristics (Bridy, 2018), building an administrative monitoring-and-compliance regime (Langvartd, 2017), or introducing more safeguards in the process of moderation (De Gregorio 2020a; Bloch Wehba 2020). In other words, the focus would move from liability to responsibility. To achieve this purpose, transparency and accountability safeguards could help to understand how speech is governed behind the scenes without overwhelming platforms with disproportionate monitoring obligations.

References

Balkin, Jack M., ‘Old-School/New-School Speech Regulation’ (2014) 127 Harvard Law Review 2296;

Belli, Luca and Zingales, Nicolo (eds), Platform Regulations. How Platforms are Regulated and How They Regulate Us (FGV Rio 2017);

Bridy, Annemarie, ‘Remediating Social Media: A Layer-Conscious Approach’ (2018) 24 Boston University Journal of Science and Technology Law 193.

Bloch-Wehba, Hannah, ‘Automation in Moderation’ SSRN (29 January 2020) https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3521619

De Gregorio, Giovanni (a), ‘Democratising Content Moderation: A Constitutional Perspective’ (2020) 36 Computer Law & Security Review.

De Gregorio, Giovanni (b), ‘From Constitutional Freedoms to the Power of the Platforms: Protecting Fundamental Rights Online in the Algorithmic Society’ (2019) 11(2) European Journal of Legal Studies 65:

Elkin-Koren, Niva and Perel, Maayan (a), ‘Algorithmic Governance by Online Intermediaries’ in Eric Brousseau et al. (eds), *Oxford Handbook of Institutions of International Economic Governance and Market Regulation* (Oxford University Press 2019).

Elkin-Koren, Niva and Perel, Maayan (b), ‘Separation of Functions for AI: Restraining Speech Regulation by Online Platforms’ SSRN (22 August 2019) https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3439261;

Flew, Terry et al., ‘Internet Regulation as Media Policy: Rethinking the Question of Digital Communication Platform Governance’ (2019) 10(1) Journal of Digital Media & Policy 33;

Gorwa, Robert et al., ‘Algorithmic Content Moderation: Technical and Political Challenges in the Automation of Platform Governance’ (2020) 7(1) Big Data & Society 1;

Grimmelmann, James, ‘The Virtues of Moderation’ (2015) 17 Yale Journal of Law & Technology 42;

Gillespie, Tarleton, Custodians of the Internet. Platforms, Content Moderation, and the Hidden Decisions that Shape Social Media (Yale University Press, 2018);

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

Keller, Daphne, ‘Internet Platforms: Observations on Speech, Danger, and Money’ (2018) Hoover Institution's Aegis Paper Series, No. 1807;

Klonick, Kate, ‘The New Governors: The People, Rules, and Processes Governing Online Speech’ (2018) 131 Harvard Law Review 1598;

Langvardt, Kyle, ‘Regulating Online Content Moderation’ (2017) 106(5) Georgetown Law Journal 1353.

Pasquale, Frank, ‘Platform Neutrality: Enhancing Freedom of Expression in Spheres of Private Power’ (2016) 17 Theoretical Inquiries in Law 487;

Roberts, Sarah T., Behind the Screen. Content Moderation in the Shadows of Social Media (Yale University Press 2018);

Suzor, Nicolas, Lawless: The Secret Rules That Govern Our Digital Lives (Cambridge University Press 2019);

Tushnet, Rebecca, ‘Power Without Responsibility: Intermediaries and the First Amendment’ (2008) 76 The George Washington Law Review 986.

Zuboff, Shoshana, Surveillance Capitalism. The Fight for a Human Future at the New Frontier of Power (Public Affairs 2019).

56. Must-carry

Must carry legislation is one of the instruments used in the regulation of cable TV to promote diversity and ensure the broadcast of channels that would not normally be included in the operators' bundle (Perry, n.d.).

The must carry rules appeared in 1965, in the face of the expansion of the cable TV service, to avoid that the growing power of the cable TV providers ended up suppressing the local broadcasters (Valente, 2013). Historically seen as a guarantee that broadcasters would be distributed by cable television providers, the must carry rules became more complex over time.

In 1968, the U.S. Court of Appeal issued the first decision (Black Hills Video Corp. v. FCC, 1968) declaring the legitimacy of the must carry, according to which it understands that its rules preserved local broadcasting without thereby violating the freedom of expression guaranteed by the First Amendment to the United States Constitution. From then on, a longstanding debate was launched on the importance of must carry rules (Pieranti & Festner, 2008).

This discussion recalls the debates on net neutrality and the obligation for internet services providers to ensure that all information that runs over the network must be equally treated (Ballard, 2011). Here we can draw a parallel between the power of cable TV services and the power of ISPs, where both have the technical capacity and economic incentives to restrict their competitors and/or favor their own business or partners (Belli, 2016). That is why the debates on freedom of expression and monopolistic tendencies are so inherent in both must carry discussions and those of network neutrality (Patrick & Scharphorn, 2015).

References

Belli, Luca. (2016). 'Net neutrality, zero rating and the Minitelisation of the internet'. *Journal of Cyber Policy, 2(1), 96–122.* doi:10.1080/23738871.2016.1238954

Pieranti, Octávio & Festner, Susana (2008). 'Estudo comparativo de regras de Must Carry na TV por assinatura'. Brasília: ANATEL.

Valente, Jonas (2013). 'Regulação democrática dos meios de comunicação'. São Paulo: Fundação Perseu Abramo.

Ballard, Tony (2011). 'Public service broadcasting: Net neutrality and the 'must carry' rules' Available at: <https://www.harbottle.com/title-215/>

Perry, Audrey (n.d.). 'Must-Carry rules' Available at: <https://www.mtsu.edu/first-amendment/article/1000/must-carry-rules>

Patrick, Andrew, & Scharphorn, Eric (2015). 'Network Neutrality and the First Amendment'. *Mich. Telecomm. & Tech. L. Rev.*, *22*, 93. Available at: <https://repository.law.umich.edu/cgi/viewcontent.cgi?article=1209&context=mttlr>

57. Non-discrimination

The term “non-discrimination” has a very specific definition and understanding under international law, however it is also used in a different sense in the context of online platforms. This entry first looks at the agreed international definitions of the term, as found in relevant legal instruments, before turning to other uses.

(i) “Non-discrimination” under international (human rights) law

The principle of non-discrimination - and, specifically, the prohibition of discrimination - has been translated into a number of international human rights instruments, most notably the International Covenant on Civil and Political Rights (ICCPR). While the ICCPR and other international human rights instruments (such as the International Covenant on Economic, Social and Cultural Rights) often prohibit discrimination in the enjoyment of the rights that they set out, the ICCPR also provides a standalone prohibition of discrimination through Article 26 which provides that: “the law shall prohibit any discrimination and guarantee to all persons equal and effective protection against discrimination on any ground such as race, colour, sex, language, religion, political or other opinion, national or social origin, property, birth or other status”.

While “discrimination” is not defined in Article 26 itself, the UN Human Rights Committee (HRC) has provided guidance on the scope of the term in its General Comment No. 18 (UN, 1989). There, the HRC states that “discrimination” should be understood as including “any distinction, exclusion, restriction or preference which is based on any ground such as race, colour, sex, language, religion, political or other opinion, national or social origin, property, birth or other status, and which has the purpose or effect of nullifying or impairing the recognition, enjoyment or exercise by all persons, on an equal footing, of all rights and freedoms”. In setting out this definition, the HRC drew upon other international human rights instruments which do define the term, such as the Convention on the Elimination of All Forms of Discrimination against Women and the International Convention on the Elimination of All Forms of Racial Discrimination. The HRC also clarified that Article 26 of the ICCPR prohibits discrimination “in law or in fact” and “in any field regulated and protected by public authorities”.

In the context of online platforms, the right to non-discrimination could be breached, for example, as a result of content moderation policies which themselves treat different groups differently, or in their enforcement disproportionately affect users with a particular characteristic; or the use of algorithms which are biased on the basis of a user’s personal characteristics, leading to discriminatory outcomes.

(ii) Other uses of the term

Outside of international law, the term “non-discrimination” is most commonly used in the context of online platforms to refer to a legal obligation not to discriminate in the conditions or quality of the services and information that it provides. The European Commission, for example, defines the term “non-discrimination” as follows: “an obligation of non-discrimination ensures that an operator applies equivalent conditions in equivalent circumstances to other undertakings providing equivalent services, and provides services and information to others under the same conditions and of the same quality as it provides for its own services, or those of its subsidiaries or partners.” (European Commission). Examples of discrimination in this context would largely be for commercial reasons, and would include differential pricing arrangements or different treatment of traffic.

References

UN Human Rights Committee, ‘General Comment No. 18: Non-discrimination’, [1989].

European Commission, ‘Shaping Europe’s digital future, Glossary’. Available at: https://ec.europa.eu/digital-single-market/en/glossary

58. Notice

Notice is a term that refers to the disclosure to a particular stakeholder of a relevant fact or situation. To be deemed a valid notice, such disclosure should contain a sufficient level of detail for the recipient to understand that information revealed and investigate the facts or situation. There are also certain formalities or general requirements to be complied with when it comes to what are deemed as qualified notices for the purpose of specific legal processes. For instance, several contracts specify that a notice of termination must be sent in writing and within a specified period. Similarly, when it comes to the information that must be provided by data controllers to data subjects, article 12 (1) of the GDPR requires it to be provided in a concise, transparent, intelligible and easily accessible form, using a clear and plain language – especially when directed to a child.

One important type of notice is the one regarding changes to a previously agreed contract. Here, cases involving the credit card and telecommunications industries provide helpful insight as to how courts evaluate modifications of standard terms, also known as “contracts of adhesion”, as there is no individual negotiation. For example, in *Badie v. Bank of America*, 67 Cal. App. 4th 779 (Cal. Ct. App. 1998), where a bank attempted to modify credit card terms by adding an arbitration procedure where one was not already part of the contract terms, a US Court found that the offeree did not receive proper notice of the modification because the proposed change was printed on an insert with the monthly bill and nothing otherwise called the change to anyone’s attention. Other companies have found out the hard way that simply providing a complete set of the proposed revised terms, without any indication as to which terms had been changed, was not sufficient notice (see e.g. *DIRECTV, Inc. v. Mattingly*, 829 A.2d 626 (Md. 2003)). On the other hand, a company that prominently announced modified terms with its monthly bill, and provided an Internet address and telephone number where the customer could access the revised terms, was found to have successfully put the customer on notice of the changed terms. *Ozormoor v. T-Mobile USA, Inc.*, 2008 U.S. Dist. LEXIS 58725 (E.D. Mich. June 19, 2008).

The appearance and placement of the notice also is important. One company was unable to enforce a notice of a contract modification that was printed on its invoice where it was the fifth item on the second page of the invoice, in ordinary type. *Manasher v. NECC Telecom*, No. 06-10749,2007 U.S. Dist. LEXIS 68795 (E.D. Mich. Sept. 18, 2007). On the other hand, in *Briceno v. Sprint Spectrum, L.P.*, 911 So. 2d 176 (Fla. Dist. Ct. App. 2005), Sprint’s notice was enforceable where it printed “Important Notice Regarding Your PCS Service From Sprint” in bold letters immediately below the amount due on the invoice. The notice also prominently discussed the changes in the contract terms and provided both a telephone number and a website where the revised terms could be found.

US Courts also have looked at whether the modification has been accepted by the offeree. For example, in *Klocek v. Gateway, Inc.*, 104 F. Supp. 2d 1332 (D. Kan. 2000), the purchaser of a Gateway computer did not see Gateway’s standard terms (and was not provided notice about the terms) until the computer was shipped to the purchaser and she opened the box. Gateway’s standard terms contained a number of provisions, including an arbitration clause. When Gateway moved to dismiss a class action lawsuit in light of the Federal Arbitration Act, the court refused to enforce the arbitration clause. The court found that the plaintiff offered to purchase the computer and Gateway accepted. Gateway's standard terms then became either an expression of acceptance or a confirmation of the offer under section 2-207 of the U.C.C. However, the court found that the rest of the provisions in Gateway's standard terms, including the arbitration clause, were not part of the original purchase agreement and were not enforceable.

More recently, in *Knutson v. Sirius XM Radio*, 771 F.3d 559 (9th Cir. 2014), the terms regarding an automobile's trial subscription to a satellite radio service were sent to the owner a month after the purchase of the automobile in an envelope marked "Welcome Kit." The Ninth Circuit refused to enforce the additional terms because there was no mutual assent to the terms. The Ninth Circuit found no evidence that the purchaser of the automobile knew that he had purchased anything from Sirius or was entering into a relationship with Sirius, let alone had agreed to the terms (which contained an arbitration clause). Therefore, continued use of the service by the purchaser did not manifest assent to the terms.

In the online context, courts that have addressed modifications generally have respected these traditional contract principles and have held that attempted modifications are unenforceable when the person to whom the modification is offered has no reason to know of the proposed changes to the agreement. As a result, online contract modifications tend to fall for failure to satisfy the notice requirement.

In evaluating online contract modification, courts have paid close attention to the differences between electronic and face-to-face or paper communications. This is a refreshing development, given that this is not always the case in opinions addressing online contract formation in the first instance. The opinion in *Campbell v. General Dynamics*, 407 F.3d 546 (1st Cir. 2005), a dispute involving an attempted modification of an employment handbook, provides an example of judicial awareness that electronic messages can get lost in the electronic shuffle. In *Campbell*, an employer attempted to modify an employment handbook by sending a mass company-wide e-mail message containing hyperlinks to the proposed changes to its employees. One of the proposed modifications was a binding arbitration clause. In holding that the modification was not effective, the court focused on the expectations of the employee receiving the modification offer. Given that the mass e-mail message did nothing to communicate its importance and that employment changes at General Dynamics were usually communicated in person by means of a signed writing, the court held that the attempted modification was not binding.

The communicative value of online interaction similarly influenced the Ninth Circuit in holding that the attempted modification in *Douglas v. U.S. District Court,* 495 F.3d 1062 (9th Cir. 2007), was ineffective. The dispute in that case arose when a phone service provider changed its online terms to add new service charges, a new arbitration clause, and a class action waiver. It did so without notifying its customers of the changes and simply posted the changes to its website. The plaintiff had agreed to automatic billing and therefore had little reason to visit the website on a regular basis. After the district court found the arbitration clause enforceable, the Ninth Circuit reversed, finding that the subscriber had not been given notice of the changes. The Ninth Circuit also felt strongly that parties to a contract have no obligation to check the terms on a periodic basis to learn whether they have been changed by the other side. This fact, plus the fact that the plaintiff would not have known where to find the changes to the terms of use even if he had visited the website, led the court to hold that the modifications were unenforceable.

The court in *Rodman v. Safeway, Inc.,* 2015 U.S. Dist. LEXIS 17523 (N.D. Cal. 2015), similarly refused to impose a duty on website users to continually check for changes to online terms. *Rodman* was another case in which the author of online terms of use posted changes to those terms on its website but made no attempt to notify its customers of the changes. The defendant attempted to justify its actions by highlighting a clause in its original terms of use that reserved the right to amend the terms at any time and imposed a duty on the customer to keep up with changes to the terms. Like the court in *Douglas*, the court in *Rodman* stressed that it is unreasonable to expect a customer to check a website regularly for changes to online terms. Moreover, the court, applying traditional contract doctrine, noted that a customer could not assent to future changes of which there was no reason to know would come.

In the context of platforms, the term notice is typically (but not only) used concerning the alleged illegality of a particular type of content or behavior, following which the platform may remove or disable access in accordance with a **notice and takedown**, a **notice and notice** procedure or some other standardized process. These notices can contain allegations either a violation of existing law or a violation of the platform´s terms of service. A civil society effort led by the Electronic Frontier Foundation (EFF) in 2014 established a number of minimum requirements for such notices, which they call “content restriction requests”, as part of the Manila Principles of Intermediary Liability. In particular, the Principles stipulate that a content restriction request pertaining to unlawful content must, at a minimum, contain the following:

1. The legal basis for the assertion that the content is unlawful.
2. The Internet identifier and description of the allegedly unlawful content.
3. The consideration provided to limitations, exceptions, and defences available to the user content provider.
4. Contact details of the issuing party or their agent, unless this is prohibited by law.
5. Evidence sufficient to document legal standing to issue the request.
6. A declaration of good faith that the information provided is accurate.

By contrast, a content restriction requests pertaining to an intermediary’s content restriction policies must, at the minimum, contain the following:

a) The reasons why the content at issue is in breach of the intermediary’s content restriction policies.
b) The Internet identifier and description of the alleged violation of the content restriction policies.
c) Contact details of the issuing party or their agent, unless this is prohibited by law.

Finally, notices in the context of online platforms may refer also to the disclosures made by platforms to users about the content that is prohibited, and the content from each specific user that is removed. This is a core principle of the Santa Clara Principles on transparency and accountability in content moderation, another civil society movement led by the EFF and a small group of organizations and advocates, establishing guidelines for content moderation that have been implemented by Reddit and endorsed by Apple, Github, Twitter, YouTube and other platforms (EFF, 2020).

The Principles require companies to provide detailed guidance to the community about what content is prohibited, including examples of permissible and impermissible content and the guidelines used by reviewers, and an explanation of how automated detection is used across each category of content. They also require minimum information to be included in the notices about why her post has been removed or an account has been suspended:

- URL, content excerpt, and/or other information sufficient to allow identification of the content removed.
- The specific clause of the guidelines that the content was found to violate.
- How the content was detected and removed (flagged by other users, governments, trusted flaggers, automated detection, or external legal or other complaint). The identity of individual flaggers should generally not be revealed, however, content flagged by government should be identified as such, unless prohibited by law.
- Explanation of the process through which the user can appeal the decision.

In addition to these requirements as to the form of notices, two important elements of the Santa Clara Principles concern the records of such notices: their availability in durable form accessible even if a user's account is suspended or terminated, and the presentation to users who flag content of a log of past content moderation requests they have submitted, along with the corresponding outcomes of the moderation processes. However, at the same time, it has been considered that the Principles fail to specifically address the peculiarities of certain practices, calling for an update (EFF, 2020). Some of the drafters of the Santa Clara Principles (Suzor, West, Quodling, and York 2020) noted that the implementation of the principles is unsatisfactory in certain respects, in particular due to (1) the prevalence of confusion from users about the exact content or behavior that triggered a sanction from the platforms; (2) the systemic failure on the part of platforms to provide good reasons to explain the decisions they reach; (3) the failure to inform users of how (and especially by whom) triggered the flagging of specific content for review by the platform´s moderation system; and (4) the confusion about who exactly makes content moderation decisions, and their possible biases. Accordingly, they call for the following disclosure in notices:

- more general demographic information about the makeup of their moderation teams, with particular regard to age, nationality, race, and gender.

- detailed information about the training and guidelines associated with the moderation process, including what processes exist to support moderators to make consistent and well- informed decisions in the context of potential ambiguity.
- Differential social impact of the inputs and outputs and the algorithms of these systems, to understand bias in moderation decisions. Analysis of this type will require large-scale access to data on individual moderation decisions as well as deep qualitative analyses of the automated and human processes that platforms deploy internally.

References

American Bar Association, ´Online Contracts: We May Modify These Terms at Any Time, Right? [20 May 2016], Available at: https://www.americanbar.org/groups/business_law/publications/blt/2016/05/07_moringiello/

Manila Principle of Intermediary Liability, Available at: https://www.manilaprinciples.org/principles

EFF, 'EFF Seeks Public Comment About Expanding and Improving Santa Clara Principles' [14 April 2020]. Available at: https://www.eff.org/press/releases/eff-seeks-public-comment-about-expanding-and-improving-santa-clara-principles

Nicolas Suzor, Sarah Meyers West, Andrew Quodling, and Jilian York, 'What Do We Mean When We Talk About Transparency? Toward Meaningful Transparency in Commercial Content Moderation', [2019] International Journal of Communication 13, 1526, 1543

59. Notice-and-notice

Notice-and-notice it is a process to deal with potential infringing content, a regime that became more widely known after being established in Canada's Copyright Act. It is an alternative to the notice-and-takedown model that works as follows: after a decision by the platform to remove content through private notification, the user is given the option to counter-notify and personally take responsibility for maintaining the content online, in which case the platform is exempt (de Souza & Schirru, 2016). According to Valente (2018) it is a model that distributes responsibilities in order to try to contemplate both the users and the (copy)rights holder, who wants a faster mechanism for notification and the possibility of removing infringing (copy)right content.

In a notice-and-notice system, providers forward to users the notifications made by right holders regarding alleged violations of rights practiced by these users, without summary removal of the content (as in the case of notice-and-takedown). Proponents of the notice-and-notice system, like the Canada model, argue that this process would eliminate flaws in the notice-and-takedown system (Geist, 2011).

ARTICLE 19 (2013) argues that this system would have good results when dealing with civil complaints regarding copyright, defamation, privacy, adult content and bullying (instead of harassment or threats of violence). In their view, this system, in the worst case scenario, would give content providers the opportunity to respond to allegations of violations of the law before any action is taken; it would contribute to reducing the number of abusive requests, as it requires a minimum of information about the allegations; and it would provide an intermediary system for resolving disputes before matters reach the courts.

References

Article 19 (2013). ‘Internet intermediaries: Dilemma of Liability’ Available at: <https://www.article19.org/wp-content/uploads/2018/02/Intermediaries_ENGLISH.pdf>

de Souza, Allan, & Schirru, Luca (2016). ‘Os direitos autorais no marco civil da internet’ *Liinc em Revista*, *12*(1). Available at: <http://revista.ibict.br/liinc/article/view/3712/3132>

Geist, Michael (2011). ‘Rogers Provides New Evidence on Effectiveness of Notice-and-Notice System’ Available at: <http://www.michaelgeist.ca/content/view/5703/125/>

Government of Canada (2017). ‘Notice and Notice Regime’ Available at: <https://www.canada.ca/en/news/archive/2014/06/notice-notice-regime.html>

Government of Canada (2018). ‘Notice and Notice Regime’ Available at: https://www.ic.gc.ca/eic/site/Oca-bc.nsf/eng/ca02920.html

Valente, Mariana (2019). ‘Direito autoral e plataformas de Internet: um assunto em aberto’ Available at:<https://www.internetlab.org.br/pt/especial/direito-autoral-e-plataformas-de-internet-um-assunto-em-aberto/>

60. Notice-and-takedown

Notice-and-takedown is a process to deal with potential infringing content based on the practice of sending an extrajudicial notification to the content provider and the immediate removal of the allegedly infringing content by the provider, without the need for a prior court order or possibility of counter-notification, before or after the content removal. Under such a mechanism, the provider may be liable if it did not remove the content immediately.

This mechanism has become predominant on the internet thanks to the Digital Millennium Copyright Act 1998 (DMCA), a U.S. law that limits the liability of online service providers for copyright infringement caused by their users if they promptly remove the offending content after being notified of an alleged infringement by copyright owners or their representatives. The enactment of this legislation immediately provoked similar adoptions in other countries, but nations that remained neutral were also regulated by the DMCA, as the major platforms apply this legislation globally, disregarding local copyright laws. Online platforms frequently use DMCA to establish a “three strikes” policy, a graduated response system that consists of three warnings to the user about posting copyrighted material, generating a serious penalty after the last warning, such as account deletion.

Explaining the relationship between this removal regime and the DMCA is relevant because notice-and-takedown mechanisms has serious problems, for instance:

a) offer serious risks to freedom of expression online, encouraging the arbitrary removal of content
b) allow short-term censorship of material whose timing is crucial, like election period.

National legislation lays out several hypotheses in which works protected by copyright can be used without the need of authorization by the copyright owner. However, it is not uncommon for copyright infringement to be the argument used for purposes of censorship, when the use of that work would be potentially lawful - which is not always easy to determine.

In big digital platforms, the responsible for assessing whether or not the posted contents complies with copyright rules is an artificial intelligence tool (*eg.* Youtube’s ContentID). This mechanism was created to analyze user generated content in search of excerpts of copyrightable works. Record companies and major film studios send copies of their original works and the system compares numerous excerpts with what is being shared on the network to find illegal copies on the platforms. The discussion about the limits of these automated tools came to light after several accounts had their contents blocked or deleted on the platforms. The problem lies in the so-called false positives, that is, when the filter allows for the claim of copyright even under lawful conditions, such as criticism and parodies.

Besides that, copyright holders can also commit abuses in the private notification procedure, as documented by the EFF's Takedown Hall of Shame project - and, without judicial review, the platform has incentives to remove content when it receives the notification, so as not to take the risk of being held responsible if it decides not to remove content that may be considered unlawful in a lawsuit. Users can file a lawsuit, too, if they understand that the removal has harmed their rights, but they have little incentive to do so, and are usually the most economically fragile part.

References

Article 19 (2013). ‘Internet intermediaries: Dilemma of Liability’ Available at: <https://www.article19.org/wp-content/uploads/2018/02/Intermediaries_ENGLISH.pdf>

Abranet (2011). ‘Contribuição para Aperfeiçoamento do Anteprojeto da Lei de Direitos Autorais’ Available at: <http://www2.cultura.gov.br/site/wp-content/uploads/2011/08/Abranet-Associa%C3%A7%C3%A3o-Brasileira-de-Internet.pdf>

de Souza, Allan, & Schirru, Luca (2016). ‘Os direitos autorais no marco civil da internet’ *Liinc em Revista*, *12*(1). Available at: <http://revista.ibict.br/liinc/article/view/3712/3132>

EFF. ‘Takedown Hall of Shame’ Available at: <https://www.eff.org/pt-br/takedowns>

Madigan, Kevin (2016). ‘Despite what you hear, Notice and Takedown is Failing Creators and Copyright Owners’ Available at: <https://cpip.gmu.edu/2016/08/24/despite-what-you-hear-notice-and-takedown-is-failing-creators-and-copyright-owners/>

Graduated Response. ‘About Graduated Responde’ Available at: <http://graduatedresponse.org/new/?page_id=5>

Valente, Mariana (2019). ‘Direito autoral e plataformas de Internet: um assunto em aberto’ Available at:<https://www.internetlab.org.br/pt/especial/direito-autoral-e-plataformas-de-internet-um-assunto-em-aberto/>

61. Notice-and-staydown

Notice and staydown (NSD) refers to a system of intermediary liability where, following a qualified notice, the intermediary is required not only to remove or disable access to an allegedly infringing content, but also to prevent further infringements by restricting the upload on the platform of the same or equivalent content. There is some ambiguity as to whether this model would require the prevention of uploads only of identical content or it would extend also to content with minor alterations (for instance a shorter version of a previously infringing video). The latter interpretation is favored at least in Europe, after the recent ruling of the European Court of Justice in Case C-18/18, *Eva Glawischnig-Piesczek v Facebook*, where the Court ruled that the prohibition of general monitoring obligation included in art. 15 of the E-commerce Directive "must be interpreted as meaning that it does not preclude a court of a Member State from:

- ordering a host provider to remove information which it stores, the content of which is identical to the content of information which was previously declared to be unlawful, or to block access to that information, irrespective of who requested the storage of that information;
- ordering a host provider to remove information which it stores, the content of which is equivalent to the content of information which was previously declared to be unlawful, or to block access to that information, provided that the monitoring of and search for the information concerned by such an injunction are limited to information conveying a message the content of which remains essentially unchanged compared with the content which gave rise to the finding of illegality and containing the elements specified in the injunction, and provided that the differences in the wording of that equivalent content, compared with the wording characterising the information which was previously declared to be illegal, are not such as to require the host provider to carry out an independent assessment of that content" (emphasis added).

Even prior to this ruling, however, NSD was already present at least to some degree in Germany, where courts had established under the doctrine of "Kern" a duty of care for hosts to review all the following infringing acts of a similar nature that are easily recognizable. This has been used to impose, for instance, the following (Husovec, 2019): (1) employ word-filtering technology for the name of the notified work, including on existing uploads,(2) use better than basic fingerprinting technology that only detects identical files, such as MD5, as a supplementary tool, (3) manually check external websites for the infringing links associated with the notified name of a work on services like Google, Facebook and Twitter or (4) use web-crawlers to detect other links on own service.

The term NSD originates from a heated discussion around the scope of the safe harbor for hosting intermediaries, which depends upon knowledge of infringing activity. As discussed in the entries on **red flag knowledge** or **willful blindness**, knowledge is occasionally found by courts even outside the qualified notice process, in the presence of facts that are sufficient to impute a culpable intention on the part of the intermediaries. However, although the implications of these doctrines might be somewhat similar to NSD, the obligations are fundamentally different in nature: the one imposed by the NSD arises automatically with the reception a valid notification, rather than following an inquiry into what is reasonable to have known considering the circumstances. This means also that NSD requires platforms to filter all uploads, in order to detect content previously identified as infringing (Kuzcerawy, 2020), which presumably will be done in automated form, due to the sheer volume of uploads. Uploads on Youtube, for instance, amount to more than 500 hours of video per minute (as of May 2019), which is strong evidence of the need for Youtube to rely on automated content recognition technologies like Content ID (deployed since 2007). Inthe view of the ECJ, automated search tools and technologies allow providers to obtain the result without undertaking an independent assessment, in particular to the extent that the notice contains the name of the person concerned by the infringement determined previously, the circumstances in which that infringement was determined and equivalent content to that which was declared to be illegal. However, one could actually question this conclusion: if a sentence or word can be considered defamatory in one context, it is not necessarily so in a different context, warranting therefore human determination at some level. This carveout from the safe harbor is likely to be even more problematic if extended to infringements of copyright or trademark law, given the challenges involved in ensuring that a machine recognizes the existence of licenses or valid defences by the alleged infringer.

In addition to the free speech concerns with the prior restraints imposed through NSD, there is substantial criticism on the economic effects of a NSD regime, in particular as it significantly raises costs for hosting platforms. While it may be convenient or even necessary for Youtube, Facebook or other large platforms to use content recognition technologies, it can be problematic to impose these requirements on smaller players, who might need in turn to obtain a license from the bigger player to fulfil their obligation. This was one of the major concerns of the proposal for a Directive on Copyright in the Digital Single Market, which required “information society service providers that store and provide to the public access to large amounts of works or other subject-matter uploaded by their users” to take measures, such as the use of *effective* content recognition technologies, to ensure the functioning of agreements concluded with rightholders for the use of their works or other subject-matter or to prevent the availability on their services of works or other subject-matter identified by rightholders through the cooperation with the service providers. The final version of the Directive changed this by requiring (a) the use of best efforts to obtain licensing agreements; (b) best efforts in accordance with high industry standards of professional diligence, to prevent the availability of the works in the sense explained above; and (c) the expeditious removal or disabling of content identified by notices *and best efforts to prevent their future uploads* in accordance with point (b). Furthermore, it removed the specific reference to content recognition technologies, while at the same time specifying that such obligations apply only to the extent that righthoders have provided the service providers with the relevant and necessary information (in the context of those technologies, this includes *in primis*

the reference files to enable the content recognition). Even more importantly, the new version of the Directive provides guiding principles on the NSD regime, by: (1) explicitly requiring Member States implementing such regime to preserve copyright exceptions and not impose general monitoring; (2) creating a three-tiered regime, where full NSD is only required for big and established players, while those who have been providing services in the EU for less than three years and which have an annual turnover below EUR 10 million would only need to comply with letter (a) above and to act upon notice for the removal of specific content, and yet those who have an average number of monthly unique visitors of such service providers exceeds 5 million would also have to demonstrate best efforts to prevent further uploads in the sense explained under letter (c); and (3) specifying that to determine the scope of the obligations imposed under this regime it must be taken into account (a) the type, the audience and the size of the service and the type of works or other subject matter uploaded by the users of the service; and (b) the availability of suitable and effective means and their cost for service providers.

References

Aleksandra Kuczerawy ,From ‘notice and take down’ to ‘notice and stay down’: risks and safeguards for freedom of expression, in Giancarlo Frosio (ed), The Oxford Handbook of Intermediary Liability Online (OUP, 2020)

Martin Husovec, The Promises of Algorithmic Copyright Enforcement: Takedown or Staydown? Which is Superior? And Why? 42 COLUM. J.L. & ARTS 53 (2018)

Christina Angelopoulos, Harmonizing Intermediary Copyright Liability in the EU: A Summary, in Giancarlo Frosio (ed), The Oxford Handbook of Intermediary Liability Online (OUP, 2020)

Giancarlo F. Frosio, Reforming Intermediary Liability in the Platform Economy: A European Digital Single Market Strategy, 112 Nw. U. L. Rev 19 (2017)

62. Nudging

Nudging refers to the use of choice architecture (the nudge) to influence the behavior of an individual or group of individuals (nudgees) without depriving them from the ability to choose a different course of action. The term was coined by Richard Thaler and Cass Sunstein with their book '*Nudge: Improving Decisions About Wealth, Heath and Happiness*', published in 2008, which offered a first conceptualization of a theory of regulation based on positive reinforcement and indirect suggestions as ways to influence the behavior and decision making of groups or individuals. The book was very impactful, leading to the rise of nudging regulation and even to the creation in 2010 of a "nudge" unit (also called behavioural insights team) within the UK government in order to generate and apply behavioural insights to inform policy, improve public services, and deliver positive results for people and communities.

Thaler and Sunstein propose to formulate public policies in a way that addresses the cognitive biases and helps improve decisions through what they call "choice architecture", i.e. the set of constraints surrounding individuals' choices. For instance, in one of their early papers, they map the letters of "NUDGES" onto five different types of design interventions:

1) iNcentives: leveraging the choosers' incentives can be a powerful mechanism to direct people. Think, for instance, about providing more salience to information that is relevant to make a decision that is otherwise underestimated, such as emphasizing the opportunity costs of buying a car.

- Understand mappings: this evokes a similar concept to the above, but referring to specific situations where the information is complex and therefore it is helpful to provide to choosers a map of possible options (for instance, in choosing the best possible cure for a disease).
- Defaults: this is probably the most commonly cited type of nudging, and refers to pre-selecting an option for the chooser while maintaining the possibility to reverse that choice.
- Give feedback: sometimes simply informing whether something is going wrong (or well) helps people redirect.
- Expect error: designing the choice architecture in a way that minimizes errors is also a nudge. The reason why this category is not subsumed within the notion of defaults is not apparent.
- Structure complex choices: sometimes choices are difficult to make, if the choice set is too large. For this reason, giving choosers the ability to structure their choice process (for instance through a filtering system that helps identifying useful ranges) can be a powerful nudge.

In a separate paper, Sunstein lists the different tools that can be used to obtain nudging effects, which appears to be an expansion (and to some extent a correction) of the previous list:

- Default Rules

- Simplification
- Use of social norms (e.g., illustrating examples of expected behaviour)
- Increase in ease and convenience (e.g. making low-cost option for healthy food visible)
- Disclosure
- Warnings (graphic or otherwise)
- Reminders
- Precommitment strategies (by which people commit to a certain course of action)
- Eliciting implementation intentions (e.g. "do you plan to vote?")
- Informing people of the nature and consequences of their own past choices ("smart disclosure").Although the above is a repetition of a largely rehearsed concepts, it is important to revisit these for two reasons. First, a constant theme running through these lists is the formulation of design choices to "de-bias" human decision-makers. This is somewhat different from the work of the fast & frugal school of behavioral economics, which endeavoured to help decision-makers by offering the best heuristics; and there is no reason in principle why heuristics could not be used in nudging to reach desired outcomes- for example, by framing options in a more visible and appealing fashion. However, the key point of criticism is that nudging tools may not always be used in a way that de-biases individuals: in fact, it can be used in a way that nurtures known biases and on that basis elicits choices that are not necessarily in the best interest of individuals. Second, the entire discussion by Thaler and Sunstein refers either explicitly or implicitly to nudging as a choice of public regulation, where the nudger can be trusted (or is at least assumed to be trusted) to pursue the general interest. But insofar as nudging is done by private entities that are not subject to the transparency and accountability safeguards that apply in the governmental context, a discussion about its boundaries and about ways in which compliance can be scrutinized becomes paramount.

It can also be noted that the definition provided by Thaler and Sunstein in their writings on the topic is not always consistent: in their book, they define nudging as "any aspect of the choice architecture that alters people's behaviour in a predictable way without forbidding any options or significantly changing their economic incentive". In doing so, they rule out the admissibility in this category of traditional regulatory tools such as bans, fines, taxes or other economic incentives (or disincentives). However, the line between a nudge and some of those categories is blurred, as the alternative course of action in all those situations remains available to the nudgee (in some instances, at the cost of violating the law) and the authors explicitly admit the possibility that nudges moderately alter one's economic incentives. They also argue that nudges are omnipresent in society (as every design choice has potential effects on individuals' behavior), and therefore the anti-nudging position is a "literal non-starter" - because at least deliberate nudges allow us to appreciate their rationale and operation. However, it is not clear that nudges can always be transparent and intelligible, and serendipity may be a value worth protecting- to let people determine their own path through random, or at least non-deliberate, nudges. Finally, Thaler and Sunstein only mention examples (such as the GPS, the retirement plan, the narrowing road design) where choice architects design, construct, or organize context without changing the original choice sets or fiddling with incentives. Yet it is clear that nudges will often have substantial impact on the range of choices or the incentives of the choosers, and not apparent how the nudger can abstain from the latter scenario. It is also not clear why Thaler and Sunstein separate economic incentives from other forms of incentives, including the prospect of pain and penalties, as that would more accurately incorporate the breadth of the endeavor of behavioural economics.

There is extensive philosophical discussion about the differences between nudging and manipulation. This discussion has been especially pronounced after the realization that the online world introduces a new type of nudge, one based on algorithmic real-time personalization and reconfiguration of choice architectures based on large aggregates of personal data: the so-called "hyper-nudging" (Yeung, 2016). In this context, where the transparency of nudges is hindered by the personalization of the nudges, the accountability of manipulative nudges increases.

There are a range of definitions that can be used to identify manipulation, for instance:

- an intentional act that successfully influences a person to belief or behavior by causing *changes in the mental processes* other than those involved in understanding (Faden & Beauchamp, 1986)
- a kind of influence that *bypasses or subverts the target's rational capabilities*, in a way that treats its objects as "tools and fools" (Wilkinson, 2014)
- directly *influencing someone's beliefs, desires or emotions,* such that she falls short of ideals for belief, desire or emotion in ways typically not in her self-interest or likely not in her self-interest in the present context (Barnhill, 2014)
- A statement or action that does not sufficiently engage or appeal to people's capacity for reflective and deliberative choice (Sunstein, 2015)
- Non-rational influence (Noggle, 1996)
- Pressure, but *not irresistible pressure* amounting to coercion (Raz 1985; Noggle 1996)
- *Trickery* to induce behavior (Noggle, 1996)
- *Hidden influence*: intentionally and covertly influencing decision-making, by targeting and exploiting one's decision-making vulnerabilities (Susser et al 2019)

We can imagine clear cases of manipulation (subliminal advertising), cases that clearly fall outside of the category (for example, a warning about deer crossings in a remote area), and cases that can be taken as borderline (a vivid presentation about the advantages of a particular mortgage, or a redesign of a website to attract customers to the most expensive products).

In order to distinguish nudging from manipulation, various authors propose criteria to set limits on the acceptability of nudges. For instance, Sunstein requires them to be de-biasing (market failure correcting), educative and non-exploitative (Sunstein 2015).

Thaler, in turn (Thaler, 2015), uses the following criteria to distinguish acceptable nudging (or "nudging for good"):

- First, the nudge is transparent and not misleading.
- Second, it is as easy as possible to opt out.

- Third, they must increase welfare.

Baldwin focuses on the proportionality of the nudge to scale of the problem, considering evidence of effectiveness and the moral considerations at stake (Baldwin 2014). He then distinguishes between simple nudges, that only engages system 1 thinking (1st degree nudge), more intrusive nudges that exploit behavioral or volitional limitations so as to bias a decision in the desired direction (2nd degree nudge), and a yet more intrusive nudge (3rd degree) where there is no ability to appreciate the influence.

He concludes that nudges will have different effectiveness depending on who the targets are, where the following characteristics are relevant: whether the individual´s objective aligns with that of the nudger, and whether their capacity to absorb the nudging information is high or low.

References

Ruth Faden & Beauchamp, A History of Informed Consent (OUP, 1986)

Karen Young, ‘Hypernudge’: Big Data as a Mode of Regulation by Design’

Information, Communication & Society (2016) 1,19

Anne Barnhill, *What is Manipulation?*, *in* MANIPULATION: THEORY AND PRACTICE 51, 65–72 (Christian Coons & Michael Weber eds., 2014)

Baldwin, Robert (2014) From regulation to behaviour change: giving nudge the third degree. The Modern Law Review, 77 (6). pp. 831-857

Pelle Guldborg Hansen and Andreas Maaløe Jespersen, ‘Nudge and the Manipulation of Choice A Framework for the Responsible Use of the Nudge Approach to Behaviour Change in Public Policy’ European Journal of Risk Regulation 1 (2013), p. 10

Cass Sunstein, The Ethics of Nudging, 32 Yale J. on Reg. (2015). Available at: http://digitalcommons.law.yale.edu/yjreg/vol32/iss2/6

T. M. Wilkinson, Nudging and Manipulation, 61 POL. STUDIES 341, 342 (2013).

Richard Thaler, ‘The Power of Nudges, for Good and Bad’ *The New York Times* (New York 31 October 2015), at https://www.nytimes.com/2015/11/01/upshot/the-power-of-nudges-for-good-and-bad.html

Richard Thaler and Cass Sunstein, *Nudge: Improving Decisions About Wealth, Heath and Happiness* (Yale University Press 2008)

Joseph Raz, *The Morality of Freedom* (Oxford: Oxford University Press, 1986).

Robert Noggle, “Manipulative Actions: A Conceptual and Moral Analysis,” *American Philosophical Quarterly* 33(1), 199

Cass Sunstein, The Ethics of Influence: Government in the Age of Behavioral Science (Cambridge: Cambridge University Press, 2016)

63. Online Advertising

Online advertising can be defined as the industry of Internet marketing/advertising, as well as the technologies (adtech) and practices characterizing this industry. Online advertising has two important features distinguishing it from traditional advertising: measurability and targetability (Goldfarb & Tucker, 2011). Among the main types of online advertising we can distinguish display advertising, search advertising and social media advertising (Goldfarb & Tucker, 2011).

Display ads are mainly used on regular websites and entail the display of banners or audio-visual ads. Search advertising entails that ads are featured at the top of search results returned from a search engine query. Both types of advertising evolved to also include so-called 'ad auctions' and 'real-time bidding', which involve the buying and selling of advertising via programmatic instantaneous auctions (Information Commissioner's Office, 2019; Google, 2020). Social media advertising uses elements of display and search advertising, combined with native advertising models such as influencer marketing (see also the entries for 'Content/Web monetization' and 'Influencers/content creators)'.

References

Avi Goldfarb & Catherine Tucker, 'Online Advertising', [2011] 81 Advances in Computers.

Information Commissioner's Office, 'Update report into adtech and real time bidding', [20 June 2019]. Available at: <https://ico.org.uk/media/about-the-ico/documents/2615156/adtech-real-time-bidding-report-201906.pdf>.

Google, 'How the Google Ads auction works'. Available at: <https://support.google.com/google-ads/answer/6366577?hl=en>.

64. Open Identity

An *online identity* is a collection of personal information about a person, associated with credentials that allow the owner of the identity to control the information and to assert their identity towards other parties over the Internet.

The identity may represent an actual person (*real world identity*), a fictitious person (*pseudonymous identity*) or an unknown set of one or more persons (*anonymous identity*).

Frameworks for the management of online identities usually perform some or all of these functions:

1. *Authentication*, i.e. the establishment and verification of credentials (passwords, biometric data etc.) to ensure that only the legitimate owner of the identity can use it;
2. *Authorization*, i.e. the request and release of permission for an authenticated identity to access a specific resource or service;
3. *Signing*, i.e. the creation of cryptographic attestations of a certain assertion by the owner of the identity;
4. *Information management*, i.e. the entering, storing and controlled distribution of the personal information that the owner associates with the identity.

An *open identity* is an online identity provided and managed through the use of open, federated standards that allow multiple identity providers to coexist, including the possibility for the identity owner to switch from a provider to another or to self-manage their identity without recurring to an external identity provider (this latter case is called *self-sovereign identity*).

Currently, the most common identity frameworks are those provided by Internet platforms, especially by Google, Facebook and Apple. These systems are widely used for registration and login into online websites and services; while they are based on an open protocol (OpenID Connect), they are not open, as the user cannot choose a different provider; e.g. a Google identity can only be used within the Google ecosystem, and no other providers can supply identities for that ecosystem.

The European Union, through the eIDAS Regulation (EU Regulation, 2014), has established an identity framework that federates national identity systems and can be used for logging into online services, typically for real world identities and public administration websites. The openness of eIDAS implementations varies across European countries.

References

(EC) Document 32014R0910 Regulation (EU) No 910/2014 of the European Parliament and of the Council of 23 July 2014 on electronic identification and trust services for electronic transactions in the internal market and repealing Directive 1999/93/EC [2014]. Available at:

https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv%3AOJ.L_.2014.257.01.0073.01.ENG

65. Open Standard

An *open standard* is a standard that is developed through open processes and can be used and implemented by every interested party under non-discriminating conditions, and if possible for free.

Different standardization organizations adopt different definitions for the term, which generally agree about the fact that the standard must have been developed through an open consensus process that does not exclude or disadvantage any stakeholder, but disagree on the intellectual property licensing requirements. Under that aspect, the definitions and the resulting policies can be broadly grouped into two categories:

1. Definitions that require the standard and the related essential intellectual property to be available for free, without requiring negotiations with intellectual property holders or the payment of royalties; this is for example the policy of the World Wide Web Consortium (Dardailler, 2007 and Weitzner, 2004);

2. Definition that require the standard and the related essential intellectual property to be available under *"fair, reasonable and non-discriminatory"* (FRAND) licensing terms, which may however include the payment of royalties and/or a discretionary negotiation with the rights holder; this is for example the policy of the ITU-T (ITU, 2005).

FRAND technologies can be a significant obstacle to projects that do not have any amount of funding or do not have the legal capabilities to deal with licensing negotiations, such as many open source projects.

The Internet Engineering Task Force *"prefers"* technologies which are not subject to patents or whose patents are royalty-free, but accepts FRAND technologies if necessary (Bradner and Contreras, 2017). A similar stance is taken by the European Union, whose definition of open standard can be found in Annex II to Regulation 2015/2012 (EU Regulation, 2012); the European Commission has repeatedly addressed the problems connected to a fair interpretation of the FRAND concept (European Comission, 2017).

References

Daniel Dardailler , ' Definition of Open Standards' (World Wide Web 2007) < https://www.w3.org/2005/09/dd-osd.html>

Daniel J. Weitzner, ' W3C Patent Policy' (World Wide Web 2004) < https://www.w3.org/Consortium/Patent-Policy-20170801/>

IPR Ad Hoc Group, ' Definition of \"Open Standards\"' (ITU 2005) < https://www.itu.int/en/ITU-T/ipr/Pages/open.aspx>

S. Bradner, J. Contreras, ' Intellectual Property Rights in IETF Technology' (IETF 2017) < https://tools.ietf.org/html/rfc8179>

(EC) REGULATION (EU) No 1025/2012 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL [25 October 2012] Available at: https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2012:316:0012:0033:EN:PDF

European Comission, *Communication From The Commission To The European Parliament, The Council And The European Economic And Social Committee* (1st, , Brussels 2017). Available at: https://ec.europa.eu/docsroom/documents/26583/attachments/1/translations/en/renditions/native

66. Optimization

Optimization refers to the practice or process of making something as effective, visible, functional, or efficient as possible. More specifically, search engine optimization refers to the practice of designing your website or page to increase the quantity, quality, or both, of organic, unpaid search results. At the same time, optimization is an predominant organizing principle of digital systems that incorporate real-time feedback from users (Kulynick et al. 2018): for instance, ride sharing applications such as Uber, which rely on optimization to decide on the pricing of rides; navigation applications such as Waze, which rely on optimization to propose best routes; banks, which rely on optimization to decide whether to grant a loan; and advertising networks, which rely on optimization to decide what is the best advertisement to show to a user. As a solution to the potentially adverse effects from manipulation of individual behavior caused by optimization systems, some have proposed the adoption of specific measures to reduce or eliminate the emergence of such effects from the design stage (see a.g. Amodei et al, 2016). As an alternative, the concept of Protective Optimization Technologies has been proposed- i.e., technological solutions that those outside of the optimization system deploy to protect users and environments from the negative effects of optimization (Kulynick et al. 2018).

References

Dario Amodei et al (2016). Available at: https://arxiv.org/abs/1606.06565

Kulynich et al (2018). Available at: https://arxiv.org/abs/1806.02711

67. Platform

According to the Merriam Webster dictionary, the concept of platform generally refers to a plan or design. The Historical Larousse dictionary suggests that the term first appeared in the French language in 1434, to define a “horizontal surface acting as a support.” This original meaning helps to construct a general characterisation of the platform as a structure on top of which something – be it a product or a service – may be built and operated.

When the substantive is qualified as “digital” or “online”, it may refer to a vast array of software applications that are frequently loosely defined. As pointed out by the European Commission (2016b), the term is frequently utilised to refer to "two-sided" or "multi-sided" markets (Rochet & Tirole, 2003; Evans, 2003). In such markets, users are brought together by a platform operator in order to facilitate an interaction (exchange of information, a commercial transaction, etc.). In the context of digital markets, depending on a platform's business model, users can be buyers of products or services, sellers, advertisers, software developers, etc.

Conspicuously, the existence of multi-sided platforms is not a phenomenon which can be defined as exclusively taking place in the online world. On the contrary, the history of businesses demonstrates that platforms emerge in a wide range of sectors, in the off-line world, as well as in the online world. In this sense, the European Commission (2016b) stresses that examples vary from markets to newspapers: both gather sellers and buyers in a common space thereby facilitating contact between two sides that would otherwise be unlikely to interact. Nevertheless, 'real life' platforms were usually limited physically and geographically (the merchandise had to be transported and stocked, a paper had limited circulation and advertisements had to be location specific etc.)

Amit Tiwana (2014) provides a useful definition of platforms as “the extensible codebase of a software-based system that provides core functionality shared by apps that interoperate with it, and the interfaces through which they interoperate.” Due to the heterogeneous nature of digital platforms, many studies on digital platforms provide examples to clarify what they refer to when discussing digital platforms. This is the case, for instance in the European Commission (2016a) Communication on Online Platforms and the Digital Single Market, where the characteristics of digital platforms are listed and described, providing examples of platforms, rather than defining them.

Notably, the aforementioned document highlights the following characteristics of online platforms:

- they have the ability to create and shape new markets, to challenge traditional ones, and to organise new forms of participation or conducting business based on collecting, processing, and editing large amounts of data;
- they operate in multisided markets but with varying degrees of control over direct interactions between groups of users;
- they benefit from ‘network effects’, where, broadly speaking, the value of the service increases with the number of users;

- they often rely on information and communications technologies to reach their users, instantly and effortlessly;
- they play a key role in digital value creation, notably by capturing significant value (including through data accumulation), facilitating new business ventures, and creating new strategic dependencies.

To provide a very broad and comprehensive definition, the Recommendations on Terms of Service and Human Rights developed by the IGF Coalition on Platform Responsibility define a platform as "as any applications allowing users to seek, impart and receive information or ideas according to the rules defined into a contractual agreement."

Overall, the majority of digital platforms share three main features: they are technologically mediated, they enable interactions between different types of users and allow those types of users to implement specific activities (de Reuver, Sørensen & Basole, 2018). Existing literature points out the existence of three broad categories of online platforms:

- Market-makers bring together two distinct groups that are interested in trading, increase the likelihood of a match, and reduce search costs.
- Audience makers match advertisers to audiences.
- Demand coordinators, such as software platforms, operating systems, and payment systems coordinate demand between different user groups (for example card holders and merchants, developers and smartphone users).

Hence, platforms provide a medium where one type of platform users can deliver value both to the other type of users and the platform itself. In this context the European Commission (2016b) points out that the demand of the different types of users is related to the supply of other types, and several kinds of interdependencies may exist between the various types of platform users:

- producers of complementary products (e.g. app developers) and end consumers (gamers),
- advertisers and readers
- shoppers and sellers
- job seekers and recruiters
- accommodation providers and accommodation seekers
- transportation providers and passengers.

Importantly, major online platforms trigger important network effects and generate revenue by recruiting one type of users (e.g. advertisers) and offering them access to another type of users (e.g. individual users of social networks).

Critically, platforms define a private ordering – through their terms of service, their technical architectures and their practices – that directly impact their users as well as the products and services built on top of them. (Belli & Veturini 2016; Belli and Sappa 2017; Belli & Zingales 2017)

References

Belli L, & Venturini J. Private ordering and the rise of terms of service as cyber-regulation. Internet Policy Review. Vol 5. N° 4. (2016). https://policyreview.info/articles/analysis/private-ordering-and-rise-terms-service-cyber-regulation

Belli L & Sappa C. The Intermediary Conundrum: Cyber-regulators, Cyber-police or both? JIPITEC. (2017). http://www.jipitec.eu/issues/jipitec-8-3-2017/4620

Belli L & Zingales N, Platform regulations: how platforms are regulated and how they regulate us. FGV: Rio de Janeiro. 2017

de Reuver, M., Sørensen, C., & Basole, R. C. (2018). The Digital Platform: A Research Agenda. Journal of Information Technology, 33(2), 124–135.

European Commission. Online Platforms and the Digital Single Market Opportunities and Challenges for Europe. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. COM/2016/0288 final. 2016a

European Commission. Online Platforms Accompanying the document Communication on Online Platforms and the Digital Single Market. Commission Staff Working Document {COM(2016) 288 final. 2016b.

Evans, D. (2003), "Some Empirical Aspects of Multi-sided Platform Industries", Review of Network Economics, 2(3), 2194-5993

Rochet, J.-C., & Tirole, J. (2003), "Platform Competition in Two-sided Markets", Journal of the European Economic Association, 1(4), 990-1029

Amrit Tiwana, Platform Ecosystems Aligning Architecture, Governance, and Strategy. 2014

van Eijk N. et al Digital platforms: an analytical framework for identifying and evaluating policy options. 2015.NO report | TNO 2015 R11271 | Final report. https://www.ivir.nl/publicaties/download/1703.pdf

68. Platform Governance

The concept of Governance refers to a 'decentred' perspective on regulation, which does not emanate solely from the state but is instead carried out by (complex, interactive) constellations of public and private stakeholders. The term has found widespread usage in the context of platforms, which are often seen to play an influential role as overseers of complex social and commercial ecosystems (e.g. Van Dijck, Poel en De Waal 2018). The result is an growing attention for "platform governance" amongst academics and policymakers (Gorwa 2019).

In the words of Tarleton Gillespie, platforms are implicated in online governance in two ways: governance *by* platforms, and governance *of* platforms (Gillespie 2016). Governance *by* platforms describes their role in facilitating and policing online behaviour, whereas governance *of* platforms describes the actions of governments and other stakeholders who contest and control platform action.

Governance *by* platforms can take many forms. Some of its most recognisable expressions are the drafting and enforcement of general rules and standards, such as Community Guidelines and Terms of Service, as well as the content moderation practices that purport to enforce these principles. But the concept of platform governance can be extended to countless other areas of platform policy, including their interactions with ad buyers such as political campaigners (Kreiss & MacGregor 2019); engagement with, and donations to, civil society and academia (Bruns 2019); treatment of content providers and influencers (Caplan & Napoli 2020; Goanta & Ranchordas 2020); and accommodations of government agencies and other public authorities (e.g. Benkler 2011). More fundamentally, the basic technical design of platform services can constitute a form of governance, to the extent that it structures and constrains the behaviour of users and other stakeholders.

Governance *of* platforms is an equally broad and varied concept. Government regulation is typically the first point of reference, from legislation and regulatory oversight to judicial action. But the aforementioned private stakeholders can also play a role in governing platforms. For instance, civil society actors can investigate and criticise platforms, either independently or as members of self- or co-regulatory regimes (Gorwa 2019). Platform users, content providers and advertisers may also be able to leverage governments or platforms to change their course, as can activists and mobilized user groups – a notable example being the recent advertiser boycott against Facebook. As Robert Gorwa highlights, these complex multi-stakeholder interactions play out across various geographical scales, with overlapping "local, national, and supranational mechanisms of governance" (Gorwa 2018).

References

Caplan, Robyn and Tarleton Gillespie. 2020. “Tiered Governance and Demonetization: The Shifting Terms of Labor and Compensation in the Platform Economy”. *Social Media + Society* 6(2).

Benkler, Yochai. 2011. “A Free, Irresponsible Press: Wikileaks and the Battle over the Soul of the Networked Fourth Estate”. Harvard Civil Rights-Civil Liberties Review 46-1. Available at: http://benkler.org/Benkler_Wikileaks_current.pdf

Bruns, Axel. 2019. After the ‘APIcalypse’: social media platforms and their fight against critical scholarly research. *Information, Communication & Society* 22(11). Available at: https://eprints.qut.edu.au/131676/

Goanta, Catalina and Sofia Ranchordás. 2020. *The Regulation of Social Media Influencers.* London: Edward Elgar.

Gillespie, Tarleton. 2016. “Governance of and by platforms”, in: Jean Burgess, Thomas Poell, and Alice Marwick (eds.), *SAGE Handbook of Social Media*. New York: SAGE Publishing. Available at: https://www.microsoft.com/en-us/research/wp-content/uploads/2016/12/Gillespie-Regulation-ofby-Platforms-PREPRINT.pdf

Gorwa, Robert. 2019. “What is platform governance?”. *Information, Communication and society* 22(6). Available at: https://gorwa.co.uk/files/platformgovernance.pdf

Gorwa, Robert. 2019. “The Platform Governance Triangle: Conceptualising the Informal Regulation of Online Content”. *Internet Policy Review* 8(2). Available at: https://policyreview.info/articles/analysis/platform-governance-triangle-conceptualising-informal-regulation-online-content

Kreiss, Daniel and Shannon MacGregor. 2019. “The “Arbiters of What Our Voters See”: Facebook and Google’s Struggle with Policy, Process, and Enforcement around Political Advertising”. *Political Communication* 36(4).

Van Dijck, Jose, Thomas Poell and Martijn de Waal. 2018. *The Platform Society: Public Values In A Connective World*. New York: Oxford University Press.

69. Platform Neutrality

Platform neutrality expresses the idea that products or services that function as platforms should not unreasonably discriminate against complements. Platform neutrality was popularized after a report issued by the French National Digital Council (Conseil National du Numérique) in 2014, which advanced numerous recommendations on the issue (CNNum 2014).

Platform neutrality draws on principles developed for utilities regulation. Utilities, because of the fundamental services they offer and because they have traditionally been (public or private) monopolies, were regulated to hold themselves out to serve the public indiscriminately, meaning that they cannot make individualized decisions on whether and on what terms to deal with each customer. Modern digital platforms are sometimes seen as performing similar fundamental roles, such as providing the necessary functionality for app ecosystems to emerge (app neutrality) or discoverability through online search (search neutrality) (Frischmann 2004). If platforms guarantee equal conditions to all complements, then the complements can compete on the merits.

Platform neutrality has been criticized on various grounds. The first is that digital platforms often do not exhibit the same characteristics as traditional utilities, namely, they are neither monopolies nor indispensable nor do they unequivocally offer the same kind of public good services as utilities and therefore they should not be subject to the same kind of rules. Secondly, discriminatory behavior can be welfare enhancing and it is therefore generally not banned in the market, unless it is unreasonable and distorts market conditions (Yoo 2004). Thirdly, some platform services by definition have to be discriminatory (e.g ranking of search results), which makes a neutrality principle impossible to implement and even counter-productive (Renda 2015).

Because of the tension between the benefits and risks of platform neutrality, regulation in this domain has so far been limited to business to business (B2B) relations and only to light touch obligations that emphasize transparency and accountability rather than banning specific types of conduct (Regulation (EU) 2019/1150 on promoting fairness and transparency for business users of online intermediation services).

References

CNNum. “Platform Neutrality: Building an Open and Sustainable Digital Environment.” Opinion No. 2014-2 (2014).

Frischmann, Brett M. “An economic Theory of Infrastructure and Commons Management.” Minnesota Law Review 89 (2004): 917.

Renda, Andrea. “Antitrust, Regulation and the Neutrality Trap: A Plea for a Smart, Evidence-based

Internet Policy.” CEPS Special Report No. 104 (2015).

Yoo, Christopher S. “Would Mandating Broadband Network Neutrality Help or Hurt Competition- A Comment on the End-to-End Debate.” Journal on Telecommunications & High Technology Law 3 (2004): 23.

70. Pornography

This entry provides a brief literature review of pornography through the lens of feminist legal theory. Chamallas (2012) segments the feminist movements in legal scholarship by (1) the generation of equality (1970s), which is often associated with liberal feminism (Chamallas, 2012, p. 19) because of the claims against formal inequality and toward individual rights such as the access to male-dominated activities; and (2) the generation of difference (1980s), which was responsible for bringing substantive inequalities to the discussion, such as the feminization of poverty and the gender gap in politics. Feminist legal scholars and activists from the 1980s are classified into two other subcategories: dominance feminism and cultural feminism. The first group was the main actor responsible for advocating for legislation to protect women's bodily integrity (Chamallas, 2012, p. 22) with campaigns against pornography.

For MacKinnon (1987), for example, pornography and the culture that portrays women as sex objects are responsible for the maintenance of sexual violence and sex discrimination – briefly, for MacKinnon, sexuality is expropriated from women by the male-dominated State, as for the Marxists, labour is expropriated from workers by the capitalist State. Her work – that defines pornography as "graphic sexually explicit subordination of women" (MacKinnon, 1991) – has inspired legislation and ordinances against it worldwide. Feminists that are influenced by this view tend to see pornography as a promotion of dehumanization and objectification of women that are in a situation of inequality – since most of the women that work in this industry are from marginalized segments (poor, black and latina women).

Other feminists, however, fear that the combat of pornography in these terms may engender a worse situation for women and for freedom of speech – also arguing that the male-dominated state could use these ordinances against minorities. In 1984, for example, a feminist "anti-censorship" task force (the F.A.C.T.) was formed by women who were against the anti-pornography movement. This would contemplate liberal concerns about individual rights and choices and also inspire the autonomy feminism movement that arose in the 1990s and focused on sex-positivity and women's own agency.

References

Chamallas, M. E. (2012). *Aspen Treatise for Introduction to Feminist Legal Theory*. Wolters Kluwer Law & Business.

MacKinnon, C. A. (1987). *Feminism unmodified: Discourses on life and law*. Harvard university press. Tn that work in this industry are from marginalized segments (poor, black and latina women).

71. Prioritization

Prioritization can refer to a form of traffic management and the way traffic flows through the internet and its connected parts, it can refer to the ranking of results in a hierarchy.

In the former, some network traffic is given precedence over others. Prioritization of some types of information over others can be based on importance. Users or edge providers can pay to optimize the transmission of traffic, the platform can prioritize some types of traffic over others, or users can pay access providers to transmit some traffic before or faster than other traffic. Some see prioritization as contradicting net neutrality principles.

Prioritization can also refer to the ordering of indexed results, with higher quality, or more important, or more relevant results being given priority over other results.

72. Proactive measures

This entry provides an overview of the concept of proactive measures, where "measures" is a term of art which includes a range of steps that can be taken as a form of governance or regulation, usually in relation to specific kinds of content or conducts. "Proactive" is a term thatis used frequently to qualify the nature of these measures taken by platforms or other intermediaries with regard to third party content. The two most common meanings are: (1) as an operational matter, acting based on the platform's own initiative, not in response to a notice or other external source of information; (2) as a legal matter, acting voluntarily and without legal compulsion.

Naturally, there is some overlap between (1) and (2), as the external source of information under **(2)** may be a judicial order or another form of notification that triggers a legal obligation for the platform to take the measures in question. In addition, legal obligations may arise independently from the existence of a specific notification, as platforms might be subject to a duty of care to prevent the dissemination of certain content in the first place: an example is the recently proposed Eliminating Abusive and Rampant Neglect of Interactive Technologies Act (EARN IT Act) of 2020, which would create an exemption to the immunity of platforms under section 230 of the Communication Decency Act by allowing civil and state criminal suits against companies who do not adhere to certain recommended "best practices" with regard to Child Sexual Abuse Material (CSAM). The EU legislation on this matter is the Audiovisual Media Service Directive (2018/1808) which among other things require**s** Member States in its art. 28b to ensure that video-sharing platform providers under their jurisdiction take appropriate measures to protect**:**

- (a) minors from programmes, user-generated videos and audiovisual commercial communications which may impair their physical, mental or moral development in accordance with Article 6a(1);
- (b) the general public from programmes, user-generated videos and audiovisual commercial communications containing incitement to violence or hatred directed against a group of persons or a member of a group based on any of the grounds referred to in Article 21 of the Charter of the Fundamental Rights of the European Union, or containing content the dissemination of which constitutes a criminal offence in the EU (namely child pornography or xenophobia**)**

(ba) the general public from programmes, user-generated videos and audiovisual commercial communications containing content the dissemination of which constitutes an activity which is a criminal offence under Union law, namely public provocation to commit a terrorist offence within the meaning of Article 5 of Dir. (EU) 2017/541, offences concerning child pornography within the meaning of Article 5(4) of Dir. 2011/93/EU and offences concerning racism and xenophobia

Similar language can be found in the proposal for a Terrorism Regulation in establishing duties of care and proactive measures on Hosting Services Providers (HSPs) to remove terrorist content**,** including to remove when appriopriate terrorist material from their services, including bydeploying automated detection tools, acting in a "diligent, proportionate and non-discriminatory manner, and with due regard for due process", and in the Christchurch call made by several governments and online service providers to address terrorist and other violent extreme content online, including a commitment by providers to adopt **"**specific measures seeking to prevent the upload of terrorist and violent extremist content and to prevent its dissemination on social media and similar content-sharing services, including its immediate and permanent removal, without prejudice to law enforcement and user appeals requirements**,** in a manner consistent with human rights and fundamental freedoms**".**

Platforms´ general concern for the adoption of proactive or "voluntary" measures is that they may lead to the establishment of knowledge that triggers an obligation to remove or disable access to content, failing which the platforms might lose the benefit of the safe harbor. For this reason, scholars have argued for the introduction of a general "good samaritan" provision, modeled upon Section 230 © of the US Communications Decency Act, which would preserve the application of the safe harbor as long as the measures are taken in "good faith" against certain types of objectionable content (Kuzcerawy, 2018; Barata, 2020)

References

Joan Barata, 'Positive Intent Protections: Incorporating a Good Samaritan principle in the EU Digital Services Act' (Center for Democracy and Technology 2020). Available at: https://cdt.org/insights/positive-intent-protections-incorporating-a-good-samaritan-principle-in-the-eu-digital-services-act/#:~:text=Close%20the%20menu-,Positive%20Intent%20Protections%3A%20Incorporating%20a%20Good%20Samaritan%20principle,the%20EU%20Digital%20Services%20Act&text=The%20%E2%80%9CGood%20Samaritan%E2%80%9D%20principle%20ensures,other%20forms%20of%20inappropriate%20content.

Aleksandra Kuczerawy , 'The EU Commission on voluntary monitoring: Good Samaritan 2.0 or Good Samaritan 0.5?' (Ku Leuven 2018) < https://www.law.kuleuven.be/citip/blog/the-eu-commission-on-voluntary-monitoring-good-samaritan-2-0-or-good-samaritan-0-5/>

73. Recommender systems

Recommender systems are algorithms aimed at supporting users in their online decision making. More specifically, in the computer science literature, a recommender system is defined as: "a specific type of advice-giving or decision support system that guides users in a personalised way to interesting or useful objects in a large space of possible options or that produces such objects as output" (Felfernig et al. 2018).

Examples of such systems are the Amazon recommender tool for products, the Netflix algorithm that suggests movies, the Facebook software that finds "friends" we might know.

A key element of recommender systems is that their suggestions are personalised, i.e. based on users' preferences. Such information can be directly obtained from users (e.g. asking specifically for her preferences) or can be generated by observation of their behaviour (Jannach et al. 2010). Most recommender systems rely on machine learning techniques, including deep neural networks (Goanta and Spanakis 2020).

From a technical point of view, four main models of recommendation systems have been identified (Aggarwal 2016): 1) collaborative **filtering** systems; 2) **content**-**based recommender systems**; 3) knowledge-based recommender systems; 4) hybrid systems.

Collaborative filtering systems perform the recommendation process based on the **user**-item interaction provided by several users. Let us assume A and B have similar tastes and that the algorithm has recorded such a similarity. A rates the movie Titanic highly, the recommender system infers that the rating of B for Titanic will be likely to be similar. Hence, the algorithm formulates Titanic as a recommendation for B.

Content-based recommender systems construct a predictive model thanks to the attributes (descriptive features) of users or items. Following in the movie example: A rated Titanic highly. Titanic is described by keywords like "drama" and "love affair". Therefore, movies that are classified in the same way (Romeo+Juliet or Pearl Harbour) would be recommended to A.

Knowledge-based recommender systems formulate recommendations based on the constraints specified by users, the item attributes and the domain knowledge. Such systems are common where items are not bought very often (so it is not efficient to rely on user-item interactions). Examples of them are tools for searching real estates, cars, touristic accommodation, etc.

Finally, hybrid systems combine one or more of the previous aspects.

Another classification proposed in the literature distinguishes recommender systems in three typologies, based on the role played by the platform in the sourcing of the content recommended (Cobbe and Singh 2019). In the so-called “open recommending” system, such as YouTube, the platform does not perform editorial control and the recommendation is elaborated from user-generated content. On the contrary, “curated recommending” is intended as a system where the platform selects, curates or approves the content. Finally, in “closed recommending” systems the platform creates itself the content to be recommended. Such a classification can be relevant when intermediary liability is at stake. While for “curated” and “closed” recommending systems the safe harbour immunity regime will be out of the picture, queries remain for “open” recommenders. Before the Court of Justice of the EU a case is currently pending to ascertain whether YouTube plays an active role by recommending videos and performing other ancillary activities (Case 500/19).

The organisation of the recommender system and its intelligibility can also give rise to direct liability of the platform vis-à-vis the content creator, such as a social media influencer. Goanta and Spanakis (2020) argue that the rules against unfair commercial practices and competition law both in Europe (the Unfair Commercial Practices Directive) and in the US (the Federal Trade Commission Act) can offer a first line of defence against the opaqueness of the algorithmic decision-making and the discretionary power exercised by platforms to the detriment of content creators. Such a framework however does not provide a full fledge of protection to the emerging actors involved in social media transactions and needs to be strengthened (Goanta and Spanakis 2020).

Recent legislative initiatives in Europe

Ranking

Knowledge-based recommender systems essentially work in response to a search query launched by a user. The output of such a model is likely to overlap with the legal definition of ranking. In Europe, the latter is intended as: "the relative prominence of the offers of traders or the relevance given to search results as presented, organised or communicated by providers of online search functionality, including resulting from the use of algorithmic sequencing, rating or review mechanisms, visual highlights, or other saliency tools, or combinations thereof" (recital 19, Directive (EU) 2019/2161. See also, Art. 3(1)(b), Directive (EU) 2019/2161 and art. 2(8), Regulation (EU) 2019/1150). To increase the **transparency** in online **marketplaces**, newly introduced provisions in the B2C and the P2B context impose an obligation to provide clear information about the main parameters and parameter weighting adopted to rank products and to disclose any paid advertising or payment specifically made for achieving a higher ranking within the search results. It is yet to be seen how these transparency requirements will be developed, considering not only the complexity and the dynamicity that ranking algorithms might reach through machine learning but also the possible limitations imposed by trade secrets (Twigg-Flesner 2018). The Commission is currently working on transparency guidelines (European Commission, 2020) to facilitate the compliance of platforms.

Ratings and reviews

Both in collaborative filtering and content-based recommender systems, the first input is given by users ratings and reviews. They can be defined respectively as scores (in a numerical form) and feedback (in a textual form) generated by the platform's users to report their experience with a product, a buyer or a service provider in a supposedly impartial manner. Some platforms provide aggregate or consolidated ratings, which sum up the single ratings or reviews in an overall assessment. Consolidated ratings can play an essential role in supporting the users' decision-making process, addressing some cognitive difficulties and the problem of information overload, i.e. the 'wall' of reviews (Busch 2016).

Ratings and reviews are not only input for recommender systems. They also represent a private ordering mechanism widely used by online platforms, such as eBay, Amazon, Uber or Airbnb, to build and maintain trust within their community and to preserve the attractiveness of their services.

Ratings and reviews can perform two main functions: (1) informative and (2) self-regulatory.

(1) First of all, they constitute a reputational mechanism that can help reduce information asymmetry between the parties and promote the overall transparency of the transaction (Smorto 2016; Busch 2016; Ranchordás 2018). They represent a source of information which, before the advent of e-commerce, could have been obtained through channels such as advertising, direct experience or recommendations of friends or acquaintances. In this sense, ratings and reviews have codified the 'word of mouth' in the business models, contracts and digital architectures of such **platforms** (Dellarocas 2003).

Recent legislative interventions in Europe have been directed to ensure the transparency of rating and review mechanisms. In the B2C context, the Directive (EU) 2019/2161 introduced the explicit prohibition to submit or commission false consumer reviews or endorsements, as well as manipulate them, in order to promote products. Furthermore, traders (including platforms) have to declare whether and how the review of a product is genuine, i.e. it is submitted by consumers who have actually used or purchased the product.

(2) The second function of ratings and reviews can be the platform’s **self-regulation**. On many platforms, users (both service providers and end-users) assess each other. This bi-directional evaluation is an incentive for users to behave according to the rules of the community and maintain a high online reputation. A series of private sanctions usually complete the rating and review systems: if the user's overall score is below the threshold set by the platform, the personal account can be suspended or deactivated. In some cases, self-regulation is the only function pursued by the platform via the rating (Ducato, 2020). Considering that ratings and reviews can be considered personal data, relating to both the individual who receives the score and the one who gives it to the other user, the data protection framework will apply to this form of **automated-decision making** processing (Ducato, 2020).

References

Consultation Results, ' Ranking transparency guidelines in the framework of the EU regulation on platform-to-business relations – an explainer' (European Commission 2020). Available at:

https://ec.europa.eu/digital-single-market/en/news/ranking-transparency-guidelines-framework-eu-regulation-platform-business-relations-explainer

Aggarwal, Charu C. 2016. Recommender Systems. Vol. 1. Springer.

Busch, Christoph. 2016. "Crowdsourcing Consumer Confidence. How to Regulate Online Rating and Review Systems in the Collaborative Economy." In European Contract Law and the Digital Single Market: The Implications of the Digital Revolution, edited by Alberto De Franceschi, 223–44. Intersentia. https://doi.org/10.1017/9781780685212.013.

Cobbe, Jennifer, and Jatinder Singh. 2019. "Regulating Recommending: Motivations, Considerations, and Principles." Forthcoming, European Journal of Law and Technology (10)3, https://ejlt.org/index.php/ejlt/article/view/686.

Dellarocas, Chrysanthos. 2003. "The Digitization of Word of Mouth: Promise and Challenges of Online Feedback Mechanisms." Management Science 49 (10): 1407–1424.

Ducato, Rossana. 2020. "Private Ordering of Online Platforms in Smart Urban Mobility: The Case of Uber's Rating System", CRIDES Working Paper Series no. 3/2020; forthcoming In Smart Urban Mobility. Law Regulation and Policy, edited by M. Finck, M. Lamping, V. Moscon, H. Reiko. Springer – MPI Studies on Intellectual Property and Competition Law.

Felfernig, Alexander, Ludovico Boratto, Martin Stettinger, and Marko Tkalčič. 2018. Group Recommender Systems: An Introduction. Springer.

Goanta, Catalina, and Gerasimos Spanakis. 2020 "Influencers and Social Media Recommender Systems: Unfair Commercial Practices in EU and US Law." TTLF Working Papers no. 54, https://ssrn.com/abstract=3592000.

Jannach, Dietmar, Markus Zanker, Alexander Felfernig, and Gerhard Friedrich. 2010. Recommender Systems: An Introduction. Cambridge University Press.

Ranchordás, Sofia. 2018. "Online Reputation and the Regulation of Information Asymmetries in the Platform Economy." Critical Analysis of Law (5): 127.

Smorto, Guido. 2016. "Reputazione, Fiducia e Mercati." Europa e diritto privato (1): 199.

Twigg-Flesner, Christian. 2018. "The EU's Proposals for Regulating B2B Relationships on Online Platforms–Transparency, Fairness and Beyond." EuCML (6): 222

74. Red Flag Knowledge

Red flag knowledge is a term of art used in the copyright context to infer knowledge on the part of an online service provider without it having received a specific notice (hence its qualification as "constructive knowledge") about infringing activity which it enables. Although US Congress did not explicitly include it in the Digital Millennium Copyright Act (DMCA), it referred to this term in the legislative history as equivalent to being "aware of facts or circumstances from which infringing activity is apparent", which triggers an obligation of expeditious removal in the safe harbor of hosting, caching services and information location tools established in the Digital Millennium Copyright Act. It clarified that the goal was to exclude from the safe harbor directories that "refer Internet users to other selected Internet sites where pirate software, books, movies, and music can be downloaded or transmitted" when infringement "would be apparent from even a brief and casual viewing."

With the DMCA in force, the term has appeared in a number of cases in US courts, leading to diverging interpretations. For instance, in *UMG Recordings, Inc. v. Shelter Capital Partners LLC*, the Ninth Circuit Court of Appeals held (citing important precedents such as *Sony Betamax* and *Napster*) that online service provide Veho´s protection under the safe harbor was not lost on ground of a general knowledge of the possibility that their platform could be used to share infringement material isn't enough to qualify as Red Flag Knowledge. However, a year later, in *Viacom International v. YouTube*, the Second Circuit Court of Appeals explained that "The difference between actual and red flag knowledge is not between specific and generalized knowledge, but instead between a subjective and an objective standard. In other words, the actual knowledge provision turns on whether the provider actually or "subjectively" knew of specific infringement, while the red flag provision turns on whether the provider was subjectively aware of facts that would have made the specific infringement "objectively" obvious to a reasonable person. The red flag provision, because it incorporates an objective standard, is not swallowed up by the actual knowledge provision under our construction of the § 512(c) safe harbor. Both provisions do independent work, and both apply only to specific instances of infringement.

By contrast, in 2016, in Capitol Records v. Vimeo—a case in which Capitol Records asserted that Vimeo, a platform that allows its users to upload videos, was "not only aware of the copyright infringement taking place on its system, but [was] actively promot[ing] and induc[ing] that infringement … [and] refusing to filter or block videos by using copyrighted recordings"—the Second Circuit held that, even where a copyright owner provides evidence that an online service provider's employee viewed "a video that plays all or virtually all of a recognizable copyrighted song," that evidence is insufficient to establish red flag knowledge: the service provider must have actually known facts that would make the *specific* infringement claimed objectively obvious to a reasonable person. The same Second Circuit, however, has also recognized that a "time-limited, targeted duty" of inquiry to determine whether there is an "objectively obvious" infringement does not run afoul of the prohibition of general monitoring in section 512(m).

Because of the confusion generated by the different articulation of red flag knowledge, the US Copyright Office has recently suggested a clarification in its report on proposed reforms to section 512 of the DMCA. In particular, it has advised clarifying the relationship between such knowledge and the prohibition of general monitoring, and called for a broader notion of knowledge which is not linked to "specific" infringing content.

References

Records, LLC v. Vimeo, LLC, 826 F.3d 788 (2d Cir. 2016).

TOTO, Carolyn, "When It Comes to the DMCA, a Red Flag Becomes Harder to Fly" (2016). Pillsbury - Internet and Social Media Law Blog. Available at https://www.internetandtechnologylaw.com/dmca-red-flag/

US Copyright Office, "Section 512 of title 17: A report of the register of copyrights" (May 2020), available at https://www.copyright.gov/policy/section512/section-512-full-report.pdf

UMG Recordings, Inc. v. Shelter Capital Partners LLC, 667 F.3d 1022 (9th Cir. 2011).

Terrica Carrington. Twenty Years of the DMCA: Notice and Takedown in Hindsight (Part II), Copyright Alliance Blog (28 October 2018), Available at https://copyrightalliance.org/ca_post/twenty-years-dmca-notice-and-takedown/

75. Regulation

Regulation refers to a set of authoritative rules designed to control or govern conduct in a particular sector or domain by restricting or enabling specific activities. There are numerous definitions for this concept, influenced more broadly by ideology and disciplinary traditions. Generally understood as a form of 'command and control' imposed by the state 'through the use of legal rules backed by (often criminal) sanctions' (Black, 2002), regulation needs to be distinguished from the broader notion of "governance". What the majority of these have in common is the understanding of regulation as a form of state action designed to influence business or social behaviour (Baldwin et al., 2012), materialized in the promulgation of a binding set of rules. Back in the 19th century, John Stuart Mill defined regulation as a 'governmental intervention in the affairs of society' (Mill, 1848) and that understanding has prevailed also in relation to Internet platforms, a newer area of regulatory concern.

The opposite move, known as 'deregulation', refers to the reduction or elimination of government power in a particular sector or across the economy in order to foster competition within the industry and reduce the inefficiencies of public regulation. Without signifying a complete withdrawal of the state from defining conditions for rule-making, deregulatory processes for the Internet economy in the early 1990s represented a balancing act between property rights regimes, existing governance structures and rules of exchange (Irion and Radu, 2014). They allowed the growth of Internet services, intermediaries and platforms in the absence of strong public regulation.

An important tension in digital regulation has been the one between hard and soft instruments, between what is legally codified and various attempts to influence behaviour via institutional mandates and modelling (Radu, 2019). While the former exerts direct influence over the conduct of the addressee, soft forms of regulation are indirect means to shape intervention in the private domain, primarily by shaping a normative order (Kettemann, 2020), integrating norms at the domestic, regional and international level.

Regulatory debates have long focused on the relationship between the regulator and the regulated entity. Over time, regulating digital technologies has grown in complexity due to the complex technical expertise required, the high levels of information asymmetry and the risk of 'regulatory capture' (the regulated industry being able to design public rules in its interest). Importantly, regulation happens through sets of practices and it is thus 'collectively mediated and legitimized by the key communities whose buy-in is necessary' (Radu 2019, p. 25).

As technology evolved worldwide, the Internet has seen a regulatory shift, away from the application of general telecommunications rules towards the creation of Internet-specific regulation from the 1990s onwards, culminating in harmonized legislation in the European Union. Plans for state-mandated regulation addressing digital platforms are back in full swing, calling into question the efficacy of self- and co-regulation models (Marsden, 2011).

See also: Co-regulation, Self-regulation

References

Baldwin, R., Cave, M., Lodge, M. (2012). *Understanding Regulation. Theory, Strategy and Practice*. Oxford, UK: Oxford University Press.

Black, J. (2002). Critical reflections on regulation. *Australian Journal of Legal Philosophy*, 27. Available at: http://www.austlii.edu.au/au/journals/AUJlLegPhil/2002/1.pdf

Irion, K., Radu, R. (2013). Delegation to independent regulatory authorities in the media sector: A paradigm shift through the lens of regulatory theory. In: Schulz, W., Valcke, P., Irion, K. (eds.), *The Independence of the Media and Its Regulatory Agencies: Shedding New Light on Formal and Actual Independence Against the National Context* (pp. 15–54). Polity Press.

Kettemann, M. (2020). The Normative Order of the Internet: A Theory of Rule and Regulation Online. Oxford: Oxford University Press.

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

Mill, J. S. (1848). *Principles of Political Economy*. London: John W. Parker, West Strand.

Radu, R. (2019). *Negotiating Internet Governance.* Oxford: Oxford University Press.

76. Remedy

A simple definition is given by legal dictionaries, emphasising the element of “recovery” or “repair”, thus referring to the end-result of a process described as an “effective grievance mechanism” (Le Docte Legal: Dictionary in Four Languages, 2011).

The term "remedy" is formally embedded in many international treaties (eg article 13 of the European Convention on Human Rights [ECHR]; article 2(3;a) of the International Covenant on Civil and Political Rights [ICCPR]; articles 12 and 23 of the Arab Charter on Human Rights [ACHR]) and national public laws.

In human rights law, the provisions on the individual right to an effective remedy are directed at states (see eg the Committee of Ministers/Council of Europe, 2016 Recommendation on Internet Freedom; section 5), as a fundamental guarantee that provides individual persons a legal means of seeking redress in connection to interferences with any of the recognised substantive human rights. As suggested by the Council of Europe’s Guide to Human Rights for Internet Users (2014, 26), ‘The remedy must be effective in practice and in law and not conditional upon the certainty of a favourable outcome for the complainant. Although no single remedy may itself entirely satisfy the requirements of Article 13 [ECHR], the aggregate of remedies provided in law may do so.’ Thus, it includes the positive obligation for the states to effectively respond to human rights issues. As stated in one of the core platform law and policy sources, the "Guiding Principles on Business and Human Rights: Implementing the United Nations “Protect, Respect and Remedy” Framework" (also widely known as the Ruggie-principles; 2011, 27): "Remedy may include apologies, restitution, rehabilitation, financial or non-financial compensation and punitive sanctions (whether criminal or administrative, such as fines), as well as the prevention of harm through, for example, injunctions or guarantees of non-repetition."

In recognition of due process concerns that exist nowadays in the many relationships between online platform providers and the user, the multi stakeholder recommendations of the Internet Governance Forum's Coalition on Platform Responsibility on the implementation of the Right to an Effective Remedy (see IGF-DCPR 2019 Outcome Document) provide the major best practices for the provision of remedies by the responsible actors. As a whole (see especially section D with the relevant provisions on 'Safeguards relating to the implementation of the remedy'), these recommendations suggest that all online platforms should provide a detailed approach in their terms of service, including the offer of measures that are commensurate with the wide scope of internet technologies' impact and with the relevant human rights issues. These safeguards seek to enhance the remedial purpose of alternative **dispute resolution** mechanisms that are provided by the platforms' terms of service - with an additional value in comparison to the above-mentioned human rights' frameworks. Thus, platforms are recommended to foresee the need for continuous **accountability** and **transparency** during the implementation of the remedy and provide the sufficient scaling of the geographical scope of the remedy in order to contribute to tackling the challenge of effectively dealing with **online harm**.

References

See also other terms in this Glossary: no … "Accountability"; no … "Appeal"; no … Dispute Resolution (online); no .. "Harm (online harm)"; no … "Liability"; no … "Moderation"; no … "Pro-active Measures"; no … "Transparency"

Le Docte: Legal Dictionary in Four Languages. 2011. edited by Hans-Werner Zehnhoff AM, Hugues Timmermans, Erika Schmatz, Yvonne Salmon. Antwerpen: Intersentia.

Committee of Ministers/Council of Europe. 2016. Recommendation CM/Rec(2016)5[1] of the Committee of Ministers to member States on Internet freedom (Adopted by the Committee of Ministers on 13 April 201 at the 1253rd meeting of the Ministers' Deputies). Available at <https://search.coe.int/cm/Pages/result_details.aspx?ObjectId=09000016806415fa#_ftn1>.

Convention for the Protection of Human Rights and Fundamental Freedoms, adopted 4 November 4, 1950, entered into force 3 September 1953, ETS No.005 Available at <www.coe.int/en/web/conventions/full-list/-/conventions/treaty/005>.

Council of Europe (2014). 'Guide to Human Rights for Internet Users.' Available at <https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=09000016804d5b31>.

IGF-DCPR 2019 Outcome Document, 'Best Practices on Platforms' Implementation of the Right to an Effective Remedy'. Available at <https://www.intgovforum.org/multilingual/index.php?q=filedepot_download/4905/1550>. (id at: <https://doi.org/10.1016/j.clsr.2019.105379>).

International Covenant on Civil and Political Rights, Adopted, GA (XXI) of 16 December 1966, entered into force 23 March 1976. Available at <https://www.ohchr.org/EN/ProfessionalInterest/Pages/CCPR.aspx>.

League of Arab States, Arab Charter on Human Rights, 22 May 2004, reprinted in 12 Int'l Hum. Rts. Rep. 893 (2005), entered into force March 15, 2008. Available at <http://hrlibrary.umn.edu/instree/loas2005.html?msource=UNWDEC19001&tr=y&auid=3337655>.

'Guiding Principles on Business and Human Rights: United Nations "Protect, Respect and Remedy" Framework', UN, HR/PUB/11/04. Available at <www.ohchr.org/Documents/Publications/GuidingPrinciplesBusinessHR_EN.pdf>. (see also Ruggie J. 2011, March 21. '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). Available at <https://documents-dds-ny.un.org/doc/UNDOC/GEN/G11/121/90/PDF/G1112190.pdf?OpenElement>.)

77. Repeat Infringer

The concept of "Repeat infringement" has been established by the Digital Millennium Copyright Act (DMCA or the Act), passed by the US in 1998. This piece of legislation originally aimed at protecting digital innovators while preserving the ability of copyright holders to prevent activities that may infringe upon their rights.

Most relevantly, DMCA section § 512 grants so-called "**safe harbour**" protections to the intermediaries that act on actual or constructive knowledge of copyright infringement and "adopt and reasonably implement, and inform subscribers and account holders [...] of, a policy that provides for the termination in appropriate circumstances of subscribers and account holders [...] who are repeat infringers."

In this perspective, "repeat infringer" refers to the number of times a user has been identified as an infringer. According the DMCA § 512 a written repeat infringer policy, consisting of a set of guidelines that detail when a users' infringing activity will result in termination of their account access. In this policy, the intermediary must explicit how often a user must be successfully accused of copyright infringement before the account for the user is terminated. (Sawicki, 2006)

Moreover, to take advantage of the safe harbour protections, an intermediary must "reasonably implement" the repeat infringement policy. As such, upon receiving a copyright infringement notice, demanding takedown of specific content, both the complaint and its outcome must be recorded, to be able to identify when the infringer repeats their infringement. The intermediary records allow to identify users who accumulate a quantity of infringements deemed as sufficient to trigger the repeat infringer policy, which imply the closure of the user account and the blocking of his or her IP address.

References

Sawicki, A. (2006) "Repeat Infringement in the Digital Millennium Copyright Act," University of Chicago Law Review: Vol. 73: Iss. 4, Article 7. https://chicagounbound.uchicago.edu/cgi/viewcontent.cgi?article=5386&context=uclrev

78. Responsibility

The concept of responsibility refers, in its simplest form, to a duty to undertake a particular action or set of actions. Such duty can be legal, but also moral, social or ethical. If it is legally enforceable, failing to fulfill the duty gives rise to **liability**. However, even where that enforcement is not available, failing to fulfil one´s responsibility can give rise to significant consequences from a legal, social and even financial standpoint. For instance, a boycott of advertisers (also known as "adpocalypse") took place in 2016 due to an alleged failure in Youtube´s responsibility to prevent ads from being associated with terrorist content. A similar boycott, known as "Stope Hate for Profit", occurred in 2020 due to Facebook´s failure to take responsibility for the incitement to violence against protesters fighting for racial justice in America in the wake of George Floyd, Breonna Taylor, Tony McDade, Ahmaud Arbery, Rayshard Brooks and many others.

Platform responsibility is the concept that brought together a variety of stakeholders leading to the establishment of the DCPR in 2014. As noted in the DCPR Outcome book in 2017, facing the proliferation of private ordering regimes in online platforms, stakeholders began to interrogate themselves about conceptual issues concerning the moral, social and human rights responsibility of the private entities that set up such regimes. The use of this notion of "responsibility" has not gone unnoticed, having been captured for example by the special report prepared by UNESCO in 2014, the study on self-regulation of the Institute for Information Law of the University of Amsterdam, the 2016 Report of the UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, the Center for Law and Democracy's Recommendations on Responsible Tech and the Council of Europe's draft Recommendation on the roles and responsibilities of Internet intermediaries.

The declination of responsibilities for a particular stakeholder is typically linked to the "role" that can be attributed to it in a particular process or system: in Internet governance, this goes back to the 2005 Tunis Agenda for Information Society, which established the attributions of different stakeholders in the management of the Internet, recognizing in particular that governments should have an equal *role and responsibility* for international Internet governance and for ensuring the stability, security and continuity of the Internet. Practically, this is a careful choice of wording as it does not articulate a corresponding liability, while still calling governments for action in a particular domain. Although the notion of responsibility can be exemplified or explained with reference to specific forms of **due diligence** or even of a **duty of care**, this typically remains the task of adjudicators to make those articulations in defining the scope of responsibility. Occasionally, this interpretation can be facilitated by authoritative guidelines. An example is the articulation of the due diligence process expected to be followed by businesses in relation to their human rights impact, in particular (a) Identifying and assessing actual or potential adverse human rights impacts that the enterprise may cause or contribute to through its own activities, or which may be directly linked to its operations, products or services by its business relationships; (b) Integrating findings from impact assessments across relevant company processes and taking appropriate action according to its involvement in the impact; (c) Tracking the effectiveness of measures and processes to address adverse human rights impacts in order to know if they are working; and (d)

Communicating on how impacts are being addressed and showing stakeholders – in particular affected stakeholders – that there are adequate policies and processes in place. This guidance is provided by the Guiding Principles on Business and Human Rights, unanimously endorsed by the UN Human Rights Council in 2011, which establish a clear separation between the *duty* of States to protect human rights, the *responsibility* of businesses to respect them, and the joint duty of both to provide effective remedies.

References

Stop Hate for Profit. Available at: < https://www.stophateforprofit.org/>

Luca Belli, Nicolo Zingales, 'Online Platforms' Roles and Responsibilities: A Call for Action', in L. Belli & N. Zingales (eds.),*Platform Regulations: How Platforms Are Regulated and How They Regulate Us* (FGV Press, 2017), 21-32

Tunis Agenda for the Information Society (18 November 2005), WSIS-05/TUNIS/DOC/6(Rev. 1)-E, Available at https://www.itu.int/net/wsis/docs2/tunis/off/6rev1.html

Guiding Principles on Business and Human Rights: United Nations "Protect, Respect and Remedy" Framework

HR/PUB/11/04

Report of the the Special Rapporteur to the Human Rights Council on Freedom of expression, states and the private sector in the digital age, A/HRC/32/38 (11 May 2016) <https://documents-dds-ny.un.org/doc/UNDOC/GEN/G16/095/12/PDF/G1609512.pdf?OpenElement>

Center for Law & Democracy, 'Recommendations for Responsible Tech' <http://responsible-tech.org/wp-content/uploads/2016/06/Final-Recommendations.pdf>

Council of Europe, Recommendation CM/Rec(2017x)xx of the Committee of Ministers to member states on the roles and responsibilities of internet intermediaries. https://rm.coe.int/recommendation-cm-rec-2017x-xx-of-the-committee-of-ministers-to-member/1680731980

79. Revenge pornography / Non-consensual intimate images

NB: While the term "revenge pornography" is commonly used, many argue that it is inappropriate since it suggests a degree of complicity or consent on the part of the person in the images. Instead, terms such as "non-consensual intimate images" are now considered to more appropriate describe the phenomenon and is the term used here.

This entry: (i) sets out examples of definitions of the term and (ii) provides examples of existing regulatory responses to non-consensual intimate images.

(i) Examples of definitions of the term

Non-consensual intimate imagery can be broadly understood as the distribution of sexually explicit images or video of an individual without their consent. Unlikes other forms of intimate imagery, such as consensual pornography, which existed prior to the internet, the distribution of non-consensual intimate imagery is a more recent phenomenon, spawned by "changes in technology, including the advent of social media and websites that feature user-generated content, in conjunction with the ease of taking and sharing digital photographs and video (particularly on smartphones, tablets, and mobile devices)" (Magaldi et al., 2020).

Legal definitions of the phenomenon, particularly those that criminalise such distribution, often include further elements and exceptions, for example, a requirement that there be an intention to cause harm or distress, that the individual in the image or video had a reasonable expectation of privacy, and that no legitimate purpose to the distribution (such as for law enforcement purposes).

(ii) Existing regulatory responses

A number of states and subnational jurisdictions governments have prohibited distributing non-consensual intimate images through the creation of criminal offences (There are a number of databases of criminal laws in place, for example: The Center for Internet and Society and InternetLab – see references). While the specific wording may vary, common to all criminal offences are the core requirements that (i) a person disseminates an image or video of another person, (ii) that image or video is sexually explicit; and (iii) the dissemination is done without the consent of the other person.

As noted above, some states have also included further elements and exceptions. In Canada, for example, the images or video must have been taken with a "reasonable expectation of privacy" (section 162.1 of the Criminal Code). In England and Wales, the person distributing the images or videos must have done so with "an intention of causing [the] individual distress" (section 33 of the Criminal Justice and Courts Act 2015), and in South Africa, there must have been an "intent to cause them harm" (section 18F of the Films and Publications Act).

Beyond outright criminalisation, there are few examples of regulatory responses, particularly when it comes to the liability of online platforms which are used to share non-consensual intimate images. One example is Article 21 of Brazil's Civil Marco da Internet, which provides for a specific liability regime in cases involving "the breach of privacy arising from the disclosure of images, videos and other materials containing nudity or sexual activities of a private nature, without the authorisation of the participants". In such instances, a platform can be held liable for such material where they fail to remove the content, in a diligent manner, and within its own technical limitations, when notified of it by the individual involved or their legal representative (i.e. a "notice and takedown" regime).

References

Magaldi, Jessica A. and Sales, Jonathan S. and Paul, John, 'Revenge Porn: The Name Doesn't Do Nonconsensual Pornography Justice and the Remedies Don't Offer the Victims Enough Justice' [January 29, 2020]. Oregon Law Review, Vol. 98, No. 1, 2020. Available at: https://ssrn.com/abstract=3527819

The Center for Internet and Society, 'Revenge Porn Laws across the World', available at: https://cis-india.org/internet-governance/blog/revenge-porn-laws-across-the-world

InternetLab, 'How do countries fight the non-consensual dissemination of intimate images?' [2018]. Available at: https://www.internetlab.org.br/en/inequalities-and-identities/how-do-countries-fight-the-non-consensual-dissemination-of-intimate-images/

80. Right to explanation

The concept of right to explanation refers to the informational duties owed by a data controllers to a data subject in relation to automated decisions based on profiling. The basic provision for the construct of the "right to explanation" is Article 22 of the General Data Protection Regulation (GDPR), which establishes a right for any data subject 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". This provision is the evolution of article 15 of the Data Protection Directive (DPD), in turn finding its historical root in the French law of 1978 "on computing, files and freedoms" which provided a *broader* right not to be subject to *any* decision involving an appraisal of human behavior based solely on the automated processing of data which describes the profile or personality of the individual. The same law also granted the right to know and challenge the information and reasoning used in such processing in case the data subject opposed the results.

While the scope of this right is much narrower both in art 15 DPD and art 22 GDPR, one can discern from the Directive's *Travaux Préparatoires* the same concern for human dignity- specifically that humans maintain the primary role in 'constituting' themselves instead of relying entirely on (possibly erroneous) mechanical determinations based on their "data shadow" 49. Arguably, that concern underlies art. 22 GDPR despite the more specific focus in its *Travaux Préparatoires* on the risks of decisions *based on profiling*, which is defined in art. 4 (4) as "any form of automated processing of personal data consisting of using those data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements". Accordingly, the explicit mention of profiling could be interpreted simply as illustrative of one of the possible risks involved in automated processing.

Another important difference of art 22 GDPR from the text of art 15 DPD is the requirement of "suitable" safeguards in case of application of any derogations to the right established art 22 (1), which are admitted only where provided by a law to which the controller is subject. "Suitable measures" to safeguard the data subject's rights and freedoms and legitimate interests are also required when applying one of the exceptions to the rule laid down in art 22 (1), i.e. necessity for entering into a contract or performance thereof, and data subject's explicit consent –a novelty introduced by the GDPR. Detailing the application of these exceptions, but arguably also informing the application of derogations, article 22 (3) specifies that "suitable measures" consist *at least* of the right of the data subject "to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision".

This is an aspect that has triggered significant discussion on the existence of a right to explanation52, as the suitable safeguards listed in Recital 71 of the GDPR include the right "to obtain an explanation of the decision reached after such assessment", but this wording is conspicuously absent from the text of art. 22 (3). Regardless of the qualification of the right provided by art 22 as one to “an explanation” (as we’ll call it here for sake of simplicity), to “information”or to “legibility”, it is indisputable that the article seeks to put data subjects in the position to appreciate at least to a certain degree the logic of any algorithm relied upon to take measures which significantly affect them. Besides the obvious question of what is the requisite degree of transparency and granularity of an explanation, the provision leaves ambiguous two important issues: (1) whether art 22 implies a prohibition of processing personal data without fulfilling the relevant criteria, or rather a right for the data subject to actively object to such processing; and (2) whether the requisite degree of transparency and granularity for requested data33, and last but not least, the possibility for data controllers to condition access requests to the payment of a fee.

Even admitting that consumers were able to use and keep perfect track of all the revealed data, they would still largely ignore predictions and decisions made on the basis of such data, at least in the absence of proactive measures taken by data controllers to that effect. EU data protection law specifically addresses this problem establishing the right for individuals to obtain basic knowledge on the logic of any automated decisions that produces a legal effect or significantly affect them otherwise, and a right not to be subjected to such decisions outside a narrow set of circumstances. Regrettably, the right to obtain such knowledge has historically been under-enforced, mainly due to the ambiguity of article 15 of the Data Protection Directive concerning its scope of application. However, there is reason to think that this situation will change in light of the forthcoming General Data Protection Regulation (GDPR), which details theminimum safeguards that should be offered when such decisions are made.

References

Wachter, B. Mittelstadt, and L. Floridi, ‘Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation’ (2017) International Data Privacy Law (2017).

L. Edwards and M. Veale, ‘Slave to the algorithm? Why a “right to an explanation” is probably not the remedy you are looking for’ 16 Duke Law and Technology Review (2017)18; Bygrave, supra n 35.

B. Goodman and S. Flaxman, ‘European Union regulations on algorithmic decision- making and a "right to explanation"’, ICML Workshop on Human Interpretability in Machine Learning, preprint, arXiv:1606.08813 (v3) (2016); AI Magazine (2017).

Malgieri, Gianclaudio and Comandé, Giovanni, Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation (November 13, 2017). International Data Privacy Law, vol. 7, Issue 3, Forthcoming, Available at SSRN: https://ssrn.com/abstract=3088976

A. Selbst and J. Powles, ‘Meaningful information and the right to explanation’, 7 (4) International Data Privacy Law, 233

I Mendoza and L A Bygrave, 'The Right not to be Subject to Automated Decisions based on Profiling', in T Synodinou, P Jougleux, C Markou, T Prastitou (eds.), *EU Internet Law: Regulation and Enforcement* (Springer, 2017); University of Oslo Faculty of Law Research Paper No. 2017-20. Available at SSRN: https://ssrn.com/abstract=2964855, p. 6.

Malgieri, Gianclaudio, Automated Decision-Making in the EU Member States: The Right to Explanation and Other 'Suitable Safeguards' for Algorithmic Decisions in the EU National Legislations (August 17, 2018). Computer Law & Security Review, 2019 Forthcoming, Available at SSRN: https://ssrn.com/abstract=3233611 or http://dx.doi.org/10.2139/ssrn.3233611

81. Right to be forgotten

The right to be forgotten is a legal right that exists in some jurisdictions, and that is closely related to the right of privacy and the protection of personal data. This right generally enables people to require removal of information about them that is stored or otherwise processed by others. It can be argued that the right existed before the internet and the rise of online platforms, and that it could be read into pre-internet norms and case law. However, the right to be forgotten has gotten particular prominence with the development of digital information technologies. The ability to store and search great amounts of information has shifted the 'default of forgetting' information towards a 'default of remembering' (Mayer-Schönberger 2009). This required a rethinking of how we deal with private information and personal data in computer mediated interactions and spurred the development of the right to be forgotten.

The right to be forgotten gained a lot of attention when it was read into the European Union Data Protection Directive – the predecessor of the General Data Protection Regulation (GDPR) – by the European Court of Justice in the Google Spain (InfoCuria, 2014) case. In its decision, the Court acknowledged that people have the right to have search results to irrelevant or outdated information about them delisted. Not all references to such information need to be removed under the ruling, as the Court explained that the right to be forgotten does not apply in cases in which there is a public interest in accessing the information in question. That may be the case if the person in question plays a role in public life, for instance because that person is a politician or celebrity. The right to be forgotten thus requires that a balance is struck between a person's right to privacy and other people's rights to share and access information (Kulk and Zuiderveen Borgesius 2017). After the *Google Spain* case, search engines have implemented forms that enable people to do removal requests, that are then processed by these search engines. As of July 2020, Google has received almost 950.000 requests concerning a total of 3.700.000 URLs.

The right to be forgotten (or 'right to erasure') is enshrined in the European Union by means of the GDPR. It applies broadly to any kind of processing of 'personal data' – not just by search engines – in cases where such processing is unlawful or no longer necessary (Ausloos 2020). The California Consumer Privacy Act of 2018 established a 'right to deletion' as well. In both laws, freedom of expression is recognized as an exception to these rights.

References

C-131/12 - Google Spain e Google, 'Processo principal' (InfoCuria 2014) < http://curia.europa.eu/juris/liste.jsf?num=C-131/12>

J. Ausloos, 'The Right to Erasure in EU Data Protection Law', [2020] Oxford University Press.

V. Mayer-Schönberger, 'Delete: the virtue of forgetting in the digital age', [2009] Princeton University Press.

S. Kulk & F.J. Zuiderveen Borgesius, Privacy, ‘Freedom of Expression, and the Right to Be Forgotten in Europe’, [2017] *Cambridge Handbook of Consumer Privacy* (eds. Jules Polonetsky, Omer Tene, and Evan Selinger), Cambridge University Press.

82. Safe harbor

A safe harbor is an area of exemption from liability guaranteed by the legal system, a “comfort zone” that enables the pursuit of what would otherwise be considered legally risky activities. Safe harbors can be important not only to enable experimentation and innovation, but also, and particularly in the context of speech platforms, to allow the free flow of information.

The scope, limits and conditions of safe harbors for digital intermediaries are, indeed, a crucial topic of Internet law and governance. There are several models with different characteristics, but a distinction should at least be made between vertical safe harbors, whose exemption applies to only a particular type of liability (for example, liability for copyright infringement, as in the US Digital Millennium Copyright Act, or DCMA) and horizontal, which establish an exemption applicable across different types of liability. However, it is worth noting that very rarely is a safe harbor entirely horizontal, as there are typically several exempted areas: for instance, the safe harbor established in section 230 of the US Communication Decency Act (CDA) explicitly excludes from its scope federal criminal law, intellectual property law, communications privacy law, sex trafficking law, and more generally U.S. state law. Similarly, the EU E-Commerce Directive (ECD) excludes from its scope taxation, data protection law, cartel law, the activities of notaries or equivalent professions to the extent that they involve a direct and specific connection with the exercise of public authority, the representation of a client and defence of his interests before the courts, and gambling activities which involve wagering a stake with monetary value in games of chance. In addition, the Directive explicitly preserves the ability of Member States to impose reasonable duties of care to detect and prevent certain types of illegal activities, and there are also differences in the type of knowledge of illegal activity that is sufficient to fall foul of the safe harbor with regard to criminal matters (actual v. constructive knowledge that applies to civil matters).

We can identify 2 different types of models regarding the conditions to be satisfied for the enjoyment of safe harbors for intermediary liability: one that applies broadly to a range of intermediaries, and another one that identifies specific safe harbors for different types of hosts. For instance, the CDA safe harbor applies to any provider or user of an interactive computer service, defined as an information service, system, or access software provider that provides or enables computer access by multiple users to a computer server, including specifically a service or system that provides access to the Internet and such systems operated or services offered by libraries or educational institutions. Similarly, Act No. 137 of 2001 on the Limitation of Liability of Intermediaries in Japan defined as provider of a service whose purpose is to communicate third party information to other parties, and the Indian IT Act defines an intermediary (which can benefit from the safe harbor) as “any person who on behalf of another person receives, stores or transmits that record or provides any service with respect to that record and includes telecom service providers, network service providers, internet service providers, web-hosting service providers, search engines, online payment sites, online-auction sites, online-market places and cyber cafes.

By contrast, the ECD provides much more specific protections to communications conduits, content hosts, and caching services, while the US DMCA includes in addition to that the categories of information location tools and non-profit educational institutions.

Another crucial point is what conditions must be fulfilled in order for the intermediary to enjoy liability. Broadly, two different models can be distinguished: one that is premised on good faith, as in the CDA, or other forms of context-dependent knowledge; and another which is premised upon a qualified **notice**. Safe harbors can also have a combination of the two: for instance, the US DMCA introduces a **notice and takedown** framework while also carving out from the liability exemption case of both actual and constructive (and thus more context-dependent) knowledge of illegality. Similarly, the Finnish, the Chilean and the Brazilian system introduce a system based on judicial notice, but contain exceptions for situations where constructive knowledge is admitted.

83. Self injury

Self injury or self harm is often described as the act of purposely hurting physically or psychologically oneself as an emotional coping mechanism (Skegg, 2005; Daine et. al, 2013). These behaviors are not just suicidal-related, the practice of violent activities and challenges with a high risk to life and self-harm are also a part of this definition, such as feeding disorders and self-deprecation.

Currently, cyberbullying represents a great source of suffering and many people, who are afraid or because they think that no one will help or even that talking can make the situation worse, live with this type of violence (Lindert, 2017). Besides that, prejudicial beliefs about mental health are quite common, which not only contributes to the increased stigma and taboo, as it hinders the search for help when there is intense suffering (Scavacini, 2019).

The concern with self injury and the internet is not new. In fact, a survey published in 2006, in the scientific publication of the American Academy of Pediatrics, already pointed out that the spread of this practice on internet forums (Whitlock et. al, 2006). In Brazil, a survey published on 2019 showed that about 15% of internet users aged 11 to 17 years old had access to content on suicide or self-harm (Cetic.br, 2019).

Despite not being proven effective, an attempt to address this problem has been to provide through criminal law for specific penalties for those who via the internet (through social media or live broadcast) encourage someone to commit suicide or self-harm or provide material assistance to do so.

More recently, livestream suicides have re-launched discussions about the alleged lack of control over what is posted on online platforms (Ribeiro, 2019). In this sense, several platforms have consolidated strategies for dealing with self injury content. In addition to user support and artificial intelligence filters, the platforms have teams focused on collaborating with local authorities.

Finally, human moderation of this type of content is an unhealthy job (Newton, 2019; Rocha, 2019), and the number of information about workers who developed post-traumatic stress syndrome as a result of this activity is growing and it shouldn't go unnoticed (Cardoso, 2019).

References

Cardoso, Paula (2019). Precariado algorítmico: o trabalho humano fantasma nas maquinarias da inteligência artificial. Media Lab UFRJ. Available at: <http://medialabufrj.net/blog/2019/09/dobras-38-precariado-algoritmico-o-trabalho-humano-fantasma-nas-maquinarias-da-inteligencia-artificial/>

Centro de Estudos sobre as Tecnologias da Informação e da Comunicação (CETIC.br) (2020). Pesquisa sobre o uso da internet por crianças e adolescentes no Brasil: TIC kids online Brasil.

Available at: <https://cetic.br/media/analises/tic_kids_online_brasil_2019_coletiva_imprensa.pdf>

Daine, K., Hawton, K., Singaravelu, V., Stewart, A., Simkin, S., & Montgomery, P. (2013). The Power of the Web: A Systematic Review of Studies of the Influence of the Internet on Self-Harm and Suicide in Young People. PLoS ONE, 8(10), e77555. doi:10.1371/journal.pone.0077555

Facebook. 'Community Standards' Available at: <https://www.facebook.com/communitystandards/suicide_self_injury_violence>

Instagram. 'Self-Injury'. Available at: <https://help.instagram.com/553490068054878>

Lindert, Jutta. Cyber-bullying and it its impact on mental health (2017). European Journal of Public Health, Volume 27, Issue suppl_3, November, ckx187.581. Available at: <https://doi.org/10.1093/eurpub/ckx187.581>

Newton, Casey (2019). The Trauma Floor: The secret lives of Facebook moderators in America. The Verge. Available at: <https://www.theverge.com/2019/2/25/18229714/cognizant-facebook-content-moderator-interviews-trauma-working-conditions-arizona>

Ribeiro, Paulo Victor (2020). After Livestreamed Suicide, TikTok Waited to Call Police. The Intercept. Available at: <https://theintercept.com/2020/02/06/tiktok-suicide-brazil/>

Rocha, Camilo (2019). O trabalho humano escondido atrás da inteligência artificial. Nexo. Available at:: <https://www.nexojornal.com.br/expresso/2019/06/18/O-trabalho-humano-escondido-atr%C3%A1s-da-intelig%C3%AAncia-artificial>.

Scavacini, Karen (2019). Prevenção do suicídio na internet: pais e educadores. Available at: <http://www.safernet.org.br/site/themes/sn/sid2017/resources/AF_cartilha_Pais.pdf>

Skegg, Keren (2005). Self-harm. The Lancet, 366(9495), 1471–1483. doi:10.1016/s0140-6736(05)67600-3

Twitter. 'Suicide and Self-harm Policy'. Available at: <https://help.twitter.com/en/rules-and-policies/glorifying-self-harm>

Whitlock, Janis; Eckenrode, John & Silverman, Daniel (2006). Self-injurious Behaviors in a College Population. Available at: <doi:10.1542/peds.2005-2543>

84. Self-preferencing

Self-preferencing is the practice of giving preferential treatment to one's own complementary products or services when they are in competition with products and services provided by other entities using the platform (Crémer 2019). The term rose to prominence after the Google Shopping case (2017) where the European Commission accused Google of self-preferencing its own shopping results over those of other competing shopping comparison services. However, self-preferencing is thought to encompass various practices, many of which are not new, such as refusal to deal or tying.

The central concern around self-preferencing is that a dominant platform will leverage its power in the platform market either to expand its power in neighboring markets or to protect its dominant position in the home platform market (Graef 2019). The former is more common when the platform is vertically integrated and wants to establish itself or protect its position in the neighboring market, whereas the latter consideration occurs when the platform wants to protect itself from competitive entry or expansion in the home market.

Self-preferencing can manifest itself in different ways, some well-known and some newer and less well-understood. Among the traditional practices that can result in non-affiliated products and services being in a disadvantageous position compared to the platform's own products or services (or those affiliated with it) are refusal to supply, tying, abusive discrimination, and margin squeeze (Ahlborn 2020). For these practices there are well-developed legal standards.

Newer practices have included the manipulation of the ranking of affiliated products and services compared to those of non-affiliated competitors, and the use of data from competitors who rely on the platform to improve the platform's own products and services. The proper tests for when such practices should be considered problematic have yet to be developed. Among the relevant parameters that can be considered are whether the platform is indispensable to the non-affiliated products and services, whether the platform is dominant, the rationale and design of the self-preferencing, the extent of the negative effects of the self-preferencing, and whether it has any pro-competitive justifications (Ahlborn 2020, Zingales 2019).

References

Ahlborn, Christian, Will Leslie & Eoin O'Reilly. "Self-Preferencing: Between a Rock and a Hard Place." CPI Antitrust Chronicle (June 2020).

Crémer,Jacques, Yves-Alexandre de Montjoye & Heike Schweitzer. "Competition Policy for the Digital Era." Final Report for the European Commission (2019).

Case AT.39740, Google Search (Shopping), 27 June 2017.

Graef, Inge. "Differentiated Treatment in Platform-to-Business Relations: EU Competition Law and Economic Dependence." Yearbook of European Law 38 (2019): 448-499.

Zingales, Nicolo. “Google Shopping: Beware of ‘Self-favouring’ in a World of Algorithmic Nudging.” CPI Europe (February 2018).

85. Self-regulation

Self-regulation refers to rules that are autonomously developed and implemented by the thereof concerned persons, independently from any structured form of rule-making. The legitimacy of self-regulation is based on the fact that private incentives lead to a need-driven rule-setting process. Self-regulation is responsive to changes in the environment and can establish rules without regard to the territoriality principle. It is justified (i) if its application leads to a higher efficiency than provided by governmental law and (ii) if compliance with the community rules is less likely than compliance with private rules (Weber 2014: 23; Gibbons 1997: 509; Black 1996: 32 seq.).

Depending on the given circumstances, the concerned persons can belong to a single or to different market level(s). Guidelines imposing specific obligations on Internet Services Providers (ISP, for example in respect of notice and take down) only influence this professional category. In the context of some types of speech, user groups of a technology and platform “owners” might develop different soft law regimes; user groups could agree on certain expressions to be avoided, platform “owners” on control measures regarding uploaded contents thereby replacing governmental regulation (see also the contributions in Belli and Zingales, 2018: 41 et seq).

A universally accepted theory as to the “legal quality” of self-regulation has not (yet) been formulated. Since self-regulation is not enforceable through public action, such rules do not have the quality of law in the traditional sense (Weber 2002: 81-83; Guzman and Meyer 2010: 179-183; Abbott and Snidal: 2000: 429). The often used terms “contract” and “social contract” also do not fit. Self-regulation can be understood as a social control model or as a gentlemen’s agreement (Gibbons 1997: 519 et seq.). But compliance with its rules is usually more than only an ethical undertaking, even in the absence of direct sanction. Self-regulatory provisions correspond to standards that reflect the common sense behavior expected to be observed by the concerned person, i.e. – in overcoming the dichotomy to “hard law” – they represent the due diligence and good faith standards being legally enforceable (Weber 2014: 25).

The strengths of self-regulation encompass the following elements being also relevant for platforms (Weber 2002: 83/84 with further references): (i) The rules created by the participants of a specific community are efficient since they respond to real needs and mirror the technology. (ii) Meaningful self-regulation provides the opportunity to adapt the regulatory framework to the changing technology. (iii) Self-regulation can usually be implemented at reduced cost. (iv) Due to the private initiatives, the chances are high that the rules contain incentives for compliance. (v) Effective self-regulation induces the concerned persons to be open to a permanent consultation process related to the development and implementation of the rules (i.e. a mechanism that helps to accurately reflect real needs).

The weaknesses of self-regulation include the following aspects (Weber 2002: 84/85; Brown and Marsden 2003: 2; Tambini, Leonardi and Marsden 2008: 269-282): (i) Self-regulation is not generally binding in legal terms, the provision are only applicable to those persons that have accepted the regulatory regime. (ii) Self-regulation tends to be based on a case-driven approach rather than general rules. (iii) If the number of “outsiders” or “dark sheep” not acknowledging the self-regulation is large, the legitimacy of the respective rules becomes doubtful. In contrast, “outsiders” not being involved in the preparation and implementation of the rules can benefit from them free of charge (“free rider”). (iv) Self-regulatory mechanisms are not always stable and can depend on the concerned (market) segment; consequently, the applicable standards risk to remain on a low level. (v) The main problem of self-regulation lies in the lack of enforcement proceedings; non-compliance with private rules does not necessarily lead to sanctions.

In order to overcome the described weaknesses of self-regulation, new models have been developed in different forms of co-operative rule-making or co-regulation (see no. 15). In reality, self-regulation already exists in many fields, traditionally in the media and the Internet environment as well as in the banking markets, however, the regulation of platforms can also be suitably done by way of private rule-making. In this field self-regulation is usually based on Codes of Conduct and Terms of Use (mostly phrased by the providers of the platforms).

References

Abbott Kenneth W. and Duncan Snidal (2000). ‘Hard and Soft Law in International Governance’, in *International Organization* 54 (2000), pp. 421-456. Available at https://www.cambridge.org/core/journals/international-organization

Belli Luca and Nicolo Zingales (eds.) (2017), ‘Platform Regulations. How Platforms are Regulated and How They Regulate Us’, Geneva 2017

Black Julia (1996). ‘Constitutionalizing Self-Regulation’, *The Modern Law Review* 59 (1996), pp. 24-55. Available at https://www.modernlawreview.co.uk

Brown Ian and Christopher T. Marsden (2013). ‘Regulating Code: Good Governance and Better Regulation in the Information Age’, Cambridge MA/London 2013

Gibbons Llewellyn J. (1997). ‘No Regulation, Government Regulation, or Self-Regulation: Social Enforcement or Social Contracting for Governance in Cyberspace’, *Cornell Journal of Law and Public Policy* 6 (1997), pp. 475-551. Available at http://scholarship.law.cornell.edu/cjlpp

Guzmann Andrew T. and Timothy L. Meyer (2010). ‘International Soft Law’, *Journal of Legal Analysis* 2 (2010), pp. 171-225. Available at https://academic.oup.com/jla

Tambini Damian, Danilo Leonardi and Chris Marsden, ‘Clarifying Cyberspace: Communications self-regulation in the age of Internet convergence’, London 2008

Weber Rolf H. (2002). ‘Regulatory Models for the Online World’, Zurich 2002

Weber Rolf H. (2012), ‘Overcoming the Hard Law/Soft Law Dichotomy in Times of (Financial) Crisis’, *Journal of Governance and Regulation* 1 (2012), pp. 8-14. Available at https://virtusinterpress.org/-Journal-of-Governance-and-Regulation-15-.html

Weber Rolf H. (2014), ‘Realizing a New Global Cyberspace Framework’, Zurich 2014

86. Shadowban / Shadow banning

A shadowban refers to a relatively common moderation practice of lowering a user's visibility, content or ability to interact without them knowing it so that they can continue to use theplatform normally but their content is not visible to anyone else. It can refer to user-driven or platform-driven blocking, and its meaning has expanded to include visibility in algorithmically-determined platform features. Shadowbanning is often a response to toxic or undesirable interaction related to a specific user. It can be observed by, and has been attributed to, the real or perceived visibility of an account.

Shadowbans can be implemented by individuals or the tech platform itself.

Some tech platforms allow individual users to restrict the visibility and engagement of other users with one’s own account, often implemented as an anti-harassment tool. Instagram and Twitter allow users to block other users without them knowing. In 2019, Instagram launched (Grothaus, 2019) a restrict feature to enable its users to block individuals without that user being notified or made aware.

Platform-instigated shadowbans reduce the visibility of content from the affected user both in terms of the posts themselves as well as through algorithmic restrictions on amplification. For example, a shadowbanned account may no longer appear in algorithmically-determined recommendations for content, in search terms or autofill recommendations, or in a list of suggested accounts. Shadowbans can reduce search visibility or prevent an account from being indexed; or it is only visible to that specific user.

Users who are shadowbanned can continue using their accounts as usual. Accounts that are shadow banned typically appear to be functioning normally for the affected user, but its content is undiscoverable and it may be unable to engage in certain ways. For example, on Reddit, the votes of shadowbanned accounts do not count, while on Twitter their engagement will be visible only to the user but not to the account holder or public.

The concept of restricting content from a user without their awareness has a long history in internet culture as an approach for reducing or deterring undesirable content or communication, such as spam, trolling, or harassment. The approach as a moderation technique has been around as long as there have been public discussion spaces, though the term did not come into use until the mid-2000s. Reddit (Reddit, 2015) used shadow banning until 2015 to “punish” (Shu , 2015) users who broke the rules and deter spam. Shadowbanned accounts could continue to post but their content would only be visible to that user, meaning that they kept posting content without realizing their accounts were banned. In 2016, the right-wing media outlet Brietbart published what it called an expose revealing that Twitter (MILO, 2016) and Facebook (Bokhari, 2018) used shadowbanning to silence conservative voices. A 2018 Vice (Thompson, 2018) article, that was later discredited (Stack, 2018), claimed that Republican figures had been intentionally removed from Twitter’s auto-populated drop-down search menu. The term shadowban entered the popular lexicon as a highly politicized term when U.S. President Trump used the term to condemn unfounded claims that social media firms were shadowbanning Conservative users and threatened (Molina, 2018) to investigate social media companies.

References

Michael Grothaus, ' Here’s how to shadow ban your Instagram bullies with the new ‘Restrict’ feature' (Fast Company 2019). Available at: < https://www.fastcompany.com/90412178/instagram-rolls-out-restrict-feature-allowing-users-to-shadow-ban-bullies>

Reddit, 'On shadowbans.' (Reddit 2015) < https://www.reddit.com/r/self/comments/3ey0fv/on_shadowbans/>

Catherine Shu, ' Reddit Replaces Its Confusing Shadowban System With Account Suspensions' (Tech Crunch 2015) < https://techcrunch.com/2015/11/11/reddit-account-suspensions/?guccounter=1&guce_referrer=aHR0cHM6Ly9lbi53aWtpcGVkaWEub3JnLw&guce_referrer_sig=AQAAALQDh-jllBW53kqlmJiwGIKcCkdHfrRIaOy8-cKg9LUg4_EKR_kLqqDNe5E4nw7pvSg3BUJXad_CqtAqq2RbH7vlBts8e3pLWy34BtXwhIjSP6qrMpP1NcmiJacUCExGZNjcHpfaKBIjZ5VXIgG67OecnM5FZNxn53kwNfMcHOwN>

MILO, ' EXCLUSIVE: Twitter Shadowbanning ‘Real And Happening Every Day’ Says Inside Source' (BreitBart 2016) < https://www.breitbart.com/tech/2016/02/16/exclusive-twitter-shadowbanning-is-real-say-inside-sources/>

Allum Bokhari, ' Facebook Admits to Shadowbanning News It Considers ‘Fake’' (BreitBart 2018) < https://www.breitbart.com/tech/2018/07/13/facebook-admits-to-shadowbanning-news-it-considers-fake/>

Alex Thompson, ' Twitter appears to have fixed “shadow ban” of prominent Republicans like the RNC chair and Trump Jr.’s spokesman' (Vice 2018) < https://www.vice.com/en/article/43paqq/twitter-is-shadow-banning-prominent-republicans-like-the-rnc-chair-and-trump-jrs-spokesman>

Liam Stack, ' What Is a ‘Shadow Ban,’ and Is Twitter Doing It to Republican Accounts?' (The New York Times 2018) < https://www.nytimes.com/2018/07/26/us/politics/twitter-shadowbanning.html>

Brett Molina, ' Shadow banning: What is it, and why is Trump talking about in on Twitter' (USA TODAY 2018) < https://www.usatoday.com/story/tech/nation-now/2018/07/26/shadow-banning-what-and-why-trump-talking-twitter/842368002/>

87. Sharing economy

See marketplace.

88. Social network

A social network is a platform-based service that enables users to share and exchange information with other individuals also using the network. Its purpose can be the sharing and exchange of political, commercial, personal interests or mixed information. Each user has a profile, which can be added within one personal network.

A social network has features enabling users to exchange information with each other publicly or semi-publicly.

“Public” means that information sharing is unrestrained.

“Semi-public” means that a user is able to disclose (private) information to a large set of users at a single time that have been authorized to be part of the users’ network.

“Purely private” exchange of information, that allow individual to share and exchange information through direct messages or mails (even in groups), and which require to know a private identifier of an individual do typically fall outside the scope of the definition of a social network. (e.g. of identifier not publicly shared: e-mail address, phone number, private pseudonym. This contrasts with a public pseudonym or an official first name and surname)

Within the social networks, the platforms that enable the publishing of content and information through multiple sources and devices are often called “social media”.

89. Spam

The concept of Spam refers to any kind of unsolicited digital communication that is usually dispatched in bulk. Hence spam is any messages sent to multiple recipients who did not ask for them.

As pointed out by the Internet Society (2017), the main problems caused by spam are due to the combination of the unsolicited and bulk aspects; the quantity of unwanted messages swamps messaging systems and drowns out the messages that recipients do want.

The first example of Spam was sent via the ARPANET in 1978. The spam was an advertisement for a new model of computer from Digital Equipment Corporation and it created a strong precedent as the communication succeeded in increasing the purchase of advertised equipment.

Some legislation, such as EU Directive 2000/31/EC on electronic commerce, utilises the concept of “unsolicited electronic communications”, which cover most sorts of 'spam' but leaves out all politically motivated unsolicited communications.

Importantly, the use of “electronic communications” is intended to cover not only traditional SMTP-based 'e-mail' but also SMS and any form of electronic messages for which the simultaneous participation of the sender and the recipient is not required.

References

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')

Internet Society. ‘What is spam?’ [2018]. Available at: https://www.internetsociety.org/wp-content/uploads/2017/08/What20Is20Spam_0.pdf

90. Technological protection measures

Due to the easy duplicability of information transmitted in digital form, copyright holders regularly resort to technical protection measures (TPMs), such as encryption-based paywalls, to prevent acts which are not authorized by the right holder of any copyright or any right related to copyright. The legal system reinforces this type of protection by explicitly outlawing not only any acts of circumvention of TPMs, but also the manufacturing and sale of devices which have the primary purpose or effect of enabling such circumvention. In the EU, in particular, Member States are required as part of the EU Copyright Directive to provide adequate legal protection against the knowing circumvention of any “effective technological measures”, thereby referring to any technology, device or component that, in the normal course of its operation, is designed to prevent or restrict acts, in respect of works or other subject-matter . According to the Directive, technological measures shall be deemed "effective" where the use of a protected work or other subject-matter is controlled by the rightholders through application of an access control or protection process, such as encryption, scrambling or other transformation of the work or other subject-matter or a copy control mechanism, which achieves the protection objective. Furthermore, adequate legal protection is required against the manufacture or sale of any device which (a) has the purpose of circumventing a TPM; (b) has a limited commercially significant purpose other than such; or (c) is primarily designed to do so.

There are two critical aspects of this type of legal protection: first, although it is intrinsically related to copyright, it is also independent from it- specifically, any unlawful circumvention constitutes a breach regardless of whether it led to an actual copyright infringement. Second, it may interfere with the ability of users of copyrighted works to benefit from exceptions and limitations that are specifically provided in copyright legislation. In principle, the continued application of these exceptions and limitations is guaranteed through article 6 (4) of the Directive, which requires Member States to provide adequate measures to that end. However, Member States are only required to act in the absence of voluntary measures taken by rightholders, including the conclusion and implementation of agreements between rightholders and other parties.

References

Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society. OJ L 167 , 22/06/2001 p. 10

U. Gasser, ‘Legal Frameworks and Technological Protection of Digital Content: Moving Forward Towards a Best Practice Model’, [2006] 17 Fordham Intellectual Property & Media Law Journal 39,60.

U. Gasser & M. Girsberger, ‘Transposing the Copyright Directive: Legal Protection of Technological Measures in EU-Member States, A Genie Stuck in a Bottle?’ Available at http://cyber.law.harvard.edu/media/files/eucd.pdf

Zingales, Nicolo, ‘Of Coffee Pods, Videogames, and Missed Interoperability: Reflections for EU Governance of the Internet of Things’ [December 1, 2015]. TILEC Discussion Paper No. 2015-026, Available at SSRN: https://ssrn.com/abstract=2707570 or http://dx.doi.org/10.2139/ssrn.2707570

91. Terms of service

Internet intermediaries, in general, and platforms, in particular, use to regulate the services they provide through standard contracts, commonly referred to as terms of service or terms of use. Terms of Service are standardized contracts, defined unilaterally and offered indiscriminately on equal terms to any user. Since users do not have the choice to negotiate, but only accept or reject these terms, Terms of Service are part of the legal category of adhesion agreements or "boilerplate contracts".

The main feature of standard contracts is that the text is not the product of a negotiation. (Prausnitz, 1937) On the contrary, the conditions are pre-determined by and expresses the one-sided control of a single party. Over the past few years, this type of contract has become the object of numerous critiques, ranging from the unilateral definition of the provisions, the almost entire absence of negotiation between the parties, and the quasi-inexistence of the bargaining power of one party that is required to adhere to the terms. (Radin, 2012; Kim, 2013; Belli & Sappa, 2017)

Internet users' mere adherence to the terms imposed by the intermediaries gives rise to a situation where consumers mechanically 'assent' to pre-established contractual regulation. Furthermore, terms of service usually foresee that the intermediaries may continue to modify the conditions of contractual agreement unilaterally. Hence, except for the possibility to "take it or leave it", users have no meaningful say about the contractual regulation they are forced to abide by. This context of "contractual authoritarianism" (Ghosh, 2014) is further exacerbated in the Internet environment. Besides having the power to unilaterally dictate the terms of use, intermediaries also enjoy the capability to unilaterally implement, through technical means, the private ordering crafted by the contractual provisions. (Belli & Venturini 2016; Belli & Zingales 2017)

As noted by the Recommendations on Terms of Service and Human Rights developed by the IGF Coalition on Platform Responsibility, the concept of "terms of service" utilised here covers not only the contractual document 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.

References

Belli L, & Venturini J. 'Private ordering and the rise of terms of service as cyber-regulation', [2016]. Internet Policy Review. Vol 5. N° 4. Available at: https://policyreview.info/articles/analysis/private-ordering-and-rise-terms-service-cyber-regulation

Belli L & Sappa C. 'The Intermediary Conundrum: Cyber-regulators, Cyber-police or both?', [2017] JIPITEC. Available at: http://www.jipitec.eu/issues/jipitec-8-3-2017/4620

Belli L & Zingales N, ‘Platform regulations: how platforms are regulated and how they regulate us’. [2017] FGV: Rio de Janeiro.

Prausnitz. O. ‘The standardization of commercial contracts in English and continental law’, [1937] Sweet & Maxwell, London.

Radin. M.J. ‘Boilerplate: The Fine Print, Vanishing Rights, and the Rule of Law’, [2012] Princeton University Press

Kim. N.S. ‘Wrap Contracts: Foundations and Ramifications’, [2013] Oxford University Press.

Ghosh S. ‘Against Contractual Authoritarianism’, [2014] Southwestern Law School Review. Vol 44.

92. Terrorist content

The concept of cyberterrorism was initially targeted at virtual attacks to infrastructure aimed at disrupting a nation or region in the same way that traditional terrorist attacks do, with the same visibility and clear intent of spreading fear. Cyberterrorism now includes the use of social media by terrorist organizations, a phenomenon that researchers have documented over a decade ago (Weimann, 2010). Terrorist content on social media is content produced and disseminated by terrorist groups or organizations with four main goals: recruitment, training, action and public terror (Goodman, 2018). This last goal is the one that features terrorist content more prominently, with the public dissemination of a terrorist narrative, one that glamorizes terrorist groups and uses social media as magnets for attention (Awan, 2017).

Much like virtual terrorist attacks need to be disentangled from activist use of computer crimes or hacktivism (Anderson, 2008), terrorist content needs to be differentiated from online activism. Cyberactivism bears some procedural similarities with the dissemination of terrorist content in social media in the sense that both involve groups that instrumentalize online tools for political purposes with wide visibility, as cyberactivism is composed of a triggering event, a media response, viral organization and a physical response (Sandoval-Almazan & Gil-Garcia, 2014). However, especially from a substantive point of view, a distinctive element of cyberterrorism and terrorist content is that they are "typically driven by an ideology with the goal of causing shock, alarm and panic." (Veerasamy, 2020).

Relevant characteristics of terrorist content in online platforms that help to identify and profile it are its specific communication flow, its influence on users, its use of propaganda and its radicalisation language (Fernandez & Alani, 2019). Social media companies employ distinct manual and automated content moderation practices to curb terrorist content (Conway et al, 2018) and therefore usually attempt to define it in their terms of service, an exercise that is useful in the development of a concept for terrorist content.

More recently, the lines between hate speech content and terrorist content have become increasingly blurred, as many of their traits bear similarities and can be described as belonging to digital hate culture, which includes not only terrorist groups formally recognized as such, but also alt-right, white supremacist and fascist groups, all belonging to "the complex swarm of users that form contingent alliances to contest contemporary political culture and inject their ideology into new spaces (...) united by 'a shared politics of negation: against liberalism, egalitarianism, 'political correctness', and the like,' more so than any consistent ideology." (Ganesh, 2018).

References

Anderson, Kent. *Hacktivism and Politically Motivated Computer Crime*, 2008. Available at: http://rageuniversity.com/PRISONESCAPE/ANTI-TERROR%20LAWS/Politically-Motivated-Computer-Crime.pdf.

Awan, Imran. Cyber-Extremism: Isis and the Power of Social Media. *Society*, v. 54, 2017.

Conway, Maura et al. Disrupting Daesh: Measuring Takedown of Online Terrorist Material and Its Impacts. *Studies in Conflict & Terrorism*, v. 42, n. 1-2, p. 141-160, 2019.

Fernandez, Miriam and Alani, Harith. Artificial Intelligence and Online Extremism: Challenges and Opportunities. In: McDaniel, John L.M. and Pease, Ken (eds). *Predictive Policing and Artificial Intelligence*. Taylor & Francis, 2019.

Ganesh, Bharath. The ungovernability of digital hate culture. *Journal of International Affairs*, Vol. 71, No. 2, 2018.

Goodman, Anka Elisabeth Jayne. When You Give a Terrorist a Twitter: Holding Social Media Companies Liable for Their Support of Terrorism. *Pepperdine Law Review*, n. 46, 2018.

Sandoval-Almazan, Rodrigo. Gil-Garcia, J. Ramon. Towards cyberactivism 2.0? Understanding the use of social media and other information technologies for political activism and social movements. *Government Information Quarterly*, Volume 31, Issue 3, pages 365-378, 2014.

Veerasamy, Namosha. Cyberterrorism - the spectre that is the convergence of the physical and virtual worlds. Emerging Cyber Threats and Cognitive Vulnerabilities. In: Benson, Vladlena and Mcalaney, John (eds). *Emerging Cyber Threats and Cognitive Vulnerabilities*, Academic Press, 2020.

Weimann, Gabriel. Terror on Facebook, Twitter, and Youtube. *Brown Journal of World Affairs*, v. 16, issue 2, spring/summer, 2010.

93. Transparency

This entry: (i) sets out the way that the term “transparency” is used in common parlance and (ii) provides examples of existing regulation which relates to transparency.

(i) Use in common parlance

To an extent, the concept of “transparency” when it comes to online platforms should be understood as falling under the umbrella of “corporate transparency” more broadly, namely the extent to which the actions of the corporation’s actions are observable by outsiders. Some elements of transparency for an online platform will therefore be comparable to other kinds of companies, particularly transparency over the company’s finances.

The use of the term “transparency” in the context of online platforms, however, refers instead to transparency over those actions which are specific to those platforms or of particular perceived importance. In practice, these primarily include actions taken in relation to the content on the platforms, as well as users’ accounts and personal data. Reflecting this, a number of online platforms publish “transparency reports” providing data and narrative descriptions on what the platform is doing. In practice, the wide variety of different kinds of platforms, and the different kinds of actions that they take, means that a more detailed definition of what constitutes transparency for a platform is challenging; this caveat notwithstanding, examples of some of the more common elements raised during discussions of transparency among online platforms include:

- Clarity over a platform’s content moderation policies and their enforcement, including the content moderation process and opportunities for challenging decisions;
- Detail on content removals, including the quantity of content removed, and the reason for its removal (e.g. for violating a company’s own terms of service or due to a government demand);
- Clarity over a platform’s data protection policies and their enforcement; and
- Detail on data shared with third parties, including governments requests for user data.

Beyond transparency reports, online platforms may seek to increase the level of transparency over their actions in other ways, such as blog posts, pop-ups for users, or libraries of advertisements hosted with details on who funded them and to whom they were targeted.

(ii) Existing regulation

There are an increasing number of examples of national regulations which require certain forms of transparency from online platforms. These include the NetzDG (Germany), which requires platforms of a certain size to publish information on the handling of complaints about unlawful content on their platforms; and the Regulation on promoting fairness and transparency for business users of online intermediation services (European Union), which requires transparency relating to commercial relationships between online platforms and business users.

94. User

The Recommendations on Terms of Service and Human Rights developed by the IGF Coalition on Platform Responsibility provide a broad definition of a Platform User as any natural or legal person entering into a contractual relationship defined by the platform's terms of service. Hence, every platform user is directly affected by the decisions and actions taken unilaterally by the platform provider.

Platform users may be taxonomized in three broad categories, including consumers, business users and providers of complementary services.

Consumers are natural persons, regardless of their nationality, who acquire or use goods or services of a kind generally acquired or used for personal, domestic or household purposes, thus not utilising or acquiring goods and services for business purposes. (UNCTAD, 2017:6)

Business users are business enterprises that utilise a platform to offer goods or services to consumers, without being required, in principle, to operate a standalone website. As pointed out by the European Commission (2018), in addition to allowing for an online presence of business users, online platforms frequently facilitate direct communications between individual business users and consumers through an embedded online communications interface. Importantly, business users of online platforms find themselves in a situation of dependence on the platform's intermediation services, thus leading to a situation in which business users often have limited possibilities to seek redress, when unilateral actions of the platform providers lead to a dispute.

There exist numerous providers of complementary products and services on all (or connected to) the big platforms. As noted by Crémer, de Montjoye & Schweitzer (2019), large platforms often invite third parties to sell their products or services on top of the original product the company already sells. This choice allows the platform to diversify its offering, and clearly has pro-competitive aspects. However, when the hosted service competes with services offered by a dominant platform itself, the rules governing the cooperation between the two may become a prime concern of antitrust enforcement

References

European Commission. Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCILon promoting fairness and transparency for business users of online intermediation services. Brussels, 26.4.2018 COM(2018) 238 final. 2018. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52018PC0238&from=en

J. Crémer, Y. de Montjoye, H. Schweitzer, 'Competition Policy for the digital era ' (European Commission Directorate-General for Competition, 2019)

UNCTAD (United Nations Conference on Trade and Development) Manual on Consumer Protection. 2017 Available at: https://unctad.org/en/PublicationsLibrary/ditccplp2017d1_en.pdf

95. User-generated content

User-generated content (UGC) is not a new term in the context of digital platforms, but rather was popularized in the earlier days of Internet intermediaries becoming popularized as mainstream spaces online, where users could congregate to connect and form communities, engage in forum discussions and other types of online dialogue, and share their writing, art, music, and other forms of work and creativity with each other and the wider world. User-generated content refers to content distinguishable as created by average everyday users online, as opposed to more professionalized content created by traditional media gatekeepers such as the legacy news, film, and music industries. Krum, Davies and Narayanaswami have described it as content that "comes from regular people who voluntarily contribute data, information, or media that then appears before others in a useful or entertaining way, usually on the Web—for example, restaurant ratings, wikis, and videos" (Krumm et al., 2008).

The term “user-generated content” appears to have declined in popular usage over time, perhaps as a result of the increasingly professionalization of such content, given the increasing accessibility, availability, and affordability of relevant tools and technologies, combined with the fact that as digital platforms have grown in reach and significance, traditional media and other industry players have themselves had to register as “**users**” on such platforms as well, as part of their business strategies. It is precisely for this reason, however, that the term UGC may still remain helpful, to distinguish content created by individual users on platforms as opposed to content created, not by traditional media gatekeepers, but neither by governmental actors or business entities that may also have accounts and engage in online activity across digital platforms.

References

John Krumm, Nigel Davies and Chandra Narayanaswami, ' User-Generated Content' [2008] IEE CS Guest Editors Introduction.

96. Utility

According to McGregor Jr. et. al (1982), the term "utility" refers to any service that is provided to people, either directly by the government (through the public sector) or indirectly (by companies). The term is associated with a social consensus (usually expressed through democratic elections) that certain services must be available to everyone, regardless of income, physical ability or intelligence.

Even when public utilities are neither publicly provided nor publicly funded, for social and political reasons they are generally subject to regulation that goes beyond what applies to most economic sectors. Public policies when done in the public interest and motivations can also provide public services (Anderfuhren-Biget et. al, 2014).

Utilities can be associated with fundamental human rights, In most countries, the term "public utilities" often includes: electricity, waste management, public transportation, health care, among many others.

A public utility can sometimes have the characteristics of a public good (being non-rival and non-excludable), but most are services that can (according to current social norms) be sub-supplied by the market. In most cases, public utilities are services, that is, they do not involve the manufacture of goods, and they can be provided by local or national monopolies, especially in sectors that are natural monopolies.

Regarding the Internet-related debates, most of them focus on the discussion on whether broadband is a public utility, something that emerged from discussions on regulation of net neutrality. Those who argue that broadband is a public utility understand that it is an essential service for the lives of citizens and, therefore, it deserves to be subject to stricter regulation, under this classification rules of net neutrality, for example, could be enforced more effectively (Mosendz, 2014). On the other hand, there are those who find it problematic to classify broadband as a public utility, since it is a service where there is a competitive market, and stronger regulations could damage innovation and create a barrier to entry (Downes, 2016).

References

Anderfuhren-Biget, Simon; Varone, Frédéric; Giauque, David (2014). «Policy Environment and Public Service Motivation». London. Public Administration. 92 (4): 807-825. doi:10.1111/padm.12026

Downes, Larry (2016). Washington Post. 'Why treating the Internet as a public utility is bad for consumers' Available at: <https://www.washingtonpost.com/news/innovations/wp/2016/07/07/why-treating-the-internet-as-a-public-utility-is-bad-for-consumers/>

McGregor Jr., Eugene B.; Campbell, Alan K.; Macy, John W.; Cleveland, Harlan (1982). «Symposium: The Public Service as Institution». Washington. Public Administration Review. 42 (4): 304-320. JSTOR i240003

Mosendz, Polly (2014). The Atlantic. 'Is Broadband Internet A Public Utility?' Available at: <https://www.theatlantic.com/technology/archive/2014/05/is-broadband-internet-a-public-utility/362093/

97. Violence

Traditionally, violence has been understood as the intentional use of physical harm against an individual or a group (Jackman 2002). However, the rise of online digital platforms has challenged existing conceptions of violence, and expanded our understanding as an act which can be perpetrated digitally (Corb 2015). Violence, in the context of virtual interactions, can be defined as the harm caused to a person or a group of people by virtue of the use of an online space, without requiring actual physical damage to the person. Described by the Council of Europe as 'cyberviolence', this can include physical, sexual, psychological, or economic harm or suffering (Working Group on cyberbullying and other forms of online violence, especially against women and children (CBG) 2018, 5).

The growing prevalence of cyberbullying, virtual sexual harassment, and online stalking make clear that the terminology related to violence in the age of online platforms is 'still developing and not univocal'(Dubravka Šimonović 2018, 5). The Council of Europe breaks down cyberviolence into six related but distinct categories (Council of Europe 2019). Firstly, cyberharassment involves the persistent and repeated targeting of a specific person in order to cause severe emotional distress or fear of physical harm. Cyberharassment, often targeting women and girls, can involve a range of online conduct including hate speech and the release of revenge pornography. Secondly, information and communication technology related violations of privacy, which seeks to obtain or misuse data or online information, is a form of violence against the individual. Thirdly, online sexual exploitation and the sexual abuse of children is increasingly prevalent due to the accessibility that online platforms provide to abusers. Fourthly, online platform-facilitated hate crimes which can be an act of individual violence, as well as lead to communal violence. For example, online platforms make more dangerous widespread incitement to violence (Avni 2018, 30–31). Fifthly, information and communication technology enables direct threats of violence as well as actually physical violence, wherein online platforms can be used in connection with murder, kidnapping, rape, extortion, and other acts of traditional violence. Finally, some cybercrime can be considered to be an act of cyberviolence. For example, the illegal access of personal data or the denial of key online services may have physical or violent repercussions. Given the rapid growth of online forms of violence, there can be a large gap in relation to 'knowledge, reporting mechanisms, support services, law enforcement, and prevention strategies to effectively tackle online abuse, online violence, and exploitation of young people' (Blaya, Kaur, and Sandhu 2018, 99).

References

Avni, Micah, ' Incitement to Terror and Freedom of Speech' in Brill | Nijhoff (eds), *Incitement to Terrorism.* (1st, e.g. Maxwell, Leiden, The Netherlands 2018). Avaliable at: https://doi.org/10.1163/9789004359826_005

Blaya, Catherine, Kirandeep Kaur, and Damanjit Sandhu. 2018. 'Cyberviolence and Cyberbullying in Europe and India - A Literature Reviewl'. In *Bullying, Cyberbullying and Student Well-Being in Schools - Comparing European, Australian and Indian Perspectives*, edited by Peter

K. Smith, Suresh Sundaram, Barbara A. Spears, Catherine Blaya, and Damanjit Sandhu, 83–106. Cambridge University Press. Available at: https://www-cambridge-org.wwwproxy1.library.unsw.edu.au/core/services/aop-cambridge-core/content/view/BD14D631A29D7A55F41E99C37B9A8D35/9781107189393c5_83-106.pdf/cyberviolence_and_cyberbullying_in_europe_and_india.pdf.

Corb, Abbee. 2015. 'Online Hate and Cyber-Bigotry: A Glance at Our Radicalized Online World.' In *The Routledge International Handbook on Hate Crime.*, 306–17. Routledge International Handbooks. New York, NY, US: Routledge/Taylor & Francis Group. Available at: https://www-routledgehandbooks-com.wwwproxy1.library.unsw.edu.au/doi/10.4324/9780203578988.ch25.

Council of Europe. 2019. 'Cyberviolence'. Cybercrime. 2019. Available at: https://www.coe.int/en/web/cybercrime/cyberviolence.

Dubravka Šimonović. 2018. 'Report of the Special Rapporteur on Violence against Women, Its Causes and Consequences on Online Violence against Women and Girls from a Human Rights Perspective'. A/HRC/38/47. Human Rights Council. Available at: https://www.ohchr.org/EN/HRBodies/HRC/RegularSessions/Session38/_layouts/15/WopiFrame.aspx?sourcedoc=/EN/HRBodies/HRC/RegularSessions/Session38/Documents/A_HRC_38_47_EN.docx&action=default&DefaultItemOpen=1.

Jackman, Mary R. 2002. 'Violence in Social Life'. *Annual Review of Sociology* 28: 387–415 Available at: https://www.annualreviews.org/doi/abs/10.1146/annurev.soc.28.110601.140936.

Working Group on cyberbullying and other forms of online violence, especially against women and children (CBG). 2018. 'Mapping Study on Cyberviolence'. Prepared by the CBG for consideration by the T-CY at its 19thPlenary. Strasbourg, France: Council of Europe. Available at: https://rm.coe.int/t-cy-2017-10-cbg-study/16808b72da.

98. User warning (of graphic content, etc.)

A user warning is a message directed to specific users by the platform operator, usually in the form of a notice or other format, aimed at alerting users that there will be an upcoming event that warrants the attention of the user. Such changing situation may be, for instance, the upcoming appearance of explicit graphic content on a news feed or the upcoming alteration of contractual terms of the platforms. User warnings are typically utilised to convey an alert and afford a user the possibility to carefully form an informed choice before proceeding to an activity that may have unwanted and potentially negative consequences.

99. Wilful Blindness

Willful blindness is a doctrine that is used by Courts to substitute for actual knowledge, especially in criminal cases**,** where a defendant could foresee the existence of wrongdoing, but deliberately avoided making an inquiry about it. Black´s law dictionary defines it as "*Deliberate avoidance of knowledge of a crime, esp. by failing to make a reasonable inquiry about suspected wrongdoing despite being aware that it is highly probable", explaining that "A person acts with willful blindness, for example, by deliberately refusing to look inside an unmarked package after being paid by a known drug dealer to deliver it. Willful blindness creates an inference of knowledge of the crime in question. See Model Penal Code § 2. [Cases: Criminal Law 20, 314. C.J.S. Criminal Law §§ 31–33, 35–39, 700; Negligence § 913.]"*

The term has been used repeatedly in the context of copyright, to identify exceptions to the safe harbor in particular for hosting intermediaries. Its role is strongly related to the concept of **red flag knowledge**, as it may enable copyright holders to circumvent the need for online service providers to be aware of specific and identifiable infringements for the purpose of establishing the requisite (constructive) knowledge. However, the Second Circuit in the Viacom case has refused to provide that basis, by essentially merging the two doctrines and requiring in both cases knowledge and awareness of specific and identifiable infringements.

The Copyright Office, in its recent Report on Section 512 of the Digital Millennium Copyright Act, has suggested to expand the reading of this concept as applied to the conditions for the safe harbor of intermediaries, going beyond the Second Circuit´s strict and bringing it in line with the interpretation of willful blindness both in other areas of copyright infringement, and outside the copyright context.

References

US Copyright Office, "Section 512 of title 17: A report of the register of copyrights" (May 2020), Available at https://www.copyright.gov/policy/section512/section-512-full-report.pdf

Viacom Int'l, Inc. v. YouTube, Inc., 676 F.3d 19 (2d Cir. 2012).