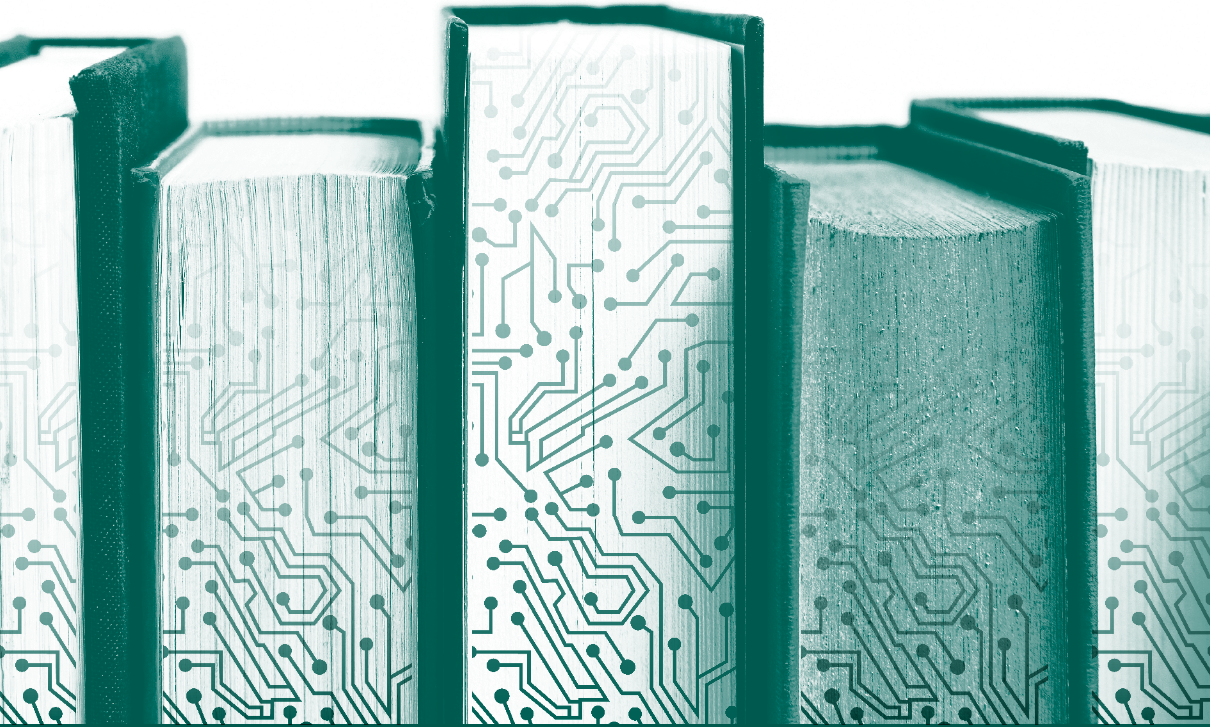
EDITORS

LUCA BELLI, NICOLO ZINGALES AND YASMIN CURZI

GLOSSARY OF PLATFORM LAW AND POLICY TERMS

PREFACE BY

PATRICK PENNINCKX



OFFICIAL OUTCOME OF THE IGF COALITION ON PLATFORM RESPONSIBILITY

IGF Internet Governance Forum

FGV DIREITO RIO

Glossary of Platform Law and Policy Terms

Luca Belli, *Nicolo Zingales* and *Yasmin Curzi*
Editors

Official Outcome of the IGF Coalition
on Platform Responsibility

Presented at the United Nations Internet Governance Forum
Katowice, Poland, December 2021

Edition produced by FGV Direito Rio
Praia de Botafogo, 190 | 13th floor
Rio de Janeiro | RJ | Brasil | Zip code: 22.250-900
55 (21) 3799-5445
www.fgv.br/direitorio

Glossary of Platform Law and Policy Terms

Luca Belli, *Nicolo Zingales* and *Yasmin Curzi*
Editors

Official Outcome of the IGF Coalition on Platform Responsibility

This Glossary is a living document and was prepared by the "Glossary Working Group" of the IGF Coalition on Platform Responsibility, a multistakeholder group under the auspices of the UN Internet Governance Forum. The views and opinions expressed herein do not necessarily reflect those of the United Nations Secretariat. The designations and terminology employed may not conform to United Nations practice and do not imply the expression of any opinion whatsoever on the part of the Organization. The members of the working group are (in alphabetical order): Luca Belli, Vittorio Bertola, Yasmin Curzi de Mendonça, Giovanni De Gregorio, Rossana Ducato, Luã Fergus Oliveira da Cruz, Catalina Goanta, Tamara Gojkovic, Terri Harel, Cynthia Khoo, Stefan Kulk, Paddy Leerssen, Laila Neves Lorenzon, Chris Marsden, Enguerrand Marique, Michael Oghia, Milica Pesic, Courtney Radsch, Roxana Radu, Konstantinos Stylianou, Rolf H. Weber, Chris Wiersma, Richard Wingfield, Monika Zalnieriute and Nicolo Zingales.

FGV Direito Rio Edition

Printed in Brazil.
1 edition finalized in December 2021.
This book is in the Legal Deposit Division of the National Library.

This material, its results and conclusions are the responsibility of the authors and do not represent, in any way, the institutional position of the Getulio Vargas Foundation / FGV Direito Rio.

Coordination: Ludmilla Totinick, Sérgio França e Nathasha Chrysthie

Organization: Luca Belli, Nicolo Zingales, Yasmin Curzi

Book cover: Andreza Moreira – Tangente Design

Layout: Andreza Moreira – Tangente Design

1st review: Luca Belli and Nicolo Zingales

2nd review: Yasmin Curzi

Cataloguing data prepared by the Mario Henrique Simonsen/FGV Library

Glossary of platform law and policy terms / Luca Belli, Nicolo Zingales and
Yasmin Curzi Editors – Rio de Janeiro: FGV Direito Rio, 2021.
324 p.

Official Outcome of the IGF Coalition on Platform Responsibility.
Presented at the United Nations Internet Governance Forum Katowice,
Poland, December 2021.

Em colaboração com: Vittorio Bertola, Giovanni De Gregorio, Rossana Ducato,
Luã Fergus Oliveira da Cruz, Catalina Goanta, Tamara Gojkovic, Terri Harel,
Cynthia Khoo, Stefan Kulk, Paddy Leerssen, Laila Neves Lorenzon,
Chris Marsden, Enguerrand Marique, Michael Oghia, Milica Pesic, Courtney
Radsch, Roxana Radu, Konstantinos Stylianou, Rolf H. Weber, Chris Wiersma,
Richard Wingfield, Monika Zalnieriute and Nicolo Zingales.

ISBN: 978-65-86060-31-7

1. Tecnologia e direito. 2. Plataforma aberta da Web - Legislação. 3. Governança
da Internet - Legislação. I. Belli, Luca, 1985-. II. Zingales, Nicolo. III. Curzi, Yasmin.
IV. Fundação Getulio Vargas.

CDD – 344.09

Elaborada por Rafaela Ramos de Moraes – CRB-7/6625

CONTENTS

INTRODUCTION 11

PREFACE 19

ABOUT THE AUTHORS 23

1 **(Digital) Access** 31
Chris Wiersma

2 **Accountability** 35
Nicolo Zingales

3 **Aggregator** 37
Clara Almeida

4 **Amplification** 39
Paddy Leerssen

5 **Appeal** 41
Luca Belli

6 **Application Program Interface** 45
Nicolo Zingales

7 **Arbitration** 47
Luã Fergus and Laila Lorenzon

8 **Automated Decision Making** 51
Rossana Ducato

9 **Bot** 59
Luã Fergus and Laila Lorenzon

10 **Child Pornography/Child Sexual Abuse Material** 63
Richard Wingfield

11 **Common Carrier** 65
Chris Marsden

12 **Content Creator/Influencer** 69
Catalina Goanta/Giovanni de Gregorio

13 **Content** 73
Richard Wingfield

14 **Content/Web Monetization** 75
Catalina Goanta

15 **Coordinated Flagging** 79
Cynthia Khoo

16 **Coordinated Inauthentic Behavior** 81
Cynthia Khoo

17 **Co-regulation** 83
Rolf H. Weber

18 **Content Curation** 85
Paddy Leerssen

19 **Dark Patterns** 87
Nicolo Zingales

20 **Data Portability** 93
Vittorio Bertola, Nicolo Zingales and Luca Belli

21 **Defamation** 101
Richard Wingfield

22 **Deindexing** 105
Courtney Radsch

23 **Demonetization** 107
Nicolo Zingales and Catalina Goanta

24 **Deplatforming/De-platforming** 109
Courtney Radsch

25 **Device Neutrality** 113
Luã Fergus and Laila Lorenzon

26 **Digital Rights** 117
Chris Wiersma

27 **Disinformation** 119
Giovanni De Gregorio

28 **Dispute Resolution [Online]** 123
Catalina Goanta

29 **Due Diligence** 125
Luca Belli

30 **Duty of Care** 127
Chris Wiersma

31 **End to End Encryption** 129
Catalina Goanta, Vittorio Bertola

32 **Fact-checking** 131
Luca Belli

33 **Federated Service** 133
Yasmin Curzi

34 **Filter** 135
Vittorio Bertola

35 **Flagging** 141
Cynthia Khoo

36 **(Digital) Gatekeeper** 143
Nicolo Zingales

37 **Disinformation (Gendered)** 147
Yasmin Curzi

38 **Governance** 149
Paddy Leerssen

39 **Harassment** 153
Terri Harel and Yasmin Curzi

40 **Harm (Online Harm)** 155
Yasmin Curzi and Cynthia Khoo

41 **Hash/Hash Database** 157
Courtney Radsch

42 **Hate Speech** 159
Yasmin Curzi

43 **Human Exploitation** 161
Luã Fergus and Laila Lorenzon

44 **Human Review** 163
Cynthia Khoo

45 **Incitement to Violence** 165
Monika Zalnieriute

46 **Inclusive Journalism** 169
Milica Pesic and Tamara Gojkovic

47 **Information Fiduciary** 173
Nicolo Zingales

48 **Infrastructure** 175
Luã Fergus and Laila Lorenzon

49 **Intermediary Liability** 177
Luã Fergus and Laila Lorenzon

50 **Internet Safety/Security** 179
Chris Wiersma

51 **Interoperability** 183
Vittorio Bertola and Nicolo Zingales

52 **Liability** 185
Nicolo Zingales

53 **Marketplace** 189
Enguerrand Marique

54 **Media Pluralism** 191
Michael J. Oghia

55 **Microtargeting** 195
Paddy Leerssen

56 **Moderation** 199
Giovanni de Gregorio

57 **Must-carry** 205
Luã Fergus and Laila Lorenzon

58 **Non-discrimination** 207
Richard Wingfield

59 **Notice** 209
Luã Fergus and Laila Lorenzon

60 Notice-and-notice 217
Nicolo Zingales

61 Notice-and-takedown 219
Luã Fergus and Laila Lorenzon

62 Notice-and-staydown 221
Nicolo Zingales

63 Nudging 225
Nicolo Zingales

64 Online Advertising 231
Catalina Goanta

65 Open Identity 233
Vittorio Bertola

66 Open Standard 235
Vittorio Bertola

67 Optimization 237
Courtney Radsch

68 Platform 239
Luca Belli

69 Platform Governance 243
Paddy Leerssen

70 Platform Neutrality 247
Konstantinos Stylianou

71 Pornography 249
Yasmin Curzi

72 Prioritization 251
Courtney Radsch

73 Proactive Measures 253
Daphne Keller and Nicolo Zingales

74 Recommender Systems 257
Rossana Ducato

75 Red Flag Knowledge 263
Nicolo Zingales

76 Regulation 265
Roxana Radu

77 Remedy 269
Chris Wiersma

78 Repeat Infringer 271
Luca Belli

79 Responsibility 273
Nicolo Zingales

80 Revenge Pornography/Non-consensual Intimate Images 277
Richard Wingfield

81 **Right to Explanation** 279
Nicolo Zingales

82 **Right To Be Forgotten** 283
Stefan Kulk

83 **Safe Harbor** 285
Nicolo Zingales

84 **Self-injury** 287
Luã Fergus and Laila Lorenzon

85 **Self-preferencing** 289
Konstantinos Stylianou

86 **Self-regulation** 291
Rolf H. Weber

87 **Shadowban/Shadow Banning** 295
Courtney Radsch

88 **Sharing Economy** 297
Enguerrand Marique and Yasmin Curzi

89 **Social Network** 299
Enguerrand Marique

90 **Spam** 301
Luca Belli301

91 **Technological Protection Measures** 303
Nicolo Zingales

92 **Terms of Service** 305
Luca Belli

93 **Terrorist Content/Cyberterrorism** 309
Ivar Hartmann

94 **Transparency** 311
Richard Wingfield

95 **User** 313
Luca Belli

96 **User-generated Content** 315
Cynthia Khoo

97 **Utility** 317
Luã Fergus and Laila Lorenzon

98 **Violence (Cyber)** 319
Monika Zaineriute

99 **User Warning** 321
Luca Belli

100**Wilful Blindness** 323
Nicolo Zingales

INTRODUCTION

A Glossary of Platform Law and Policy Terms to Foster Legal Interoperability

Luca Belli and Nicolo Zingales

At the 2019 United Nations 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 policymakers when discussing issues of platform responsibility.

In this perspective, the elaboration of this Glossary aims at providing the conceptual basis on which legal interoperability between different systems framing platform governance can be built. Through this Glossary, we aim at offering guidance on what specific platform-related concepts mean, so that different stakeholders and, particularly, policymakers may have a better understanding of such a complex set of issues, thus elaborating well-informed and, ideally, good-quality and compatible platform policies and regulations.

Importantly, this Glossary does not aim at being prescriptive, but rather at recognizing that a specific concept may have various meanings and such differences and nuances should be acknowledged and highlighted to allow stakeholders to have a more complete understanding of each issue. Indeed, the consideration of heterogeneous conceptualizations, interpretations, and approaches adopted by different (juridical) cultures holds the promise to enrich the way platform-related issues are framed by stakeholders in different countries. At the same time, this can foster the elaboration of shared or, at least, converging principles, rules and procedures by national regulators, legislators or even market players, thus facilitating

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

legal interoperability.[2] Indeed, common conceptualizations may not only inspire legislative efforts but also be used as basis on which develop cooperative and converging frameworks by national public bodies and intergovernmental organizations, while elaborating public policies, or even assist private-sector actors in the elaboration of self-regulatory instruments.

As we are fully aware of the evolving nature of many of concepts analyzed in this Glossary, we agreed that the Glossary should be considered as a "living document" that could be updated over time. As such the Glossary aims at bringing together contributions from a heterogeneous range of disciplines, stakeholder perspectives and vocabularies. Coalition stakeholders also agreed that the definitional efforts should recognize as much as possible the existence of competing and alternative views on the topic, and the Glossary contributors did their best effort to reflect such conceptual diversity in this volume.

Glossary contributors were encouraged to conceive definitions as a springboard for learning more about concepts and views, through links and references to external sources. Further reference and links to external sources will be added in the new versions of the Glossary that will be uploaded on the IGF website in the coming years, after having received and incorporated any comments arising in the platform related discussions the Coalition will organize within the IGF.

How have we organized this participatory effort?

The IGF Coalition on Platform Responsibility, as any other IGF 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. After the first round of suggestions took place and several Coalition members manifested interested in the Glossary project, the following action plan was shared for feedback, and subsequently implemented between May and October 2020:

[2] For an analysis of the concept of legal interoperability, see Weber (2014). For a discussion of how this concept can be applied to foster compatible net neutrality and data protection frameworks, see Belli & Foditsch (2016) and Belli (2020). An noteworthy approach to the concept of legal interoperability is offered by the works of Internet and Jurisdiction project. See <https://www.internetjurisdiction.net/>.

- reception of expressions of interest for the development of the Glossary and participation to the Coalition session;
- consolidation of the proposed terms and circulation of a draft list of terms to be used to compose the Glossary;
- reception of feedback on the draft list and suggestion of further terms;
- development 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;[3]
- elaboration of draft entries describing the proposed terms;
- consolidation of the draft entries into a first draft version of the glossary and request for comment on the first draft;
- consolidation of the updated version into a consolidated draft to be circulated at the IGF 2020 for further feedback from the IGF community;
- discussion of the draft at the 2020 session of the Coalition, during the IGF, and elaboration of a strategic approach aimed at maximizing the impact of the Glossary;
- a final consultation phase was organized using the IGF website as a platform for comments, between November 2020 and January 2021;
- consolidation and revision with glossary contributors and Coalition stakeholders took place until November 2021.

While this may not be the first attempt to create a glossary of platform-related terms,[4] 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

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

[4] See, for example, the Stanford Glossary, available at: <http://cyberlaw.stanford.edu/blog/2018/01/glossary-internet-content-blocking-tools>.

meaningful and inclusive discussion and cooperation among academics, policymakers, journalists, platform users and any other stakeholder with a keen interest in platform governance. To be continued!

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 stakeholders, is to have coined and promoted the concept of "Platform Responsibility".[5] 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[6] 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 behavior.

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,[7] in particular their responsibility to respect Human Rights and to grant effective grievance mechanisms.[8] 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.[9]

[5] See Belli, L., De Filippi, P., Zingales, N. (2014).

[6] See the 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. Available at: <https://www.ohchr.org/Documents/Publications/GuidingPrinciplesBusinessHR_EN.pdf>.

[7] Idem.

[8] See Belli, L., De Filippi, P., Zingales, N. (2015). Recommendations on terms of service & human rights, Outcome Document n°1. Available at: <https://tinyurl.com/toshr2015>.

[9] See Zingales N. and Belli L. (2014). Report of the "inception" meeting at the 2014 IGF. Available at: <https://www.intgovforum.org/multilingual/index.php?q=filedepot_download/4905/631>.

The resulting "Recommendations on Terms of Service and Human Rights"[10] (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".[11]

Indeed, the Recommendations served as an inspiration for (and were annexed to) both the study on Terms of Service and Human Rights,[12] 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.[13] 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.[14] 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 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

[10] See Belli L., De Filippi P. and Zingales N. (2015).

[11] See Tunis Agenda (2005) available at: <https://www.itu.int/net/wsis/docs2/tunis/off/6rev1.html>.

[12] See Venturini et al. (2016).

[13] See Belli and Zingales (2017).

[14] See <http://bit.ly/CoEinternetintermediaries>.

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.[15] 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".[16] This volume aimed at promoting a discussion on the multiform notion of platform value(s) and the term "value" was 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,[17] 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

Belli, L. (2020). Data Protection in the BRICS Countries: Enhanced Cooperation and Convergence towards Legal Interoperability. New Media Journal. Chinese Academy of Cyberspace Studies. <https://cyberbrics.info/data-protection-in-the-brics-countries-enhanced-cooperation-and-convergence-towards-legal-interoperability/>.

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

[16] A pre-print version of the Special Issue can be accessed at <https://www.intgovforum.org/multilingual/index.php?q=filedepot_download/4905/1900>.

[17] See, as an instance, Crémer, de Montjoye and Schweitzer (2019); Eyler-Driscoll, Schechter and Patiño (2019); BRICS Competition Law and Policy Centre (2019).

Belli, L., Zingales, N. (Eds) (2017). *Platform regulations: how platforms are regulated and how they regulate us.* Leeds. Available at: <https://bibliotecadigital.fgv.br/dspace/handle/10438/19402>.

Belli, L., Foditsch, N. (2016). Network Neutrality: An Empirical Approach to Legal Interoperability. In Belli, L., De Filippi P. (Eds.) Net Neutrality Compendium. Human Rights, Free Competition and the Future of the Internet. Springer.

Belli, L. B., De Filippi, P., Zingales, N. (2014). A New Dynamic Coalition on Platform Responsibility within the IGF. *Medialaws*. Available at: <http://www.medialaws.eu/a-new-dynamic-coalition-on-platform-responsibility-within-the-igf>.

Belli, L. B., De Filippi, P., Zingales, N. (2014). Recommendations on terms of service & human rights. Outcome Document n°1. *Internet Governance Forum*. Available at: <https://www.intgovforum.org/cms/documents/igf-meeting/igf-2016/830-dcpr-2015-output- document-1/file/>.

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

Crémer, J. de Montjoye, YA. Schweitzer, H. (2019). *Competition Policy for the digital era. European Commission Directorate-General for Competition*. Available at: <https://ec.europa.eu/competition/publications/reports/kd0419345enn.pdf>.

Eyler-Driscoll S., Schechter A. and Patiño C. (2019). Digital Platforms and Concentration. ProMarket and Chicago Booth Stigler Center.

Ruggie, J. (2011). Report of the special representative of the secretary-general on the issue of human rights and transnational corporations and other business enterprises: Guiding principles on business and human rights: implementing the united nations 'protect, respect and remedy'framework. *Netherlands Quarterly of Human Rights*, *29*(2), 224-253. Available at: <https://www.ohchr.org/Documents/Publications/GuidingPrinciplesBusinessHR_EN.pdf>.

Tunis Agenda for the Information Society. (2005). *WSIS-05/TUNIS/DOC/6(Rev. 1)-E*. Available at: <https://www.itu.int/net/wsis/docs2/tunis/off/6rev1.html>.

Venturini, J., Louzada, L., Maciel, M., Zingales, N., Stylianou, K., Belli, L. (2016). Terms of service and human rights: An analysis of online platform contracts. FGV Direito Rio – Revan Editora – Council of Europe. <http://bibliotecadigital.fgv.br/dspace/handle/10438/18231>.

Zingales, N. Belli, L. (2014). Dynamic Coalition on Platform Responsibility: Report of the "inception" meeting at the 2014 IGF. *Internet Governance Forum.* Available at: <https://www.intgovforum.org/multilingual/index.php?q=filedepot_download/4905/631>.

Weber R. (December 2014). Legal Interoperability as a Tool for Combatting Fragmentation, Global Commission on Internet Governance, Paper Series n°4. <https://www.cigionline.org/sites/default/files/gcig_paper_no4.pdf>.

PREFACE

Framing the Roles and Responsibilities of Platforms and their Impact on Human Rights

Patrick Penninckx

We all strive for a common understanding of the Human Rights ideals and the need of protecting our common values of Human Rights, Democracy and the Rule of Law as the basis of our societies. The international exiting milestones, like the Universal Declaration of Human Rights, the **International Covenant on Civil and Political Rights** or the European Convention of Human Rights point in the same direction when it comes to protecting these common values in changing times and new challenges.

In an era, where the world is evolving faster than ever, we should hold on to those anchors. And just as we speak and have a common understanding about the values we need to protect, it is of extreme importance to use the same approach when exploring our fast-evolving modern lives and the turmoil it brings about. Beacons are needed.

With this intention, this Glossary puts together such beacons for the digital sector.

The development and deployment of digital platforms over the past decades has taken on a whole new dimension. They have become an important part of people's everyday information and communication activities and more generally, of various aspects of our lives. They have also transformed individuals' media and news consumption habits, offering new opportunities in terms of access to information, freedom of expression, public debate, and democratic participation. In short, the ability to receive and impart information and ideas.

Platforms have not only become powerful intermediaries between content producers and their audiences but have also assumed a central position in the world economy, at a time when the whole public sphere has gone through a profound structural transformation. New powerful actors, including internet intermediaries, became strongholds in their own right.

Digital technologies, their platforms and search engines offer enhanced opportunities for expression, access to information and communication. At the same time, they are not neutral. Some assume an active curatorial or editorial role in the dissemination of online content. They exercise editorial functions through content moderation and prioritisation, which mostly involve automated tools using artificial intelligence, complemented by human moderators.

The content dissemination itself is facilitated and oriented through access to personal data, while algorithms are able to select privileged sources and types of information and provide them to the audience in a highly personalised and targeted manner.

This allowed intermediaries to establish themselves rapidly as dominant players on the information and communication market and to have a strong influence on shaping of the public opinion. This also implies that the interests of the platforms' owners – predominantly financial, but potentially also political – can have significant societal ramifications.

Finally, as always, there are two sides of the coin and the use made of the tools we have at our disposal is a choice.

In fact, most of platforms' content moderation is aimed at restricting access to illegal content (hate speech, online terrorist propaganda, child sexual abuse material, etc.). The platforms' curatorial and editorial role also extends to dealing with legal but contentious or harmful content which can range from disinformation (which may also originate from official sources,) to obscene, vulgar or other offensive material.

Beyond the opportunities they offer, platforms pose new challenges not only for the realisation of human rights and fundamental freedoms, including the freedom of expression and information, the right to private life and the protection of personal data, but also for the functioning of democratic societies.

This has many state actors, as well as the international community, to bring forward proposals for rethinking and reforming their freedom of expression, media and communication frameworks. There is a need of rebalancing the relationship between the states, individuals, online platforms, the traditional media and other information and communication channels.

The changes in these regulatory and liability frameworks carry a risk of excessive control and a systemic chilling effect on freedom of expression and information. Examples include wholesale blocking of internet websites and vaguely framed restrictions allowing for abuse. This has become particularly apparent during the recent global pandemic.

Thus these changes require preliminary impact assessments and an inclusive and participatory approach. In today's complex media and communication ecosystem it is a shared responsibility of both public and private actors to contribute to an open and free environment for expression, information, communication and public participation, involving free speech, free access to/exchange of quality and pluralistic information, and free, yet responsible and balanced public discourse.

The Council of Europe engaged in a reflection on related roles and responsibilities of platforms, states and other stakeholders and has developed detailed and nuanced guidance in this field from the early days.

In particular, the Committee of Ministers' *Recommendation on the roles and responsibilities of internet intermediaries* sets out obligations for States and specific responsibilities for intermediaries. The objective is to properly integrate human rights and the rule of law into the governance of platforms. States are to formulate regulations to effectively safeguard human rights of users vis-à-vis intermediaries. Intermediaries, in turn, have the responsibility to conform to the international human rights standards. This includes the requirements of transparency of their content moderation, independent oversight and effective remedies for potential violations.

A recent Guidance note for States and other actors in the area of content moderation by online platforms elaborates further on the responsibilities of internet intermediaries and way for states to provide the most appropriate frameworks to ensure due respect of human rights and rule of law in this area.

More specific guidance is under preparation on election communication and media coverage of electoral campaigns. Guiding principles for media and communication governance are intended to help states address the shift from established channels to social networks and related risks (manipulation of public opinion, lack of public trust, information disorder). Moreover, a future Recommendation on addressing hate speech, including online expressions of hate, is soon to be finalised.

Such policies need to be developed in a constructive, open and inclusive multi-stakeholder dialogue to find effective and sustainable solutions. The Council of Europe has been cooperating closely with civil society for decades. More recently, it also established a partnership with leading technology firms and their associations, enabling their representatives to sit side-by-side with governments and civil society when shaping policies related to digital technologies, in the perspective of the respect of human rights and supporting democracy and the rule of law.

Ultimately it is for the internet users themselves to make this equation to work. Therefore, the Council of Europe is systematically integrating the Media and Information Literacy (MIL) perspective in its standard setting instruments, with a view to empowering individuals and equipping them with the critical skills and knowledge necessary to navigate and make autonomous choices in the complex digital sphere.

A multi-stakeholder approach is in this complex and volatile setting the most effective way forward to safeguard and promote human rights, rule of law and democracy both offline and online.

In conclusion, the Platform Responsibility concept and approach as presented in the Glossary of Platform Law and Terms of Service finds reflects largely the key issues identified in the Council of Europe’s work on the roles and responsibilities of platforms and their impact on human rights, and hence becomes a valuable instrument in the continuous efforts to promote our common values in the digital era.

Patrick Penninckx,
Head of the Information Society Department, Council of Europe

ABOUT THE AUTHORS

Luca Belli

Luca Belli, PhD, is Professor of Internet Governance and Regulation at Fundação Getulio Vargas (FGV) Law School, where he heads the Center for Technology and Society (CTS-FGV) and the CyberBRICS project, and associated researcher at Centre de Droit Public Comparé of Paris 2 University. He is co-founder and co-coordinator of the IGF Coalition on Platform Responsibility and Director of the Latin-American edition of the Computers Privacy and Data Protection conference (CPDP LatAm). Before joining FGV, Luca worked as an agent for the Council of Europe Internet Governance Unit and served as a Network Neutrality Expert for the Council of Europe. He is author of more than 50 academic publications which have been quoted by numerous media outlets, including The Economist, Financial Times, Forbes, Le Monde, BBC, The Hill, China Today, O Globo, Folha de São Paulo, El Pais, and La Stampa. Luca holds a PhD in Public Law from Université Panthéon-Assas, Paris 2.

Nicolo Zingales

Nicolo Zingales is Professor of Information Law and Regulation at the law school of the Fundação Getulio Vargas in Rio de Janeiro, and coordinator of its E-commerce research group. He is also an affiliated researcher at the Stanford Center for Internet and Society, the Tilburg Law & Economics Center and the Tilburg Institute for Law and Technology, co-founder and co-chair of the Internet Governance Forum's Dynamic Coalition on Platform Responsibility.

Yasmin Curzi de Mendonça

Researcher at the Center for Technology and Society at the FGV Law School. PhD Candidate at the Rio de Janeiro State University, with a CAPES grant. She holds a Master's Degree in Social Sciences from PUC-Rio, also with a CAPES grant. Yasmin holds Bachelor's Degrees in both Law and Social Sciences from FGV-Rio, with an exchange period at the Université Sorbonne (Paris-IV). She is a former assistant researcher at the Center for Law and Economics from the FGV Law School, and former researcher at the Directory for Analysis of Public Policy. As an attorney, Yasmin also has experience with legal counseling, having worked with the NGO Soul Sisters (Brazil, São Paulo), and with the NGO Stop Street Harassment (Washington-DC).

Clara Almeida

Research Assistant at the Center for Technology and Society (CTS-FGV). Master of Laws candidate in Regulatory Law at FGV Direito Rio. Bachelor of Laws from FGV Direito Rio (2019).

Vittorio Bertola

Vittorio Bertola is the Head of Policy & Innovation at Open-Xchange, a leading provider of open source email and DNS solutions, where he develops and promotes new technical standards and advocates an open Internet based on user choice, privacy and federation. In the last twenty years he was involved with several Internet startups, including the early pan-European digital music platform Vitaminic, and he served in many positions in national and international Internet governance organizations, including as ICANN Board liaison and Chairman of the At-Large Advisory Committee, and as a member of the United Nations' Working Group on Internet Governance.

Luã Fergus Cruz

Luã Fergus Cruz is a researcher at the Brazilian Institute for Consumer Protection (Instituto Brasileiro de Defesa do Consumidor, IDEC). He's also a postgraduate student in Science Communications at State University of Campinas (Unicamp).

Catalina Goanta

Assistant Professor in Private Law at the Faculty of Law. During February 2018 – February 2019, I was a Niels Stensen fellow and visited the University of St. Gallen (The Institute of Work and Employment) and Harvard University (The Berkman Center for Internet and Society). She is also a non-residential fellow of the Stanford Transatlantic Technology Law Forum.

Tamara Gojkovic

Tamara Gojkovic is the Secretary General of the Youth for Exchange and Understanding (YEU) in Belgium and also the Vice President of the Life Long Platform. She holds a BA in Media and Communications and a Master's Degree in Japanese Language and Literature from the University of Belgrade. From 2020 to 2021 she worked as Head of Operations at the Media Diversity Institute. She has broad experience in project management (Prince2 Practitioner certified), fundraising, campaigning, organisational development and management, strategic planning, policy development and advocacy.

Giovanni de Gregorio

Giovanni De Gregorio is postdoctoral researcher working with the Programme in Comparative Media Law and Policy at the Centre for Socio-Legal Studies at the University of Oxford. His research focuses on digital constitutionalism, platform governance and digital policy.

Terri Harrel

Terri Harel is the Executive Director at OnlineSOS.org. At OnlineSOS, she led content development and research for an in-depth report about the current state of online harassment and its effects on media, journalists and civil society. Before joining OnlineSOS, Terri consulted for nonprofit and tech organizations and, before that, led product marketing at Classy.org, a fundraising platform for nonprofits. She's a graduate of the University of California, Berkeley with a degree in Political Economy.

Ivar Hartmann

Ivar Hartmann is an Associate Professor at Insper Learning Institution in São Paulo, Brazil. His research and teaching areas comprise cyberlaw, legal data science and constitutional law. He was previously an Assistant (2013-2018) and Associate (2018-2020) Professor at FGV Law School in Rio de Janeiro, where he coordinated the Center for Technology and Society (CTS-FGV) and the Legal Data Science Nucleus. Ivar holds an MsC from the Catholic University of Rio Grande do Sul, Brazil, an LL.M. from Harvard Law School, and an S.J.D. from the Rio de Janeiro State University.

Michael J. Oghia

Michael J. Oghia is a Belgrade-based consultant, editor, researcher, speaker, and ICT sustainability advocate working within the digital policy & infrastructure, Internet governance, and media development ecosystems. He is a third-culture kid (TCK) and a connector at heart with more than a decade of professional experience in conflict resolution, journalism & media, policy, and development across five countries: The United States, Lebanon, India, Turkey, and Serbia.

Daphne Keller

Daphne Keller directs the Program on Platform Regulation at Stanford's Cyber Policy Center, and was formerly the Director of Intermediary Liability at CIS. Her work focuses on platform regulation and Internet users' rights. She has published both academically and in popular press; testified and participated in legislative processes; and taught and lectured extensively. Her recent work focuses on legal protections for users' free expression rights when state and private power intersect, particularly through platforms' enforcement of Terms of Service or use of algorithmic ranking and recommendations. Until 2015 Daphne was Associate General Counsel for Google, where she had primary responsibility for the company's search products. She worked on groundbreaking Intermediary Liability litigation and legislation around the world and counseled both overall product development and individual content takedown decisions.

Cynthia Khoo

Cynthia Khoo is an Associate at the Center on Privacy and Technology at Georgetown Law, where she leads on worker surveillance and the civil rights implications of commercial data practices, including algorithmic discrimination. She is a Canadian technology and human rights lawyer who joined the Center after accumulating years of experience in technology law, policy, research, and advocacy with various digital rights NGOs and through her sole practice law firm. Cynthia is also a fellow at the Citizen Lab (University of Toronto). She holds a J.D. from the University of Victoria and an LL.M. from the University of Ottawa.

Stefan Kulk

Stefan Kulk is an assistant-professor at Utrecht University. His research focuses on the role and influence of online services providers in our information societies. He is specialized in online tort law, privacy, and intellectual property. Stefan wrote his PhD-thesis on the liability for illegal content of online service providers, such as internet access providers, search engines, and platforms. During his time as a PhD student, he spent three months at the Berkman Klein Center for Internet and Society of Harvard University. Stefan also wrote several pieces on the right to be forgotten, including a chapter in the Cambridge Handbook of Consumer Privacy (together with Frederik Zuiderveen Borgesius).

Paddy Leersen

Paddy Leerssen is a PhD Candidate in information law at the University of Amsterdam. His research focuses on the regulation and governance of social media platforms, with a particular focus on transparency and data access.

Laila Lorenzon

Laila Lorenzon works at the Data-Pop Alliance and is an International Relations student at Federal University of Rio de Janeiro, Brazil. She is currently a member of the Brazilian Chapter of Internet Society and helps the events and communications team by promoting ISOC Brazil projects and activities. Her main interests are related to gender studies, education and digital citizenship, privacy policies and data protection, Internet Governance. She has worked as a research intern for the Digital Rights Foundations, as assistant researcher at the CyberBRICS Project, and carried out educational projects, being an ambassador of the Digital Citizen Program, created by the NGO Safernet and Facebook.

Enguerrand Marique

Dr. Enguerrand Marique is an Assistant Professor in Conflict Solving Institutions and Digital conflict resolution at Radboud University Nijmegen (The Netherlands) and a guest lecturer at the UCLouvain (Belgium) and Université Saint-Louis, Brussels (Belgium). His current research interests address conflict resolution between users and platforms, EU harmonization policies and digital governance.

Chris Marsden

@ChrisTMarsden is Professor of Internet Law at the University of Sussex and an expert on Internet and new media law, having researched and taught in the field since 1995. Chris researches regulation by code – whether legal, software or social code. He is author of five monographs on Internet law including “Net neutrality: From Policy to Law to Regulation” (2017), “Regulating Code” (2013 with Prof. Ian Brown), “Internet Co-regulation” (2011). He is author of many refereed articles, book chapters, professional articles, keynote addresses, and other scholarly contributions. His current funded research is into Trusted Autonomous Systems (UKRI-EPSRC 2020-24).

Milica Pesic

Milica Pesic is the Executive Director of the Media Diversity Institute (MDI). She has been working in the diversity and media field for more than 20 years, designing and supervising multi-national, multi-annual programmes in Europe, NIS, MENA, South Asia, the Sahel, Sub-Sahara, West Africa, China and Cuba. She has co-designed an MA Course in Diversity and the Media which is jointly run by the MDI and University of Westminster. A journalist by profession, she has reported for the BBC, Radio Free Europe, the Times HES, TV Serbia and other media. She holds an MA in International Journalism from City University, London. Prior to MDI, she had worked for New York University, the IFJ (Brussels) and the Alternative Information Network (Paris). Her current interests are in MIL and CVE. MDI has branches in the US, the Western Balkans, Belgium and the South Caucasus.

Courtney Radsch

Courtney Radsch is an American Journalist. She holds a Ph.D. in international relations and is author of Cyberactivism and Citizen Journalism in Egypt: Digital Dissidence and Political Change. She has also worked as the advocacy director for the Committee to Protect Journalists until 2021.

Roxana Radu

Roxana Radu is an Internet governance and digital policy expert. She is a Research Associate at the Global Governance Centre, Graduate Institute (Geneva) and a non-residential fellow at the Centre for Media, Data and Society, Central European University (Vienna). She is the winner of the 2017 Swiss Network for International Studies Award for her PhD (summa cum laude), obtained at the Graduate Institute. Her interdisciplinary research and publications focus on international governance and Internet policy-making, ranging from the politics of technical standards to artificial intelligence.

Konstantinos Stylianou

Konstantinos Stylianou is an Associate Professor in Competition Law and Regulation at the University of Leeds. His areas of focus are the law and policy of digital markets, antitrust, and blockchain. He has worked on projects funded by the EU, Swedish Competition Authority, Google, Facebook, and Thailand's National Broadcasting and Telecommunications Commission among others, and he has been involved with the Hellenic Competition Commission and the Greek Government on the reform of competition law in Greece. He holds an S.J.D. from the University of Pennsylvania, where he studied as an Onassis Scholar, an LL.M. from Harvard University, where he studied as a Fulbright Scholar, and an LL.M. and LL.B. from Aristotle University.

Rolf H. Weber

Prof. Dr. Rolf H. Weber is Professor of international business law at Zurich University acting there as co-director of the Research Program on Financial Market Regulation, the Center for Information Technology, Society, and Law and the Blockchain Center. Furthermore, he was Visiting Professor at Hong Kong University and he is practicing attorney-at-law in Zurich. Prof. Weber is member of the Editorial Board of several Swiss and international legal periodicals and frequently publishes on issues of global law. His main fields of research and practice are IT- and Internet, international trade and finance as well as competition law.

Chris Wiersma

Chris Wiersma is an independent researcher and senior adviser-jurist. Chris specializes in Information Law, especially having expertise in media law, IP/copyright, data protection and the impact of digital technologies on human rights, in the context of European legislation. Previously, Chris held positions as scientific staff at the Universities of Amsterdam and Ghent, where he taught classes in law and political/social sciences. His work has been published in Dutch and English in journals such as *Mediaforum* (deLex), *Auteurs en Media* (Larcier), *Nordic Journal of Human Rights* (Taylor&Francis/Routledge) and *Communication Law and Policy* (idem). Recent research on ORCiD at 0000-0002-5137-6046.

Richard Wingfield

Richard is Head of Legal at Global Partners Digital, an international human rights organisation working to enable a digital environment underpinned by human rights. Richard oversees the organisation's legal, policy and research function, building its understanding of the application of international law to internet and digital policy, developing its policy positions, and monitoring trends and developments across the world. Richard also oversees the organisation's engagement in key legislative and legal processes at the national, regional and global levels, as well as its engagement with the tech sector.

Monika Zalnieriute

Dr. Monika Zalnieriute is a Senior Lecturer and Australian Research Council DECRA Fellow at UNSW Sydney. She holds a PhD in Law from European University Institute in Florence, Italy. Previously, Monika led a research stream on 'Technologies and Rule of Law' at the Allens Hub for Technology, Law and Innovation at UNSW Law Sydney, and held a Postdoctoral Fellowship at the University of Melbourne, where she worked on the digital rights and discrimination of marginalized groups online.

1 (Digital) Access

Chris Wiersma

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 organization.

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) or even the development of network infrastructure itself, also known as 'network self-determination' (Belli, 2017). These factors should be taken up together for describing 'meaningful connectivity' (A4AI, 2020).

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 emphasized 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, 2020:51).

The lack of relevant skills also limits awareness of potential benefits from digitization. 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 "Digital Education Action Plan" of the European Commission, 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 emphasized in the platform governance communities. As a key term, digital access thus refers to issues of facilitating access to content and content-producing organizations. 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" (ROAM-principles, UNESCO, 2014).

The Office of the Special Rapporteur for Freedom of Expression of the Inter-American Commission on Human Rights (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 sense, the increasing number of court rulings holding that government shutdowns of the internet are illegal should also be mentioned, such as the judgment of the Community Court of Justice of the Economic Community of West African States, in *Amnesty International Togo v. The Togolese Republic* (ECOWAS, Court of Justice, 2020; Pollicino, 2020).

Similarly, the Council of Europe's "Human Rights Guidelines for Internet Service Providers" (2008) sought to raise awareness for those entities providing access and 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). As previously developed statements, through the opinions of several courts and other institutions throughout the world (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

Alliance for Affordable Internet – A4AI. (2020). *Meaningful Connectivity: A New Target to Raise the Bar for Internet Access.* Available at: <https://a4ai.org/meaningful-connectivity/>.

Belli, L., De Filippi, P. (2015). *Net neutrality compendium: Human rights, free competition, and the future of the Internet*. Springer. Available at: <https://www.ohchr.org/Documents/Issues/Expression/Telecommunications/LucaBelli.pdf>.

Belli, L. (2017). *Network self-determination and the positive externalities of community networks. Community Networks: The Internet by the People, for the People. Official Outcome of the UN IGF Dynamic Coalition on Community Connectivity.* FGV Direito Rio. Available at: <https://www.intgovforum.org/multilingual/index.php?q=filedepot_download/4391/1132>.

Carpentier, N. (2007). Participation and interactivity: changing perspectives. The construction of an integrated model on access, interaction, and participation. In: *New media worlds*. Oxford University Press. 214-230.

Council of Europe. (2019). *Digital Citizenship Education Handbook.* Available at: <https://rm.coe.int/16809382f9>.

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

Council of Europe. (2008). 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.* Available at: <https://rm.coe.int/16805a39d5>.

European Commission. (2020). *Digital education action plan.* Available at: <https://ec.europa.eu/education/education-in-the-eu/digital-education-action-plan_en>.

Hilbert, Martin. (2016). The bad news is that the digital access divide is here to stay: Domestically installed bandwidths among 172 countries for 1986–2014. *Telecommunications Policy*, 567-581. Available at: <http://doi.org/10.1016/j.telpol.2016.01.006>.

Lanza, Edison. (2017). 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. OEA/Ser.L/V/II, CIDH/RELE/INF.* Available at: <http://www.oas.org/en/iachr/expression/docs/publications/INTERNET_2016_ENG.pdf>.

Pollicino, Oreste. (2020). The Rights to Internet Access: Quid Iuris? 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>.

Prodi, Romani. (2020). *La connessione sia un diritto umano (letter to Sassoli). 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/>.

Ribble, M. (2011). *Digital citizenship in schools: Nine elements all students should know.* International Society for Technology in Education.

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

UN. (2019). Broadband Commission. *The State of Broadband 2019*. Available at: <https://broadbandcommission.org/publications/Pages/SOB-2019.aspx>.

UNESCO. (2014). Fostering freedom online: the role of Internet intermediaries. *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. (June 25, 2020). Application n°. ECW/CCJ/APP/61/18IN. ECOWAS Court of Justice. Available at: <https://africanlii.org/node/7135>.

Websites:

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

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

2 Accountability

Nicolo Zingales

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 (OECD, 2014), 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'Donnell, 1998). The latter is imposed upon organizations 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 maximized when individuals can act via civic organizations ('civil society') or media. Horizontal accountability is maximized 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 organizations put in place appropriate technical and organizational 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

Husovec, M. (2017). *Injunctions against intermediaries in the European Union: accountable but not liable?* (Vol. 41). Cambridge University Press.

O'Donnell, G. A. (1998). Horizontal accountability in new democracies. *Journal of democracy*, *9*(3), 112-126.

Organization for Economic Co-operation and Development – OECD. (2014). *Accountability and Democratic Governance. Orientations and Principles for Development*. Available at: <https://www.oecd-ilibrary.org/development/accountability-and-democratic-governance_9789264183636-en>.

3 Aggregator

Clara Almeida

The term 'aggregators' refers to digital platforms that have as an economic activity the matching of providers with consumers. To be considered an aggregator, a digital platform must present three defining characteristics: (1) direct relationship with users; (2) zero marginal cost for serving users; and (3) demand-driven multi-sided networks with decreasing acquisition costs (Thompson, 2017). The term was coined by Ben Thompson (2015) in his 'Aggregation Theory', which explains why aggregators came to dominate the markets in which they compete in (Abrol, 2017), by controlling the relationship with consumers on a given value chain.

According to Thompson's 'Aggregation Theory', any consumer market's value chain is divided into (1) suppliers, (2) distributors and (3) user (consumer) experience. Before the digital era, competitors wished to increase their profits by forming horizontal monopolies at one stage of the chain, or vertical monopolies by integrating stages. The formation of a vertical monopoly depended on controlling distribution. The internet, however, made possible the costless distribution of digital goods, ending the need of distributors to maintain exclusive relations with suppliers. Also, the transaction costs came down to zero, making it possible for distributors to connect directly with consumers at scale (Thompson, 2015).

Since distributors no longer need to compete for exclusive relationships with suppliers, the suppliers' priority becomes reaching consumers. The success of a business now no longer depends on controlling the distribution stage, but the user experience. Aggregators are able to directly access users, therefore, aim to provide the best user experience to attract more users. The more users accessing a platform, the more suppliers the platform will gather. With more suppliers, the user experience is enhanced, which keeps a virtuous cycle that benefits the aggregator (Thompson, 2015).

This cycle creates strong winner-takes-all effects on the market, enabling the aggregator to become a horizontal monopoly over the last stage of the chain. Given the characteristics of aggregators, they are able to potentially scale users worldwide. The suppliers, however, lose their bargaining power, because they lose influence in the relationship with the customer, since customers make their decisions based on the experience provided by the platforms they access. Suppliers adapt to platform specifications and demands, no longer distinguishing themselves from their competitors (Abrol, 2017).

Thompson (2017) also classifies the aggregators in four different levels:

1. Supply acquisition: aggregators which market power comes from their relationship with users, but they have to acquire their supply (e.g., Netflix);
2. Supply transactions costs: aggregators that do not need to acquire their supply, but do incur in transaction costs to bring suppliers onto to their platform (e.g., Uber);
3. Zero supply costs: aggregators that do not own the supply, but also do not incur in transaction costs to bring supplier to their platform (e.g., Google as a search engine)
4. Super-aggregators: aggregators that intermediate multi-sided market combining users, suppliers, and advertisers, incurring zero marginal costs on all the sides (e.g., Facebook).

References

Abrol, Anuj. (2017). An introduction to Aggregation Theory. *Medium*. <https://anujabrol.com/an-introduction-to-aggregation-theory-589b360ac373>.

Thompson, Ben. (2015). Aggregation Theory. *Stratechery*. <https://stratechery.com/2015/aggregation-theory/>.

Thompson, Ben. (2017). Defining Aggregators. *Stratechery*. <https://stratechery.com/2017/defining-aggregators/>.

4 Amplification

Paddy Leerssen

In the context of platform governance, 'amplification' refers to actions (typically by platforms) that magnify the visibility or reach of certain information.. 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 (The White House, 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. (2018). *Communication on Tackling Online Disinformation: A European Approach.* Available at: <https://www.whitehouse.gov/presidential-actions/executive-order-preventing-online-censorship>.

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

Kaye, D. (2019b). *Promotion and protection of the right to freedom of opinion and expression.* Available at: <https://www.ohchr.org/Documents/Issues/Opinion/A_74_486.pdf>.

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

5 Appeal

Luca Belli

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 regarding 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 defenses addressed in the proceeding that is challenged – usually defined as ‘first-instance proceeding’ (ALI, 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 analyzed (so-called ‘questions of fact’) and matters of procedure or discretion.

In general terms, appeals can:

1. constitute a repeat exercise of the lower body's decision-making, both in terms of fact-finding and the application of the relevant law/rules to those facts;
2. 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);
3. 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 is 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:

1. 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 utilized improperly.
2. When the procedure was applied unfairly.
3. When the decision is not reasonable.

4. When the decision is not proportional.
5. When the decision is not compatible with Human Rights obligations.
6. When the decision contradicts the legitimate expectations of an individual or entity subject to a given set of rules.
7. 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/UNIDROIT. (2006). *Principles of Transnational Civil Procedure.* Available at: <https://www.unidroit.org/instruments/civil-procedure/ali-unidroit-principles/>.

Marshall, P. D. (2011). A comparative analysis of the right to appeal. *Duke J. Comp. & Int’l L.*, *22*, 1.

6 Application Program Interface

Nicolo Zingales

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* (2012), 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

de Souza, C. R. et al. (2004). Sometimes you need to see through walls: a field study of application programming interfaces. In: *Proceedings of the 2004 ACM conference on Computer supported cooperative* work, 63-71.

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

Van Rooijen, A. (2010). *The software interface between copyright and competition law: a legal analysis of interoperability in computer programs.* Kluwer Law International BV.

Case Law:

Oracle Am., Inc. v. Google Inc. (June 2012). No. 3:10-cv-. 3561, N.D. Cal., ECF No. 1211.

Websites:

MyData (2017). *Declaration of Principles.* Available at: <https://mydata.org/declaration>.

7 Arbitration

Luã Fergus and Laila Lorenzon

The World Intellectual Property Organization (WIPO, 2020a) 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 more straightforward definition given by the Cambridge Dictionary states that the arbitration process is a way of “solving an argument between people by helping them to agree to a standard and acceptable solution”. It is essential to highlight that both sides in the dispute must 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 the reason why this procedure is often chosen rather than the litigation process. An impartial third party, the arbitrator, resolves the disputes, and their decision is legally binding for all parties.

As for the online process of arbitration, its premise follows the same path. The difference is that conflicts can be resolved entirely online by video calls for hearings and software for uploads of evidence (documents, photos, videos, etc.). Thus, online arbitration makes it possible to resolve disputes without one having to appear in person, and, by that, it minimizes the costs of the process.

The most prominent example of online arbitration is in disputes over Internet domains (Mania, 2015). One can make a parallel between domain names on the Internet and the system of business identifiers protected by intellectual property rights and has existed long before the arrival of the Internet. Arbitration processes are helpful to solve conflicts regarding both issues. The most common reason for disputes over Internet domain names comes from the practice known as ‘cybersquatting’ – when a random person registers a domain name under famous people or business trademarks and offers them for sale at prices far beyond the cost of registration.

Under the Uniform Domain Name Dispute Resolution Policy (UDRP), "any domain name registered in the international domains, such as .com, is subject to this dispute resolution mechanism" (WIPO, 2020b). First, the parties must consent to solve their domain disputes and submit the dispute details to the chosen institution. Then, the Complaint can file a case against the entity by filling a form on WIPO's website. Subsequently, the entity can file a Response. The WIPO is responsible for the administrative process, with the role to ensure the absence of interests' conflicts by choosing panelists or experts with impartiality and independence from a "roster of independent individuals qualified for deciding such cases" (WIPO, nd). Finally, the system notifies the parties of the decision – which they must follow.

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. ICANN adopted the Uniform Domain Name Dispute Resolution Policy (UDRP) based on the report's recommendations. Under the standard dispute clause of the Terms and Conditions for registering a gTLD domain name, the registrant must submit to the UDRP proceedings. In addition, the Protocol on Cybersecurity in International Arbitration (Cybersecurity Protocol) guides reasonable information security measures that the parties and arbitrators can take, particularly considering increasingly virtual hearings and paperless document transfer.

An example of alternative online arbitration practices for resolving domain name conflicts is the 'Sistema de Administração de Conflitos de Internet', a.k.a. SACI-Adm, developed by the Internet Steering Committee in Brazil (CGI.br). This method serves to resolve 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. SACI-Adm procedures' scopes are limited to the domain's cancellation and transfer requests, and its aspects are similar to those presented in the UDRP. The main difference between them is that in the Brazilian regulation, the .br Information and Coordination Centre (NIC.br) doesn't allow the transfer of the domain name in conflict with the beginning of the arbitration with the SACI-Adm procedure until its termination (Angelini, 2012).

There's also the 'baseball-style arbitration process used between Google and news publishers in Australia. In this form, the organization selects an arbitrator to decide the main issues in dispute between two or more parties. Such arbitration is named 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 essential to state that there's a difference between Online Dispute Resolution (ODR) and online arbitration. ODR is a vast field and encompasses many types of dispute resolution practices that use online methods and tools to explore the convenience and efficiency of internet communications and make disputes more accessible and faster. It 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). Online arbitration is an essential part of ODR, on the other hand. Two or more parties can solve any disagreement originating from their contractual relationship online – domain names, Business to Business (B2B) cross-border e-commerce disputes, and traditional cross-border commercial disputes (Amro, 2019).

References

Amro, I. (2019). *Online Arbitration in Theory and in Practice: A Comparative Study of Cross-border Commercial Transactions in Common Law and Civil Law Countries.* Cambridge Scholars Publishing. 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>.

Angelini, Kelli. (2012). *SACI: o Sistema Administrativo de Conflitos de Internet implementado para domínios no ".br".*

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

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

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

Petrauskas, F., Kybartienè, E. (2011). Online dispute resolution in consumer disputes. *Jurisprudencija*, *18*(3).

World Intellectual Property Organization – WIPO. (2020a). *What is Arbitration?* Available at: <https://www.wipo.int/amc/en/arbitration/what-is-arb.html>.

World Intellectual Property Organization – WIPO. (2020b). *WIPO's Anti-"Cybersquatting" Service Surpasses 50,000 Cases amid COVID-19 Surge*. Available at: <https://www.wipo.int/pressroom/en/articles/2020/article_0026.html>.

World Intellectual Property Organization – WIPO. (nd). *Frequently Asked Questions: Internet Domain Names*. Available at: <https://www.wipo.int/amc/en/center/faq/domains.html#5>.

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

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

8 Automated Decision Making

Rossana Ducato

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 is 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 behavioral advertising system that shows a particular 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 broad 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 them (Article 22 GDPR). In the literature, several doubts have been raised regarding the exact meaning of 'decision', 'solely automated', 'legal effects', and 'similarly affect' (Mendoza; Bygrave, 2017; Bygrave, 2020). The Working Party Article 29 (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 affect 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 adverse effects which may seriously impact the behavior of individuals, e.g., potentially leading to discrimination (for example, a system denying someone an employment opportunity) (WP29, 2018). However, the provision does not seem to entail an evaluation of the negative impact on groups (Veale and Edwards, 2018). Still, several authors have argued in favor of expanding the data protection framework from the individual level to the collective one (Taylor, Floridi, 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 performing, 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 must 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 their point of view; 3) to contest the decision. The implementation of such measures has been differently embraced by Member States (Malgieri, 2019).

Finally, if the ADM involves processing particular categories of data (defined in 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 High-Level Expert Group on Artificial Intelligence has stressed the importance of explainability 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 them. 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 Afro-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 in a significative more negative way a protected category if compared to others in a similar situation. For instance, to anchor the earnings of platform's drivers to the distance and time they travel appear to be a neutral decision. However, studies show that women drive at a lower average speed; therefore, they are likely to take fewer rides, and, consequently, their pay is substantially lower than their male colleagues (Cook et al., 2018).

Nevertheless, the anti-discrimination legal framework suffers important limitations since it covers only specific sectors and certain protected grounds. This situation is particularly critical in the digital discrimination brought by ADM, as the latter often transcends the traditional protected attributes (Borgesius, 2018; Xenidis, Senden, 2020). In online behavioral advertising, for example, people might be discriminated against because the inferential analytics draws correlations among apparently neutral data, thus exposing individuals to price discrimination or exclusions from lucrative job ads without them even being aware of how and based on which criteria they have been profiled (Wachter, 2020).

In the field of consumer protection, ADM has been recently taken into account in relation to the transparency of online marketplaces. The "Omnibus Directive", amending the Consumer Right Directive (Directive 2011/83/EC), established that when the price is personalized based on 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 the 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, organized or communicated by the trader". The concept of ranking is constructed in a technologically neutral way; therefore, it might consist of an ADM.

Another sector regulating ADM is medical devices. When a software standalone 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 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 rightsholder 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 that will lead to establishing upload filters and limiting 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, M., Bonnet, G. (2020). Legal and technical feasibility of the GDPR's quest for explanation of algorithmic decisions: of black boxes, white boxes and Fata Morganas. *European Journal of Risk Regulation*, *11*(1), 18-50.

Brkan, M. (2019). Do algorithms rule the world? Algorithmic decision-making and data protection in the framework of the GDPR and beyond. *International journal of law and information technology*, *27*(2), 91-121.

Bygrave, Lee. (2020). 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.

Cerf, Vint et al. (2018). *Joint Letter to the European Parliament*. Eletronic Frontier Foundation. Available at: <https://www.eff.org/files/2018/06/13/article13letter.pdf>.

Cook, C. et al. (2018). *The gender earnings gap in the gig economy: Evidence from over a million rideshare drivers*. National Bureau of Economic Research.

Directive 2019/790. 17 April 2019 on Copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC. European Parliament, Council of the European Union. <https://eur-lex.europa.eu/eli/dir/2019/790/oj>.

Edelman, B., Luca, M., Svirsky, D. (2017). Racial discrimination in the sharing economy: Evidence from a field experiment. *American economic journal: applied economics*, *9*(2), 1-22.

Edwards, L., Veale, M. (2017). Slave to the algorithm: Why a right to an explanation is probably not the remedy you are looking for. *Duke L. & Tech. Rev.*, *16*, 18.

European Commission. (2018). *Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679*, WP251, rev. 01. Available at: <https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612053>.

Finck, M. (2019). Smart Contracts as Automated Decision-Making under Article 22 GDPR. *International Data Privacy Law*, *9*, 1-17.

Goodman, B., Flaxman, S. (2016). *EU regulations on algorithmic decision-making and a 'right to explanation*'. Preprint.

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

Kaminski, M. E., Malgieri, G. (2020). Algorithmic impact assessments under the GDPR: producing multi-layered explanations. *International Data Privacy Law.* 19-28.

Kretschmer, M. et al. (2019). *The Copyright Directive: Articles 11 and 13 must go*, Statement from European Academics in advance of the Plenary Vote on 26 March 2019.

Malgieri, G., Comandé, G. (2017). Why a right to legibility of automated decision-making exists in the general data protection regulation. *International Data Privacy Law*.

Malgieri, G. (2019). Automated decision-making in the EU Member States: The right to explanation and other "suitable safeguards" in the national legislations. *Computer law & security review*, *35*(5).

Mantelero, A. (2018). AI and Big Data: A blueprint for a human rights, social and ethical impact assessment. *Computer Law & Security Review*, *34*(4). 754-772.

Mendoza, I., Bygrave, L. A. (2017). The right not to be subject to automated decisions based on profiling. In *EU Internet Law*. Springer, Cham. 77-98.

Quintais, J. et al. (2019). *Safeguarding user freedoms in implementing Article 17 of the copyright in the Digital Single Market Directive: recommendations from European Academics.* Available at: <https://www.dekuzu.com/en/docs/European-Academics-article-17-DSMD-SSRN-id3484968.pdf>.

Quintais, J. (2020). The new copyright in the digital single market directive: A critical look. *European Intellectual Property Review*. Available at: <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3424770>.

Reda, Julia. (2019). *EU copyright reform: Our fight was not in vain*. Available at: <https://juliareda.eu/2019/04/not-in-vain>.

Regulation (EU). 2017/745 of the European Parliament and of the Council of 5 April 2017 *on medical devices*, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC>.

Romero Moreno, F. (2020). Upload filters and human rights: implementing Article 17 of the Directive on Copyright in the Digital Single Market. *International Review of Law, Computers & Technology*, *34*(2), 153-182.

Samuelson, P. (2020). Pushing Back on Stricter Copyright ISP Liability Rules. *Michigan Technology Law Review, Forthcoming*.

Schmon, Christoph. (2020). *Copyright Filters are on a Collision Course with EU Data Privacy Rules*. Available at: <https://www.eff.org/deeplinks/2020/02/upload-filters-are-odds-gdpr>.

Taylor, L., Floridi, L. Sloot, Bart Van der, eds. (2016). *Group privacy: New challenges of data technologies.* Vol. 126. Springer.

Veale, M., Edwards, L. (2018). Clarity, surprises, and further questions in the Article 29 Working Party draft guidance on automated decision-making and profiling. *Computer Law & Security Review*, *34*(2), 398-404.

Wachter, S., Mittelstadt, B., Russell, C. (2017). Counterfactual explanations without opening the black box: Automated decisions and the GDPR. *Harv. JL & Tech.*, *31*, 841.

Wachter, S., Mittelstadt, B., Floridi, L. (2017). Why a right to explanation of automated decision-making does not exist in the general data protection regulation. *International Data Privacy Law*, *7*(2), 76-99.

Wachter, S. (2020). Affinity Profiling and Discrimination by Association in Online Behavioral Advertising. *Berkeley Tech. LJ*, *35*, 367.

Xenidis, R., Senden, L. (2019). EU non-discrimination law in the era of artificial intelligence: Mapping the challenges of algorithmic discrimination. In: Bernitz, U. et al. (eds), *General Principles of EU law and the EU Digital Order (Kluwer Law International, 2020)*, 151-182.

Zuiderveen Borgesius, Frederik. (2018). Discrimination, artificial intelligence, and algorithmic decision-making. Study for the Council of Europe. Available at: <https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73>.

9 Bot

Luã Fergus and Laila Lorenzon

‘Bot’ is a tech slang for ‘robot’. In this sense, they share the same conceptual core: a reprogrammable machine built to perform a variety of tasks (RIA, n.d.). In the specific case of online bots, they are labelled 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 soon replaced by this increasing 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, n.d.). The first category is subdivided in: ‘social bots’ simulate 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, especially 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’); and, finally, ‘helpful’ or ‘informational bots’, that surface helpful information and usually push notifications and breaking news stories.

The examples mentioned above are usually utilized to help/optimize human actions and tasks. In the opposite way, ‘malicious bots’ can be: ‘scrapers bots’ that are designed to steal content or vast 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 and attack websites or entire networks. In this latter case, devices that are affected are called 'zombies' and infected networks, 'botnets', a 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. For example, in September 2016, a botnet called Mirai was responsible for one of the biggest cyber-attacks in history when it launched a DDoS attack on the servers of Dyn, one of the primary DNS providers, which resulted in a blackout for various internet services (Antonakakis et al., 2017). Finally, a type of malicious bot that has pervasively dominated digital policy debates recently is the 'impersonators bots', a type of bot mimics human behavior predominantly to manipulate public opinion, spread disinformation, and exercise social control (Bessi, Ferrara, 2016; Howard et al., 2018).

References

Antonakakis M., et al. (2017). Understanding the Mirai Botnet. In: *26th {USENIX} security symposium. {USENIX} Security* 17. 1093-1110.

Bessi, A., Ferrara, E. (2016). Social bots distort the 2016 US Presidential election online discussion. *First Monday*. 21(11-7).

Bort, J. (2014*).* Bill Gates: People don't realize how many jobs will soon be replaced by Software Bots. *The Business Insider*, USA.

Delaney, K. J. (2017). The robot that takes your job should pay taxes, says Bill Gates. *Quartz*.

Franklin, S., Graesser, A. (1996). Is it an Agent, or just a Program? A Taxonomy for Autonomous Agents. *In International workshop on agent theories, architectures, and languages.* 21-35. Springer, Berlin, Heidelberg.

Gorwa, R., Guilbeault, D. (2020). Unpacking the social media bot: A typology to guide research and policy. *Policy & Internet*. 12(2). 225-248.

Jones, S. (2015). How I learned to stop worrying and love the bots. *Social Media+ Society*. 1(1).

Howard, P. N., Woolley, S., Calo, R. (2018). Algorithms, bots, and political communication in the US 2016 election: The challenge of automated political communication for election law and administration. *Journal of information technology & politics*. 15(2). 81-93.

Roberts, E. (2020). *Bad bot report 2020: Bad bots strike back*. Imperva blog.

Websites:

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/>.

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

Richard Wingfield

NB: While the term ‘child pornography’ has been used 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 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.

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.

Existing regulatory responses

Most states have sought to comply with their obligations under international law to prohibit child sexual abuse material through criminalization, 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 criminalization, 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 organizations, 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. (2018). *Child Sexual Abuse Material: Model Legislation & Global Review*. 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

Chris Marsden

‘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 restrict the commodities that they will carry. Carriers may refuse to carry dangerous goods, improperly packed goods, and goods that they are unable to carry due to size, legal prohibition, or lack of facilities (Longley, 1967; Ridley, Jasper, Whitehead, 1982); Encyclopædia Britannica, 2009).

This definition offers several reasons not to common carry that can be extended to Internet Access Providers – spam and viruses, for instance, may be refused. In common law countries such as the United Kingdom and the 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 defects 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 Denial of Service (DoS) attacks. It even might be suggested that streaming video streams can be ‘jettisoned’ in order to allow other traffic to progress during peak time 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 government and transport networks or facility, in which an exchange is made: for the privileges of classification as a common carrier, private actors are granted the rights and benefits that an ordinary private carrier would not. Barbara Cherry explained that 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 the common carriage.

The United Kingdom Carriers Act of 1830 was the first legislation for the 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': UK Carriers Act 1830 Chapter 68), 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 includes provisions for common carriage and 'Parliamentary trains' (low-cost trains that stopped at all stations, later known as 'milk trains' because they collected milk from all stations 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 (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 (1994:435) 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* (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.* (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 may require new regulation.

References

Cherry, Barbara. (2008). Back to the Future: How Transportation Deregulatory Policies Foreshadow Evolution of Communications Policies. *The Information Society*. Volume 24 Issue 5 pp 273–291 <https://doi.org/10.1080/01972240802356059>.

Encyclopædia Britannica. Common Carrier. Available at: <http://www.britannica.com/EBchecked/topic/128177/common-carrier>.

Longley, H. N. (1967). *Common Carriage of Cargo*. Matthew Bender & Co.: New York.

Noam, Eli M. (1994). Beyond liberalization II: the impending doom of common carriage. *Telecommunications Policy*. 435-452.

Ridley, Jasper and Geoffrey Whitehead. (1982). *The Law of the Carriage of Goods by Land, Sea and Air*, 6th ed., Shaw: Crayford, Kent.

UK. (1830). Carriers Act. Chapter 68. Available at: <http://www.opsi.gov.uk/RevisedStatutes/Acts/ukpga/1830/cukpga_18300068_en_1>.

Case Law:

Lane v. Cotton (1701) 1Ld.Raym. 646.

Nat. Ass'n of Reg. Utility Com'rs v. F.C.C., 525 F.2d 630 (D.C. Cir. 1976).

Western Union Tel. Co. v. Call Pub. Co., 181 U.S. 92, 21 S. Ct. 561, (1901).

12 Content Creator/Influencer

Catalina Goanta/Giovanni de Gregorio

During the past ten 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 utilized 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).

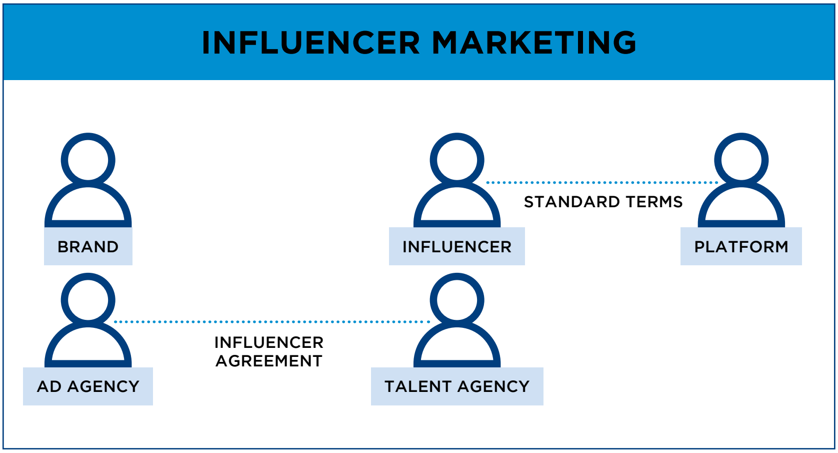


Figure 1 – Goanta & Wildhaber, 2020

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 fewer 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 programs 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 transactions 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 behavior or political ideas for commercial gain.

References

De Gregorio, G., Goanta, C. (2020). *The Influencer Republic: Monetizing Political Speech on Social Media*. Available at: <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3725188>.

European Commission. (2018). *Behavioural Study on Advertising and Marketing Practices in Online Social Media.* Final Report. Available at: <https://ec.europa.eu/info/sites/default/files/osm-final-report_en.pdf>.

Goanta, C., Wildhaber, I. (2020). Controlling influencer content through contracts: a qualitative empirical study on the Swiss influencer market. In: *The Regulation of Social Media Influencers*. Edward Elgar Publishing.

Riefa, C., Clausen, L. (2019). *Towards Fairness in Digital Influencers' Marketing Practices*. Journal of European Consumer and Market Law, 64.

Stokel-Walker, C. (2019). Instagram: Beware of bad influencers. The Guardian. Available at: <https://www.theguardian.com/technology/2019/feb/03/instagram-beware-bad-influencers-product-twitter-snapchat-fyre-kendall-jenner-bella-hadid>.

Word of Mouth Marketing Association – WOMMA. (2017). The WOMMA Guide to Influencer Marketing. Available at: <http://getgeeked.tv/wp-content/uploads/uploads/2018/03/WOMMA-The-WOMMA-Guide-to-Influencer-Marketing-2017.compressed.pdf>.

Websites:

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

13 Content

Richard Wingfield

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).

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.

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). The UK’s Draft Online Safety Act (published in 2020) also uses the term ‘content’, and defines it as “anything communicated by means of an internet service, whether publicly or privately, including written material or messages, oral communications, photographs, videos, visual images, music and data of any description”. 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, although its proposals for a Digital Services Act (published in 2020) use the term ‘content’ in the context of ‘illegal content’.

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*.* *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>.

14 Content/Web Monetization

Catalina Goanta

'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 narrower 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 must 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 on 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 content creators/influencers), bring views, likes and clicks to platforms that are kept 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 are 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 their own branded goods to their fan base.

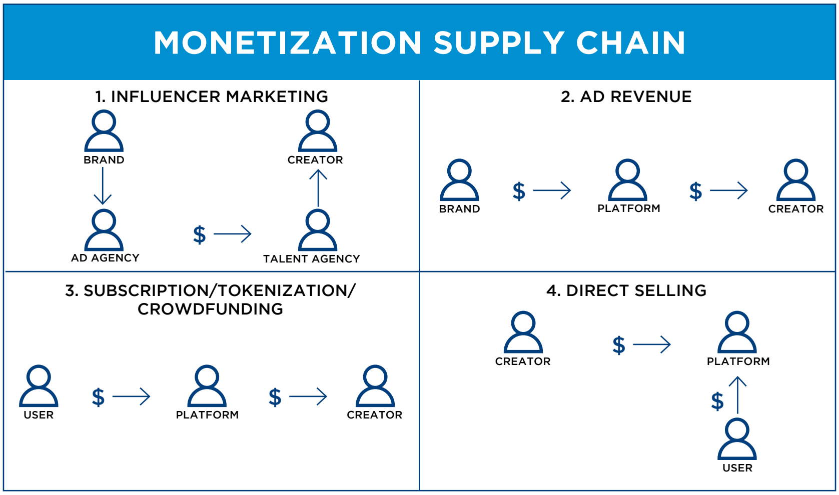


Figure 1 – Monetization business models

The Web Monetization protocol

In the past decades, information transfer protocols have been used to get as much information as possible on user behavior. Under the guise of personalizing advertising to fit individual needs and preferences, behavioral 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 that 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. (2020). Review of online targeting: Final report and recommendations. Available at: <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/864167/CDEJ7836-Review-of-Online-Targeting-05022020.pdf>.

Melendez, Ken. (2019). *The State of Web Monetization.* Medium. Available at: <https://coil.com/p/kenmelendez/The-State-of-Web-Monetisation/KTVijO7ah>.

Statista. (2020). Social media – Statistics & Facts. Available at: <https://www.statista.com/topics/1164/social-networks/>.

Websites:

Solid. <https://solid.mit.edu>.

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

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

15 Coordinated Flagging

Cynthia Khoo

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 the 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, “[i]n 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 an ostensible broader political purpose that goes beyond engaging in discriminatory online abuse for its own sake. For instance, 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”. Here, the coordinated attacks are explicitly framed by the perpetrators as part of a broader political objective (in contrast to other instances of discriminatory online abuse and harassment, which also advance broader political objectives but are not always acknowledged to have this effect):

> 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, 2012:1972-73).

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, 2016:421).

References

Crawford, K., Gillespie, T. (2016). What is a flag for? Social media reporting tools and the vocabulary of complaint. *New Media & Society*. 18(3), 410-428.

Fiore-Silfvast, B. (2012). User-generated warfare: A case of converging wartime information networks and coproductive regulation on YouTube. *International Journal of Communication*, *6*, 24.

16 Coordinated Inauthentic Behavior

Cynthia Khoo

'Coordinated inauthentic behavior' (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 behavior 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 behavior 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 (2020) highlights:

> 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.

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 behavior' 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 being CIB for its own internal purposes. As douek (2020) explains,

> 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.

References

douek, e. (2020). *What does "coordinated inauthentic behavior" actually mean?* Slate. Available at: <https://slate.com/technology/2020/07/coordinated-inauthentic-behavior-facebook-twitter.html>.

Facebook. (2020). *Coordinated Inauthentic Behavior Report.* Available at: <https://about.fb.com/news/2020/04/march-cib-report/>.

Gleicher, Nathaniel. (2018). *Coordinated Inauthentic Behavior Explained.* Available at: <https://about.fb.com/news/2018/12/inside-feed-coordinated-inauthentic-behavior/>.

17 Co-regulation

Rolf H. Weber

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 rulemaking are called (i) co-regulation, (ii) regulated self-regulation, (iii) directed self-regulation or (iv) audited self-regulation (Weber, 2014). The most common model balancing the interests of international organizations, States, businesses, and civil society is 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 a model – coined by Grainger (1999) and 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; Marsden, Meyer, Brown, 2020).

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 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, Brown, 2020).

Many examples for co-regulatory mechanisms exist, 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. In addition, co-regulation can have a positive impact on the behavior of the concerned actors. Joint efforts of various stakeholders also allow the government to assess the standards' representativeness and to judge the best practices' appropriateness.

References

Grainger, G. S. (1999). *Broadcasting, Co-Regulation and the Public Good*. Conseil supérieur de l'audiovisual. Available at: citese-erx.ist.psu.edu.

Hoffmann-Riem W. (2000). *Regulierung der dualen Rundfunkordnung: Grundfragen*. Nomos-Verlag-Ges, Baden Baden.

Marsden, C., Meyer, T., Brown, I. (2020). Platform Values and democratic elections: How can the law regulate digital disinformation?. *Computer Law & Security Review*, 36.

Senn, M. (2010). *Non-state Regulatory Regimes: Understanding institutional transformation.* Springer Science & Business Media.

Weber, R. H. (2015). *Realizing a New Global Cyberspace Framework*. Berlin, Heidelberg.

Case law:

Magyar Tartalomszolgáltatók Egyesülete (MTE) and *Index.hu zrt v. Hungary* (ECHR 2016).

18 Content Curation

Paddy Leerssen

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 (2016) define curation as the "production, selection, filtering, annotation or framing of content". 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 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*.

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, R., Freuler, J. O., Fagan, C. (2018). *The invisible curation of content: Facebook's news feed and our information diets*. Available at: <http://webfoundation.org/docs/2018/04/WF_InvisibleCurationContent_Screen_AW.pdf>.

Cobbe, J., Singh, J. (2019). Regulating recommending: motivations, considerations, and principles. *Considerations, and Principles*.

Helberger, N., Kleinen-von Königslöw, K., Van Der Noll, R. (2015). *Regulating the new information intermediaries as gatekeepers of information diversity*.

Napoli, P. M. (2015). Social media and the public interest: Governance of news platforms in the realm of individual and algorithmic gatekeepers. *Telecommunications Policy*. 39(9), 751-760.

Napoli, P. M. (2019). *Social media and the public interest.* Columbia University Press.

Thorson, K., Wells, C. (2016). Curated flows: A framework for mapping media exposure in the digital age. *Communication Theory*. *26*(3), 309-328.

Van Couvering, E. (2010). Search engine bias: the structuration of traffic on the World-Wide Web. *Doctoral dissertation. The London School of Economics and Political Science (LSE).* Available at: <http://etheses.lse.ac.uk/41>.

19 Dark Patterns

Nicolo Zingales

This entry discusses: (I) the notion of ‘dark patterns’, 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, 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, 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.

(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 seven 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 cancelling), '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' shopping 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 an apparent ad) and 'Confirmshaming' (framing choice in a 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).

Another useful taxonomy is the one developed by Mathur et al. (2020), who identify five dimensions alongside which dark patterns can be measured: asymmetric burden, covertness, deceptiveness, hiding of information, and restrictiveness of available choices.

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” (CNIL, 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.

Some work has been done on the connection between dark patterns and data protection: user studies, which conduct experiments aimed at gauging the impact of specific dark patterns (see e.g. Utz et al., 2019); measurement and detection studies, which through semi-automated techniques aim to measure the prevalence of dark patterns in a specific domain (e.g., Mathur et al., 2020); and finally, compliance studies, aiming to examine the compatibility of certain dark patterns with existing law (e.g., Nowens et al., 2020).

References

Authority for Consumers and Markets – ACM. (2020). *Guidelines on the Protection of the online consumer.* Boundaries of online persuasion. 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. (2019). *Shaping Choices in the Digital World From dark patterns to data protection: the influence of UX/UI design on user empowerment*. 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, C. et al. (2016). Tales from the Dark Side: Privacy Dark Strategies and Privacy Dark Patterns. *Proc. Priv. Enhancing Technol.*, *2016*(4), 237-254.

Conti, G., Sobiesk, E. (2010). Malicious interface design: exploiting the user. In: *Proceedings of the 19th international conference on World Wide Web*. 271-280. Available at: <https://doi.org/10.1145/1772690.1772719>.

Greenberg, S., et al. (2014). Dark patterns in proxemic interactions: a critical perspective. In *Proceedings of the 2014 conference on Designing interactive systems*. 523-532.

Hausman, D. M. en B. Welch (2010). Debate: To Nudge or not to nudge. *Journal of Political Philosophy*, 123-136.

Luguri, J., Strahilevitz, L. J. (2021). Shining a light on dark patterns. *Journal of Legal Analysis*, *13*(1), 43-109. Available at: <https://ssrn.com/abstract=3431205>.

Mathur, A., et al. (2019). Dark patterns at scale: Findings from a crawl of 11K shopping websites. *Proceedings of the ACM on Human-Computer Interaction*, *3*(CSCW). 1-32. Available at <https://arxiv.org/abs/1907.07032>.

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

Nouwens, M., et al. (2020). Dark patterns after the GDPR: Scraping consent pop-ups and demonstrating their influence. In: *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1-13.

Transatlantic Consumer Dialogue and Heinrich Böll Stiftung. (2020).

Privacy in the EU and US: Consumer experiences across three global platforms. Available at: <https://eu.boell.org/en/2019/12/11/privacy-eu-and-us-consumer-experiences-across-three-global-platforms>.

Utz, C., et al. (2019). (Un)Informed Consent: Studying GDPR Consent Notices in the Field. Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security (London, United Kingdom). CCS '19. Association for Computing Machinery, New York, NY, USA, 973–990. Available at: <https://doi.org/10.1145/3319535.3354212>.

20 Data Portability

Vittorio Bertola, Nicolo Zingales and Luca Belli

The current debate on (data) portability finds its predecessor in the discussions on Mobile Number Portability (MNP), which emerged towards the end of the 1990s. MNP requires that, when switching from one provider to another, mobile telephone users must be able to keep their telephone number(s). The world's first country to introduce MNP was Singapore in 1997, followed by the UK, Hong Kong and the Netherlands in 1999 (Buehler; Haucap, 2004). As of the early 2000s, many countries had created MNP requirements, especially in Europe.

The rationale driving the introduction of MNP is rather straightforward and understanding it is essential to grasp the importance of the current debate on data portability. Indeed, MNP aims at fostering competition, while maximising consumer interests. Before, the introduction of MNP, consumers of mobile telecommunications services were required to give up their number and adopt a new one when switching providers. In practice, this situation represented a considerable transaction cost dissuading users to switch to competing providers thus jeopardising healthy competition. Importantly such cost included informing all their personal and professional networks about their new number, missing potentially valuable calls from people that still had the old number, etc.

Hence MNP is considered to be beneficial for multiple reasons (Viard, 2004). Besides avoiding incurring in the above-mentioned cost, consumers switching provider thanks to MNP are more likely to use the services they prefer the most, at the condition that suit them the most. The resulting increase in competition among providers become beneficial for both consumers who decide to export their number and those who decide not to do so.

In the same perspective, 'data portability' is aimed at empowering consumers, providing them more control over their personal data, while fostering competition between service providers. 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 behavior, such as activity logs; it does not include further elaborations, such as inferred or derived data.

The Guidelines also identify 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 the 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 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 the threshold of substantiation of risk is, 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 standardized data exchange remain in their infancy*" (CERRE, 2020).

In its European conception, the right to data portability has a consumer dimension. Indeed, GDPR prescribes that this right only applies when the data processing is carried out by automated means and based either on data subjects' consent or the execution of a contract. Considering that consent and contract are the most common bases for processing consumer personal data, we can argue that, at the EU level, portability primarily aims at improving consumer welfare and fostering competition amongst consumer-facing services.

The concrete implementation of data portability, however, is not as straightforward as it may appear, due to some normative limits and the practical complexity of the issue, which relies upon the sound interoperability as an essential precondition. The European example is telling in this regard. Article 20 GDPR prescribes that a data subject willing to enjoy data portability will receive personal data "*which he or she has provided to a controller, in a structured, commonly used and machine-readable format.*" By receiving the data in such interoperable format, the data subject will be able to enjoy the right to transmit those data to another controller" without *hindrance from the controller to which the personal data have been provided.*"

However, the abovementioned interoperable format is not defined by GDPR, nor the regulation identifies an authority with remit to identify such format. Although an earlier version of the RDP, which was included in article 18.3 of the non-final version of GDPR, attributed the authority to define to the appropriate format European Commission, the final version adopted by the EU Legislator has delegated definition of such format to the market. As emphasised by Paul de Hert and colleagues (2018), the EU Commission's role in the first proposal was crucial to achieve interoperability and effective implementation of the RDP as "the European Commission's role was conceived as a progressive specification of data format, but also for "technical standards, modalities and procedures for the transmission of personal data". In other words, it could have helped to conform normative standards to technological developments, and it could have fostered a concrete and effective development of interoperability of all digital services. Unfortunately, in the final version, such reference to the "standardisation" role of the European Commission has been removed."[1] The agreement over format will depend on specific sectors of the (digital) economy and quintessentially relies on the evolutions of interoperability debates. However, it is important to mention that two formats are already sufficiently widespread to be considered as good candidates for the role of "commonly used format", which can provide the interoperability needed to facilitate data portability:

- XML (Extensible Markup Language), which has been utilised for more than 3 decades, and is an integral part of every web application. Be it a configuration file, mapping document, or a schema definition, XML facilitates data interchange by giving structuring data and facilitating dynamic configurations.
- JSON, which aims at storing data in a map format relatively neat and easy to comprehend. JSON is said to be slowly replacing XML because of several benefits like ease of data modelling or mapping directly to domain objects, offering a relatively predictable and understandable structure.

Importantly, portability debates are not limited to Europe and the right to data portability has been enshrined also in other data protection frameworks, that followed GDPR, such as the Brazilian General Data Protection Law, better known as "LGPD", in its Portuguese acronym. LGPD enshrines data portability in article 18.V, granting every data subject the right to "the portability of data to another service or product provider upon express request, in accordance with the ANPD

[1] See De Hert et al. (2018).

regulations". Conspicuously, portability of personal data does not include data that have been anonymised by the controller.

According to LGPD, the Brazilian Data Protection Authority, better known as "ANPD", plays a fundamental role to enable data portability. Indeed, article 40 of LGPD prescribes that "the supervisory authority may establish interoperability standards for purposes of portability, free access to data and security, and on the retention time of the registrations, especially in view of the need and transparency."

However, it is important to note that the ANPD "may" define such format but is not obliged to do so. Critically, according to its regulatory agenda, released in January 2021, the establishment of such formats is not even mentioned in the list of regulatory initiatives to be undertaken in the first two years of activities of ANPD. Lastly, it is also important to temper an excessively rosy picture, regarding data portability, stressing that, bedsides numerous advantages, it can also entail risks, notably regarding cybersecurity, but also the potential to jeopardise third-party rights, intellectual property rights and trade secrets, to name a few. One of the most significant risks associated with the new data portability right is identity theft. The GDPR recognises that a controller cannot comply with a data portability request if it cannot identify the data subject and that it may request additional information to confirm the identity of the data subject. The Article 29 Working Party view, now integrated by the European Data Protection Board, is that data controllers should implement authentication procedures to confirm the identity of the data subject and mitigate risks (Article 29 WP, 2017).

As regards the risks that portability measure may have on third party rights, Graef, Husovec and Purtova (2018) have noted that the inappropriate implementation of the RDP may lead to questions about the protection of intellectual property rights, particularly if the data to be shared or APIs to be offered involve trade secrets or other protected information (Graef, Husovec, Purtova, 2018).

References

Brasil. ANPD. (28/01/2021). *No Dia da Proteção de Dados, ANPD publica agenda regulatória bianual da autoridade para 2021-2022.* Available at: <https://www.gov.br/anpd/pt-br/assuntos/noticias/no-dia-da-protecao-de-dados-anpd-publica-agenda-regulatoria-bianual-da-autoridade-para-2021-2022>.

Article 29 WP, *Guidelines on the right to data portability*, 16/ EN, WP 2042, rev01, as last revised and adopted on 5 April 2017.

Buehler, S., and Haucap, J. (2004). Mobile Number Portability. In: *Journal of Industry, Competition*, and Trade. 4, 223-228. Available at: <https://www.itu.int/ITU-D/finance/work-cost-tariffs/events/tariff-seminars/lithuania-04/haucap%20.pdf>.

De Hert, P., Papakonstantinou, V., Malgieri G., Beslay, L., Sanchez. I. (2018). The right to data portability in the GDPR: Towards user-centric interoperability of digital services. *Computer Law & Security Review*. 193–203.

European Council. (2016). Regulation 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. Available at: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32016R0679>.

Graef, I., Husovec, M., Purtova, N. (2018). Data portability and data control: lessons for an emerging concept in EU law. *German Law Journal*, *19*(6), 1359-1398. Available at: <https://ssrn.com/abstract=3071875 or <http://dx.doi.org/10.2139/ssrn.3071875>.

Krämer, J., et al. (2020). *Making data portability more effective for the digital economy: Economic implications and regulatory challenges*. Centre on Regulation in Europe (CERRE).

Malgieri, G. (2016). *Trade Secrets v. Personal Data*: a possible solution for balancing rights. *International Data Privacy Law*, *6*(2), 102-116.

Viard, V.B. (2004). Do Switching Costs Make Markets More or Less Competitive? The Case of 800-Number Portability, Stanford Graduate School of Business Research Paper Series Paper No. 1773(R1). HTU <http://ssrn.com/abstract=371921UTH>.

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

21 Defamation

Richard Wingfield

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

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 (ordinarily 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 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 defamatory if done with actual malice, which means that there was a real willingness to harm the defamed person.

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

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, Delfi AS 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 were not clearly unlawful, the platform was not operated on a commercial basis, and the platform took general measures to prevent defamatory comments.

References

ARTICLE 19. (2017). *Defining Defamation: Principles on Freedom of Expression and Protection of Reputation*. Available at: <https://www.article19.org/data/files/medialibrary/38641/Defamation-Principles-(online)-.pdf>.

Ligabo, A. (2007). 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, Paragraph 47.

UN. (2011). UN Human Rights Committee. (General comment N°. 34: Article 19: Freedoms of opinion and expression, UN Doc. CCPR/C/CG/34, Paragraph 47.

22 Deindexing

Courtney Radsch

See also deplatforming/de-platforming.

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, the “right to be forgotten” 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

Nicolo Zingales and Catalina Goanta

See also deplatforming/de-platforming.

Demonetization refers to 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 programs, 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

Caplan, R., Gillespie, T. (2020). Tiered governance and demonetization: The shifting terms of labor and compensation in the platform economy. *Social Media+Society*, 6(2).

Koi, C. (2020). YouTube Help. *Wrongfully demonetized, how many months without revenue on average?* Available at: <https://support.google.com/YouTube/thread/56781687?hl=en>.

Goanta, C., Ranchordás, S. (2020). *The Regulation of Social Media Influencers*. Edward Elgar Publishing.

Lobato, R. (2016). The cultural logic of digital intermediaries: YouTube multichannel networks. *Convergence*, 22(4), 348-360.

24 Deplatforming/De-platforming

Courtney Radsch

‘Deplatforming’, or ‘de-platforming’, 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, YouTube or Twitter;
- Commerce platforms such as Amazon of the Apple Store;
- 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 behavior, 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 of 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 Instagram 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. (2018). *Removing Myanmar Military Officials from Facebook*. <https://about.fb.com/news/2018/08/removing-myanmar-officials/>.

Facebook. (2018). *Update on Myanmar*. Available at: <https://about.fb.com/news/2018/08/update-on-myanmar/>.

Chandrasekharan, E., et al. (2017). You can’t stay here: The efficacy of reddit’s 2015 ban examined through hate speech. *Proceedings of the ACM on Human-Computer Interaction*, 1-22.

Clifford, B., Powell, H. C. (2019). De-platforming and the Online extremist’s dilemma. *Lawfare Blog*, 6. Available at: <https://www.lawfareblog.com/de-platforming-and-online-extremists-dilemma>.

25 Device Neutrality

Luã Fergus and Laila Lorenzon

'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 achieving equal access to applications, contents, and services, which is essential to achieve an open Internet. It is a fundamental civil rights issue, ensuring 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 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 favor their content or demote competitors.

Its main principle is the idea that the consumers should have the right to uninstall software, apps and content they are not interested in. Thus, companies should also give them 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) 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 the bill "S.2484 Disposizioni in materia di fornitura dei servizi della rete internet per la tutela della concorrenza e della libertà di accesso degli utenti". This bill was approved at the Chamber of Deputy, and it is still waiting to be voted in Senate. It addresses Device Neutrality in Article 4, stating that that:

> Users have the right to find online, in a format suitable to the desired 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 an 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). Similarly, device neutrality defends that the users should have the right to non-discrimination of the services or apps in their devices regardless of the hardware companies, Net Neutrality defends the right to non-discrimination by Internet service providers, regardless of the content or applications utilized by the Internet users, unless such discriminatory treatment is necessary and proportionate to the achievement of a legitimate aim (Belli, De Filippi, 2015; Belli, 2017). Thus, one's about equal access to applications and the other about equal access to the Internet.

Another set of rules that could go in the opposite direction of device neutrality is the anti-circumvention laws, which provides penalties for those who wish to make changes to their devices and operational systems. At first, these rules were created to protect 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, n.d.).

References

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

Belli, L., De Filippi, P. (2016). General introduction: Towards a multistakeholder approach to network neutrality. In: *Net Neutrality Compendium*. Springer, Cham. 1-7. Available at: <https://www.ohchr.org/Documents/Issues/Expression/Telecommunications/LucaBelli.pdf>.

Belli, L. (2017). Net Neutrality, zero-rating and the Minitelisation of the Internet. *Journal of Cyber Policy*, *2*(1), 96-122.

Borchert, Katharina. (2016). *EU Copyright Law Undermines Innovation and Creativity on the Internet.* Available at: <https://blog.mozilla.org/en/mozilla/eu-copyright-law-undermines-innovation-and-creativity-on-the-internet-mozilla-is-fighting-for-reform/>.

Doctorow, Cory. (2019). *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. 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. (2017). *After Net Neutrality, Device Neutrality.* Available at: <https://www.hermescenter.org/net-neutrality-device-neutrality/>.

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

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

26 Digital Rights

Chris Wiersma

In a framework of 'digital citizenship' (see 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 I 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 consensus resolutions (2012, 2014, 2016, 2018, 2021) 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, e.g., Vesnic-Alujevic et al., 2021; also, Shulga-Morskaya 2020).

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

Ribble, M. (2011). *Digital citizenship in schools: Nine elements all students should know*. International Society for Technology in Education.

UN. (2012). The Promotion, Protection and Enjoyment of Human Rights on the Internet. A_HRC_20_L13. Geneva: United Nations. Available at: <http://ap.ohchr.org/documents/E/HRC/d_res_dec/A_HRC_20_L13.doc>.

UN. (2014). The Promotion, Protection and Enjoyment of Human Rights on the Internet. A/HRC/RES/26/13. Geneva: United Nations. Available at: <https://ap.ohchr.org/documents/dpage_e.aspx?si=A/HRC/RES/26/13>.

UN. (2016). 'The Promotion, Protection and Enjoyment of Human Rights on the Internet'. A/HRC/RES/32/13. Geneva: United Nations. Available at: <https://ap.ohchr.org/documents/dpage_e.aspx?si=A/HRC/RES/32/13>.

UN. (2018). 'The Promotion, Protection and Enjoyment of Human Rights on the Internet'. Geneva: United Nations. A/HRC/38/L.10. Available at: <https://ap.ohchr.org/documents/alldocs.aspx?doc_id=29960>.

UN. (2021). A/HRC/47/L.22. Geneva: United Nations. Available at: <https://undocs.org/A/HRC/47/L.22>.

Shulga-Morskaya, T. (2020). *E-démocratie. Aménager la démocratie représentative?* (L'Harmattan).

Vesnić-Alujević, L. (2021). 'Imagining democratic societies of the future: Insights from a foresight study', *Futures & Foresight Science*, 3:e60, <https://doi.org/10.1002/ffo2.60>.

Websites:

Nine Elements of Digital Citizenship. <https://www.digitalcitizenship.net/nine-elements.html>.

27 Disinformation

Giovanni De Gregorio

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, Derakhshan, 2017). False information would include information disseminated as intentionally false and impossible to verify to mislead the public (Allcott, 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" (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, 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 recognize it as such' and underlines that disinformation does not include illegal speech" (e.g., hate speech).

Disinformation is not a phenomenon of the digital age. Its digital dissemination is a novelty. The digital dimension entails the worldwide reach of online content beyond territorial boundaries and the 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 tends to gravitate toward sensationalism, the need for constant novelty, and the aim of achieving profits instead of professional ethical standards and civic responsibility (Marwick; 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 defining how information is organized online. Beyond media strategy to disseminate disinformation, scholars have emphasized the role of the political context. Indeed, the role of technology platforms, bots and foreign spies has tended to be overemphasized (Benkler, 2018). Political parties and, in particular, populism movements, have relied on strategies of disinformation to support their political ideas (Bayer et al., 2019), and political micro-targeting contributes to this purpose (Dobber 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; 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; 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 criminalization 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 scrutinize 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; 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, H., Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of economic perspectives*, 31(2), 211-36.

Balkin, J. M. (2018). Free speech is a triangle. *Colum. L. Rev.*, 118, 2011.

Bayer, J., et al., E. (2019). Disinformation and propaganda – impact on the functioning of the rule of law in the EU and its Member States. *European Parliament, LIBE Committee, Policy Department for Citizens' Rights and Constitutional Affairs*.

Benkler, Y., Faris, R., Roberts, H. (2018). Network propaganda: Manipulation, disinformation, and radicalization. In: *American politics*. Oxford University Press.

Chesney, B., Citron, D. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. *Calif. L. Rev.,* 107, 1753.

De Gregorio, Giovanni, Perotti, Elena. (2019). Tackling Disinformation around the World: a new policy report. *World Association of News Publishing Focus*.

Dobber, Tom & Ó Fathaigh, Ronan & Zuiderveen Borgesius, Frederik J. (2019). The regulation of online political micro-targeting in Europe. Internet Policy Review, 8, 4.

European Commission. (2018). A multi-dimensional approach to disinformation. *Report of the independent High-Level Group on fake news and online disinformation*.

Ireton, C., Posetti, J. (2018). *Journalism, fake news & disinformation: handbook for journalism education and training.* UNESCO Publishing. Available at: <https://unesdoc.unesco.org/ark:/48223/pf0000265552>.

Majó-Vázquez, Silvia et al. (2020). Volume and Patterns of Toxicity in Social Media. Conversations during the COVID-19 Pandemic. *Reuters Institute*. Available at: <https://reutersinstitute.politics.ox.ac.uk/volume-and-patterns-toxicity-social-media-conversations-during-covid-19-pandemic>.

Martens, Bertins et al. (2018). The Digital Transformation of News Media and the Rise of Disinformation and Fake News. *JRC Digital Economy Working Paper*, 2.

Marwick, Alice, Lewis, Rebecca. (2017). Media Manipulation and Disinformation Online. *Data & Society*.

Pollicino, O., De Gregorio, G., Laura, S. (2020). *Europe at the Crossroad: The Regulatory Conundrum to Face the Raise and Amplification of False Content in Internet*.

Pitruzzella, Giovanni and Pollicino, Oreste. (2020). *Disinformation and Hate Speech. A European Constitutional Perspective.* Bocconi University Press.

Robinson, Olga et al. (2019). *A Report on Antidisinformation Initiative.* Available at: <https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2019/08/A-Report-of-Anti-Disinformation-Initiatives>.

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

Tandoc Jr, Edson C. et al. (2018). Defining "Fake News": A Typology of Scholarly Definitions, 6(2), *Digital Journalism*, 137.

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

28 Dispute Resolution [Online]

Catalina Goanta

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-Kohle, 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 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

Citron, D., Jurecic, Q. (2018). *Platform Justice: Content Moderation at an Inflection Point*. Aegis Series Paper n°. 1811. Available at <https://www.lawfareblog.com/platform-justice-content-moderation-inflection-point>.

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

Google. (2020). *What is a Content ID claim?* Available at: <https://support.google.com/YouTube/answer/6013276?hl=en>.

Hörnle, J. (2019). *Cross-border internet dispute resolution.* Cambridge University Press.

Katsh, M. E., Rabinovich-Einy, O. (2017). *Digital justice: technology and the internet of disputes*. Oxford University Press.

Loebl, Z. (2019). *Designing Online Courts: the Future of Justice is open to all.* Wolters Kluwer.

Mistelis, L. (2003). ADR in England and Wales: a successful case of public private partnership. *ADR Bulletin*, 6(3), 53-55.

Thomson, S. Sherr, A. (2012). Definitions of Online Dispute Resolution. In: Gramatikov, M. (Ed.). (2012). *Costs and quality of online dispute resolution: a handbook for measuring the costs and quality of ODR*. Maklu.

Tworek, H., Ó Fathaigh, R., Bruggeman, L., Tenove, C. (2020). D*ispute resolution and content moderation: Fair, accountable, independent, transparent, and effective*.

29 Due Diligence

Luca Belli

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 minimize 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 authorization 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, P. M. (1977). Due diligence in the international law of liability. *Legal Aspects of Transfrontier Pollution*, 369.

30 Duty of Care

Chris Wiersma

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 e.g., 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

De Streel, A. et al. (2020). *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. Available at<https://d.docs.live.net/67c027bb92cc8900/%253c:<https://www.europarl.europa.eu/RegData/etudes/STUD/2020/652718/IPOL_STU(2020)652718_EN.pdf>.

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>.

Le Docte, E., Am Zehnhoff, H.-W. (2011). Le Docte : viertalig juridisch woordenboek; Dictionnaire de termes juridiques en quatre langues; Rechtswörterbuch in vier Sprachen; Legal dictionary in four languages.

van Eijk, T.M. van Engers, C. Wiersma, C. Jasserand W. Abel. (2010). Moving Towards Balance: A Study into Duties of Care on the Internet. *Institute for Information Law Research Paper No.2012-16.* Available at: <https://www.ivir.nl/publicaties/download/Moving_Towards_Balance.pdf>.

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

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

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

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

31 End to End Encryption

Catalina Goanta, Vittorio Bertola

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

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

Lee, Micah., Grauer, Yael. (2020). Zoom Meetings aren't End-to-end Encrypted, Despite Misleading Marketing. *The Intercept.* Available at: <https://www.loopinsight.com/2020/03/31/zoom-meetings-arent-end-to-end-encrypted-despite-misleading-marketing/>.

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

SSD.EFF.ORG. (2018). A Deep Dive on End-to-End Encryption: How Do Public Key Encryption Systems Work? *Surveillance Self-Defense*. Available at: <https://ssd.eff.org/module/deep-dive-end-end-encryption-how-do-public-key-encryption-systems-work>.

SSD.EFF.ORG. (2020). End-to-end encryption. *Surveillance Self-Defense*. Available at: <https://ssd.eff.org/en/glossary/end-end-encryption>.

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

32 Fact-checking

Luca Belli

Fact-checking has gained prominence as a concept, in light of the global proportions gained by the divulgation of fabricated and misleading content, frequently categorized 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 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

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

33 Federated Service

Yasmin Curzi

Federation refers to a group of individuals, servers or applications that operate in a decentralized manner, following a common protocol. It typically involves the exchange of data and services between two or more cloud service providers, sometimes for security reasons (logging into a website through another service's more secure protocols), but also for collaborative purposes – e-mails, a prime example of federated services, can be sent across service providers –, among others. One peculiarity is that the decentralized structure may enable the sharing of knowledge derived from data without actually transferring such data from one server (or from an individual's device) to another. Specifically, federated learning (also known as collaborative learning) is a machine learning technique that trains an algorithm across multiple decentralized edge devices or servers holding local data samples, without exchanging them.

A more extensive use of federated services could encourage user independence and reduce market concentration, since users would tend to be less dependent on market dominant services (e.g., the case of WhatsApp in Brazil, where using the service is almost necessary for ordinary digital social interactions). In a more extensive implementation, federated services could further improve users' security and privacy through federated computing, with even further data decentralization. I.e., instead of data being stored in massive central servers, data could remain in the users' devices and be accessed and processed locally – the result of the computation could then be privately aggregated before being sent to the server, ensuring the proper anonymization is taking place. Overall, it could result in a reduction in the amount of information that is stored by companies, whose often liberal use of user data is hard to assess and (potentially) punish.

Another off-shoot of federated services are the so-called federated networks, e.g. 'Mastodon', which is a decentralized, open-source software for networking services. The infrastructure of these types of networks composed of several other networks which have their own structure, with its own content moderation features, terms of service and community policies. They are often part of a larger 'Fediverse', which allows interactions between users and members in different communities. Nevertheless, Mastodon has general rules that apply to all communities, not unlike a constitution which applies to all states of a federated country. It is a good example to be followed vis-à-vis interoperability features, since it allows users to communicate with other communities without having to integrate them.

34 Filter

Vittorio Bertola

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 endpoints 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 centers 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, *rendezvous 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*.

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

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

Borgolte, Kevin. Chattopadhyay, Tithi. Feamster, Nick. Kshirsagar, Mihir. Holland, Jordan. Hounsel, Austin. Schmitt, Paul. (2019). *How DNS over HTTPS is Reshaping Privacy, Performance, and Policy in the Internet Ecosystem.* Available 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, OJ L Document 32019L0790>.

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.

European Parliament and Council (EC) Regulation (EU) 2018/302 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, OJ L Document 32018R0302.

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

35 Flagging

Cynthia Khoo

'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

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

36 (Digital) Gatekeeper

Nicolo Zingales

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 communication occurs in an environment more closely akin to a public sphere. As a result of this, she identifies two different categories: (1) *Internet gatekeepers*, which are those gatekeepers that control the flow of information; and (2) *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 (2021) 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 et al. (2015) reinforce this understanding, focusing on the control of critical resources, rather than on access to and supply of information, as a measure of their 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:

1. 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 (Cremer et al., 2019)
2. intermediation power, linked to having "unavoidable trading partner" status (Competiton Law 4.0 report, 2019)
3. 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. 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.

The work of these reports has fed into recent proposals, and particularly in the case of the Cremer et al. report, it has led to a formalization of the term gatekeeper in the proposed Digital Markets Act. The Act establishes in its Article 3 that a provider of "core platform services" (a term which refers to platforms operating in specific markets, such as online intermediation services, online search services, online social networking services, video-sharing platform services, number-independent communication services, operating systems, cloud computing services, and advertising services) shall be designated as gatekeeper if: (a) it has a significant impact on the internal market; (b) it operates a core platform service which serves as an important gateway for business users to reach end users; and (c) it enjoys an entrenched and durable position in its operations or it is foreseeable that it will enjoy such a position in the near future.

Furthermore, the same article establishes a presumption that a provider of core platform services satisfy the above criteria in the following way:

(a) the requirement of (a) where the undertaking to which it belongs achieves an annual EEA turnover equal to or above EUR 6.5 billion in the last three financial years, or where the average market capitalisation or the equivalent fair market value of the undertaking to which it belongs amounted to at least EUR 65 billion in the last financial year, and it provides a core platform service in at least three Member States; (b) the requirement of (b) where it provides a core platform service that has more than 45 million monthly active end users established or located in the Union and more than 10 000 yearly active business users established in the Union in the last financial year; and (c) the requirement in paragraph 1 point (c) where the thresholds in point (b) were met in each of the last three financial years. Finally, Article 3 (6) establishes that in designating an entity as gatekeeper, the Commission shall take into account the following elements:

(a) the size, including turnover and market capitalisation, operations and position of the provider of core platform services;

(b) the number of business users depending on the core platform service to reach end users and the number of end users;

(c) entry barriers derived from network effects and data driven advantages, in particular in relation to the provider's access to and collection of personal and non-personal data or analytics capabilities;

(d) scale and scope effects the provider benefits from, including with regard to data;

(e) business user or end user lock-in;

(f) other structural market characteristics.

References

Citron, Danielle. (2014). *Hate Crimes in Cyberspace.* Harvard University Press.

Commission Competition Law 4.0. (2020). *A New Competition Framework for the Digital Economy*. BMWi.

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

Crémer, J. de Montjoye, YA. Schweitzer, H. (2019). *Competition Policy for the digital era. European Commission Directorate-General for Competition*. Available at: <https://ec.europa.eu/competition/publications/reports/kd0419345enn.pdf>.

Furman, J., Coyle, D., Fletcher, A., Marsden, P., McAuley, D., Unlocking digital competition. Report of the Digital Competition Expert Panel (March 2019), available at <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/785547/unlocking_digital_competition_furman_review_web.pdf>.

Helberger, N., Kleinen-von Königslöw, K., Van Der Noll, R. (2015). *Regulating the new information intermediaries as gatekeepers of information diversity*.

Kadri, Thomas E. (2021). *Digital Gatekeepers. Texas Law Review*. 99.

Kraakman, R. H. (1986). Gatekeepers: The anatomy of a third-party enforcement strategy. *Journal of Law, Economics, & Organization*, 2(1), 53-104.

Laidlaw, E. B. (2010). A framework for identifying Internet information gatekeepers. *International Review of Law, Computers & Technology*, 24(3), 263-276.

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

Van Loo, R. (2020). The New Gatekeepers: Private Firms as Public Enforcers. *Va. L. Rev*., 106, 467.

37 Disinformation (Gendered)

Yasmin Curzi

Disinformation is a worldwide phenomenon that mostly targets public persons such as politicians, journalists, actors and actresses, influencers, etc. Nevertheless, its impacts on its targets are felt differently by its victims and spectators. As shown by a Heinrich Böll Stiftung policy brief on the subject (Judson, 2021), gender and sexuality directly affect how the victims are affected and how these attacks are carried out. The recent initiative of taking an intersectional perspective regarding this phenomenon may help with the development of public policies that can more properly address it.

Regarding the definition of 'gendered disinformation', studies conducted in different countries (e.g., for Brazil, see Azmina; InternetLab, 2020; for Germany and Russia, see Wilfore, 2021) are showing that the attacks against women and LGBTQI+ people are often connected with stereotypes linked to their sexual identity/affect orientation, while the attacks against men aim at their ideas, opinions and past activities in public life. As Judson (2021) clarifies "[g]endered disinformation is manipulated information that weaponises gendered stereotypes for political, economic or social ends". It also includes fake and doctored sexual images, coordinated abuse, and caricatures. The current definition provided by the organization 'She Persisted' is that

> [G]endered disinformation is the spread of deceptive or inaccurate information and images against women political leaders, journalists and female public figures, following story lines that draw on misogyny, as well as gender stereotypes around the role of women in order to undermine their perceptions of their participation in public life.

The conceptualization of 'gendered disinformation' is part of a broad initiative related to making explicit political gender-based violence that takes place in imperfect democracies. According to Kristina Willfore (2021), "[g]endered disinformation attacks online are a well-known tactic that illiberal actors around the world – including Russia, Hungary and Brazi – have developed to undermine their opponents". The background for this type of political violence is the historical segregation of minorities from public institutions. Racism, sexism and LGBTphobia in these spaces is not new. Nevertheless, social media can be a major amplifier of disinformation campaigns, providing attackers with tools to sponsor and boost problematic content on a very large scale, as well as fostering the reach and permanence of such content in an unprecedent way. It also increases the challenges minorities face in belonging and occupying such spaces – both online and offline. The main effect of this phenomenon is to undermine equal participation, thus also potentially undermining the democratic institutions.

Gendered disinformation reports also shows that language and semiotics are vital for its spread. As a report by Judson et al (2020) on disinformation campaigns in Poland points out, state-sponsored campaigns, linked to the country's conservative far right, seek to redefine social movements' terminologies, such as 'women's rights' by linking it to the abortion agenda and then equating it with 'killing children'. It is a strategy to modify societal comprehension about this subject. It also targets LGBTQI+ people in the country, linking their agenda with assault and aggression.

In Brazil and in the UK, reports show that female politicians are often called 'hysterical', 'stupid', 'immoral', and other. These offenses are taken to another level if the woman in question is black – the stereotypes of the 'ugly/angry/mad black women' are then deployed, putting into question their capabilities and mental stability.

References

Azmina; InternetLab. (2020). MonitorA: Report on online political violence on the pages and profiles of candidates in the 2020 municipal elections. Available at: <https://azmina.com.br/wp-content/uploads/2021/03/5P_Relatorio_MonitorA-ENG.pdf>.

Judson, Ellen. (2021). Gendered disinformation: 6 reasons why liberal democracies need to respond to this threat. *Heinrich Böll Stiftung.* Available at: <https://eu.boell.org/en/2021/07/09/gendered-disinformation-6-reasons-why-liberal-democracies-need-respond-threat?utm_campaign=thinktech_6&utm_medium=email&utm_source=RD+Station&dimension1=democracy#_ftn53>.

Judson, Ellen. Atay, Asli. Krasodomski-Jones, Alex. Lasko-Skinner, Rose. Smith, Josh. (2020). Engendering hate: the contours of state-aligned gendered disinformation online. Demos, UK. Available at: <https://demos.co.uk/wp-content/uploads/2020/10/Engendering-Hate-Report-FINAL.pdf>.

Wilfore, Kristina. (2021). The gendered disinformation playbook in Germany is a warning for Europe. Available at: <https://www.brookings.edu/techstream/the-gendered-disinformation-playbook-in-germany-is-a-warning-for-europe/>.

Website:

She Persisted Organization. Available at: <https://www.she-persisted.org/why>.

38 Governance

Paddy Leerssen

The concept of governance offers a 'decentered' 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" (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).

Despite these various shadings and permutations, several authors tend to see 'governance' as broadly synonymous with regulation, though connoting an emphasis on the role of private actors in processes of rulemaking and enforcement. However, the equivalence between governance and regulation seems to be visible only in English-language literature.

As noted by Belli (2016; 2019) the use of the term governance in reference to the Internet frames the mechanisms that stimulate the interaction and association of different stakeholders in a political space where divergent ideologies and economic interests are confronted. In this context, governance can be considered as the set of processes and institutions that organize comparison of heterogeneous ideas and perspectives and, ideally, promote the collaborative proposal of new regulatory instruments aimed at solving specific problems. On the contrary, regulation can be considered as the collection of the different regulatory instruments that are the product of governance. In the view of Frison-Roche (2003), the objective of regulation is to foster equilibrium and ensure the proper functioning of complex systems, characterized by the presence of a plurality of actors, animated by divergent purposes and interests. In the latter case may be contractual, such as the terms and conditions (Venturini, Belli, 2016) defining the rules for the use of web platforms (Belli, Zingales, 2017), mobile applications and Internet access networks, or technical, such as the algorithms, standards and protocols defining the software and hardware architectures that determine what users can and cannot do in the digital environment (Reidenberg, 1998; Lessig, 2006; Belli, 2016).

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, Gollatz (2016). This broader, non-regulatory conception of governance encompasses all forms of coordination and interaction which lead to the creation of rules and principles that guide conduct, such as, for instance, standard setting by Internet Service Providers or online platforms. However, to prevent 'governance' from expanding to cover 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

Belli, Luca. (2019). Internet Governance and Regulation: A Critical Presentation. In: Belli, Luca, Cavalli, Olga. *Internet Governance and Regulations in Latin America.* FGV Direito Rio. Available at: <https://www.gobernanzainternet.org/book/book_en.pdf>.

Belli, L., Zingales, N. (2017). *Platform regulations: how platforms are regulated and how they regulate us.* Leeds. Available at: <https://bibliotecadigital.fgv.br/dspace/handle/10438/19402>.

Belli, Luca. (2016). *De la gouvernance à la regulation de l'Internet.* Paris: Berger-Levrault.

Black, J. (2008). Constructing and contesting legitimacy and accountability in polycentric regulatory regimes. *Regulation & governance*, 2(2), 137-164. Available at: <https://onlinelibrary.wiley.com/doi/full/10.1111/j.1748-5991.2008.00034.x>.

Hofmann, J., Katzenbach, C., Gollatz, K. (2017). Between coordination and regulation: Finding the governance in Internet governance. *New Media & Society*, 19(9), 1406-1423.

Lessig, Lawrence. (2006). Code and Other Laws of Cyberspace. Version 2.0. Aufl. New York.

Reidenberg, J. R. (1997). Lex informatica: The formulation of information policy rules through technology. Tex. L. Rev., 76, 553. Available at: <https://gorwa.co.uk/files/platformgovernance.pdf>.

Van Dijck, J., Poell, T., De Waal, M. (2018). *The platform society: Public values in a connective world*. Oxford University Press.

Van Eeten, Michel and Milton Mueller (2013). *Where is the governance in Internet governance?*.

Venturini, J., Belli, L. 2016*. Terms of service and human rights: an analysis of online platform contracts*. Revan, in collaboration with the Council of Europe and FGV Direito Rio. Available at: <https://bibliotecadigital.fgv.br/dspace/handle/10438/19402>.

39 Harassment

Terri Harel and Yasmin Curzi

This entry presents: (I) a brief history of the concept of harassment; (II) some of its variations – e.g., harassment in the working place, in the streets; and (III) the concept of online harassment or cyberharassment.

(i) The initiative of feminist lawyers in the US, at the 1980s, for the typification of harassment as a sex discrimination act, under Title IV of the Civil Rights Acts of 1964, was a novelty that made it possible for women to bring cases of abuse and harassment in the workplace to the courts. It was seen as a 'feminist invention' because of the deep naturalization of violent behaviours against women and LGBTQI+ people in society. The deconstruction of such violences is still an issue, but women have achieved their legal recognition (Honneth, 1996) in many countries, and can demand reparation.

(ii) Naming experiences 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 was increasingly highlighted in global movements of 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 that differentiates it from harassment in workplaces: the perpetrators are almost always anonymous to the victims, making it difficult for punitive institutions to address it or even for the law to typify it properly in most cases. Both harassment in the workplace and 'street harassment' are seen by the literature (MacKinnon, 2018; 1986) as a form of power expression of the privileged against minorities, taking advantage of situations of vulnerability that exist because of several inequalities within society.

(iii) As a means of communication, 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" (2011:678).

'Online harassment' is an umbrella term that refers to a set of specific, damaging behaviours and tactics. Tactics include, but aren't limited to, coordinated behavior, coordinated attacks, cyberstalking, dogpiling, dog-whistling, doxing, vile and hateful comments, mob harassment and other harms. It occurs through several techniques (Jhaver et al., 2018:15), and has as its main victims those who diverge from the 'conservative status quo' (Fladmoe, Nadim, 2017). The degree of violence in the comments also grows according to the race, gender identity and sexuality of the target.

Franks (2011) differentiates cyberharassment from mere insults or juvenile behavior by pointing out that it targets, in most cases, individuals that belong to subordinate groups and profoundly affects their lives. The study "O Reino Sagrado da Desinformação" (The Sacred Kingdom of Disinformation), has pointed out that the far-right agenda has been pushed on social media in Brazil aimed at manipulating public opinion with disinformation. The study, developed by the Brazilian independent news organization "Gênero e Número" (2019), further points out that this agenda preferentially attacks female journalists and politicians by producing 'gendered disinformation', i.e., disinformation that has, at its core, stereotypes which diminish their female targets (such as the 'mad', 'crazy', 'ugly and resented' woman). Caplan and Marwick (2018) dubbed the networks of masculinists that often lead these anti-feminist campaigns the 'manosphere'.

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*.

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., et al. (2018). Online harassment and content moderation: The case of blocklists. *ACM Transactions on Computer-Human Interaction*, 25, 2, 1-33. Available at: <https://dl.acm.org/doi/abs/10.1145/3185593?casa_token=4L_hW6bPP48AAAAA:gojOpiq_6mGlpU_vN2Xg5HrXRLPX-vysyNHwmmKFYck4yUFv_w-qcTMILy5OnmqsYgFh3OKOHKep_w>.

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

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, 543-559.

40 Harm (Online Harm)

Yasmin Curzi and Cynthia Khoo

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 driving 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. The UK government white paper (UK Government, 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' means and who these "clear" or "less clear" content or activities harm. That leaves the content or activity to stand independently, 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. For example, 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

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

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

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

41 Hash/Hash Database

Courtney Radsch

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 to eradicate 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

GIFCT. (2020a). *GIFCT Transparency Report*. Available at: <https://gifct.org/wp-content/uploads/2020/10/GIFCT-Transparency-Report-July-2020-Final.pdf.

Gregoire, C. (2016). *First Microsoft PhotoDNA update adds Linux and OS X support, detections up to 20 times faster*. Microsoft on the Issues.

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

Microsoft. (2020). *PhotoDNA*. <https://www.microsoft.com/en-us/photodna>.

42 Hate Speech

Yasmin Curzi

This entry aims to (I) present the definition of hate speech according to human rights law and the specialized literature; and (II) draw a distinction between hate speech and harm.

Hate speech is not a new phenomenon, and neither are the attempts to address it. The International Covenant on Civil and Politics Act (ICCP, United Nations, 1966) prohibited (art. 20) "[the] *advocacy* of *national, racial or religious* hatred that constitutes incitement to discrimination, hostility or violence". Many other legal instruments subsequently tried to encompass 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 to "*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 to cultural changes, being incorporated by domestic legislations worldwide.

Despite the advantages to the existence of general definitions capable of encompassing all of what are to be considered hate manifestations, in a universalist approach, the lack of objectivity leaves enormous discretionary space for judges and punitive institutions in applying the legislation. This can lead to several issues for law enforcement. Similarly, when it comes to online hate speech, there is a huge space for the actions of social network companies to define what they consider to be hate speech and apply content moderation in the online environment.

Considering the online fora 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 Responsibility and Transparency Draft Bill) so that platforms commit to the maintenance of a safe digital sphere and the protection of democratic 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 '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 addressing of hate speech.

References

Citron, Danielle. (2014). *Hate Crimes in Cyberspace.* Harvard University Press.

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

Franks, M. A. (2020). *The Cult of the Constitution*. Stanford University Press.

United Nations (General Assembly). (1966). International Covenant on Civil and Political Rights. *Treaty Series*, *999*, 171.

43 Human Exploitation

Luã Fergus and Laila Lorenzon

There is no agreed definition of 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 – monetary or not – for obtaining the consent of others over whom one has authority) for 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 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 sold often as an illegal trade.

Among the major online platforms, Facebook has an extensive exclusive section dedicated to human exploitation in its Community Standards. In addition to the activities mentioned above, it also officially condemns content related to children sailing for illegal adoption, orphanage trafficking and orphanage voluntourism.

Furthermore, platforms usually prohibit content geared towards the recruitment of potential victims, facilitating human exploitation, and 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, onboard fishing boats, in the sex industry or in private homes, generating billions of dollars a year. In addition, 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 can help combat trafficking, e.g., using text analysis tools to identify a writing pattern in sexual ads (Mzezewa, 2017).

Finally, it is important to notice the literature discussing whether social media content moderators suffer a form of human exploitation in jobs where they have to see the most violent content daily and decide whether it should remain online or not (Roberts, 2016). Even though that there is little literature 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 (Cardoso, 2019).

References

Allain, J. (2012). *Slavery in international law: Of human exploitation and trafficking*. Martinus Nijhoff Publishers.

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, T. (2017). Hacks That Help: Using Tech to Fight Child Exploitation. *The New York Times*. <https://www.nytimes.com/2017/11/24/style/sex-trafficking-hackathon.html>.

Roberts, S. T. (2016). *Commercial content moderation: Digital laborers' dirty work*.

UN. (2000). UN General Assembly. Protocol to Prevent, Suppress and Punish Trafficking in Persons, Especially Women and Children, Supplementing the United Nations Convention against Transnational Organized Crime. 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>.

44 Human Review

Cynthia Khoo

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 which content 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, or their labour, are a crucial component of commercial content moderation systems (Roberts, 2016). They 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)—or conversely, ensuring continued systematized approval and circulation of erroneously under-moderated content, such as technology-facilitated gender-based violence, abuse, and harassment, as well as that aimed at other and intersecting historically marginalized groups.

References

Caplan, R. (2018). *Content or Context Moderation? Artisanal, Community-Reliant, and Industrial Approaches*. Data & Society. <https://datasociety.net/wp-content/uploads/2018/11/DS_Content_or_Context_Moderation.pdf>.

Keller, D. (2018). *Internet Platforms: Observations on Speech, Danger, and Money*. Hoover Institution. <https://www.hoover.org/sites/default/files/research/docs/keller_webreadypdf_final.pdf>.

Roberts, S. T. (2016). *Commercial Content Moderation: Digital Laborers' Dirty Work*. In Noble, S.U. and Tynes, B. (Eds.), *The intersectional internet: Race, sex, class and culture online* (pp. 147-159). New York: Peter Lang. <https://ir.lib.uwo.ca/cgi/viewcontent.cgi?article=1012&context=commpub>.

45 Incitement to Violence

Monika Zalnieriute

Traditional incitement to violence, enshrined in Article 20 of the ICCPR and criminalized in many domestic jurisdictions, has taken on new significance due to the proliferation of online platforms and the growth of global terrorism (Bayefsky, 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, §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 marginalization' (Google, 2020, §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" (Šimonović, 2018:5).

References

Avni, Micah. (2018). Incitement to Terror and Freedom of Speech. *Incitement to Terrorism*, 30– 36. Available at: <https://doi.org/10.1163/9789004359826_005>.

Bayefsky, Anne F., and Laurie R. Blank, eds. (2018). *Incitement to Terrorism*. Available at: <https://brill-com.wwwproxy1.library.unsw.edu.au/view/title/36109>.

Facebook. (2020). Violence and Criminal Behavior. *Community Standards*. Available at: <https://www.facebook.com/communitystandards/violence_criminal_behavior>.

Google. (2020). Terms and Policies – Currents Help. Available at: <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. Available at: <https://doi.org/10.1163/9789004359826_016>.

Izsák, Rita. (2015). *Report of the Special Rapporteur on Minority Issues*. A/HRC/28/64. Available at: <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. Available at: <https://doi.org/10.1163/9789004359826_017>.

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

Olteanu, A., Castillo, C., Boy, J., Varshney, K. (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>.

Šimonović, D. (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.

Twitter. (2019). *Violent Threats Policy.* Available at: <https://help.twitter.com/en/rules-and-policies/violent-threats-glorification>.

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

46 Inclusive Journalism

Milica Pesic and Tamara Gojkovic

In increasingly diverse societies, the label ‘developing democracies’ refers to countries in transition to democracy. The need for fair, accurate and responsible journalism stands at the top of the requests for rebuilding media for a 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 should involve the media/journalistic sector. Journalism is a vehicle to public conversation and civic action, and strengthening journalism training and education contributes to strengthening its value for society.

The UNESCO Media Development Indicators (MDI), tailored to identify how media reflect the diversity of society to fulfil its democratic potential, underline the importance of the presence of minority groups in mainstream media. Other free speech organizations, with a critical approach, have been recognizing the significance of diversity, pointing out 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, along with the OSCE Representative on Freedom of the Media, the Organization of African States Special Rapporteur on Freedom of Expression, and the ACHPR Special Rapporteur on Freedom of Expression and Access to Information, made a joint Declaration on Promoting diversity in the Broadcast Media. The declaration stressed that:

> [t]he 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 an inclusive society, ‘society for all’, overrides differences of race, gender, class, generation, and geography and ensures inclusion, equality of opportunity, and capability of all members of the society. Among the prerequisites for an inclusive society – respect for all human rights, freedom, the rule of law, and a solid civic society – equal access to public information and tolerance for cultural diversity education, are also critical. Furthermore, it provides opportunities to learn the history and culture of one's own and other societies, which cultivates the understanding and appreciation of different societies, cultures, and religions. In addition, the 'inclusive society' implies a radical set of changes through which society restructures itself to embrace all of its members.

Journalism that can relate to the idea of inclusive society can be called 'inclusive journalism'. Inclusive journalism challenges the status quo to prevent media from intentionally or unintentionally spreading prejudice, intolerance, and hate. This idea is rooted in and inseparable from the political notion of inclusive democracy.

Used interchangeably, inclusive democracy and inclusive society indicate a 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: (1) 'external', when there is an exposed exclusion of groups and individuals from the decision-making process; and (2) 'internal', when "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 an increasingly diverse society is to develop inclusive communicative competence. This ability involves reflective thinking, the 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' and an understanding of the relationship between the individual and community.

Unfortunately, most university journalism programs are so focused on developing academic disciplines by integrating theory and practice that they dislocate journalism from its natural embeddedness into the community. Mensing (2011) notes that most university journalism programs preserve the structure of education based on the industrial model of journalism. The author 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" (2011:16). 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 several academic disciplines (e.g., Sociology, Social Psychology and Political Science) that deal with the issue of social diversity and may offer valuable insights to journalism students, such as theories of media power and social function.

In all modules developed through the MDI's inclusive journalism program, the question of assessment is a vital tool for enabling students to engage critically with social diversity and acquire the skills necessary to conceptualize and produce a brilliant 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, 1999:152) requires an active demonstration of knowledge of contemporary journalism. It deals with functional expertise by setting up tasks that are an exciting and challenging learning experience for students rather than taking it as a judgmental instrument in the academic analysis of media. The assessment that includes formative and summative procedures might take different forms, as outlined in some of the modules developed within the MDIs inclusive journalism curriculum framework.

References

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

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

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

Young, Iris Marion. (2002). *Inclusion and democracy* [electronic resource] / Iris Marion Young. Notes: Electronic reproduction. Oxford: Oxford University Press. (Oxford scholarship online). Publisher: Oxford: Oxford University Press. Internet.

47 Information Fiduciary

Nicolo Zingales

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.

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

Balkin, J. M., & Zittrain, J. (2016). A Grand Bargain to Make Tech Companies Trustworthy. *The Atlantic*. <https://www.theatlantic.com/technology/archive/2016/10/information-fiduciary/502346/>.

Khan, L. M., & Pozen, D. E. (2019). A skeptical view of information fiduciaries. *Harv. L. Rev.*, *133*, 497.

48 Infrastructure

Luã Fergus and Laila Lorenzon

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 (e.g., 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). *The Guardian*. 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. et al. (2014). Computer Security Incident Response Teams (CSIRTs): An Overview. Global Cyber Security Capacity Centre. 1-23.

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. (2020). *Infrastructure definition*. <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/>.

49 Intermediary Liability

Luã Fergus and Laila Lorenzon

This entry defines what an 'intermediary' is, while mainly referring to a dedicated entry on liability.

In essence, intermediaries are entities providing services that enable internet communication between different users. 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 (e.g., 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.

References

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

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

50 Internet Safety/Security

Chris Wiersma

The terms internet safety and internet security are closely connected. In the words of the European Commission (2020:1), "[s]ecurity 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" (i2018:2). Examples of the implementation of this strategy by the US government are the executive orders by President Trump (2015-2020) 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). These 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 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 N. 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 UKs recently proposed online safety laws as put forward in the "Online Harms White Paper" (2019), which proposed a broad (statutory) **duty of care**. See also harm.

References

Council of Europe, *Convention on Cybercrime*, adopted 23 November 2001, entered into force 1 July 2004, ETS n°.185. Available at <https://www.coe.int/en/web/conventions/full-list/-/conventions/treaty/005>.

European Commission. Communication from the commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions, on the *EU Security Union Strategy*. Available at: <https://ec.europa.eu/info/sites/info/files/communication-eu-security-union-strategy.pdf>.

European Commission. (2016). The EU Code of conduct on countering illegal hate speech online. Available 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. EU Law on Cybercrime. Available at: <https://ec.europa.eu/home-affairs/what-we-do/policies/cybercrime_en>.

Huawei and the Chinese Government. (2019). *New IP protocol system.* Available at: <https://www.itu.int/md/T17-TSAG-C-0083>.

Internet Society. (2020). *ITU World Telecommunication Standardization Assembly 2020*. Available at: <https://www.internetsociety.org/resources/doc/2020/itu-wtsa-2020-background-paper/>.

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

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

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. (2020). *Executive Order on Addressing the Threat Posed by WeChat.* Available at <https://www.federalregister.gov/documents/2020/08/11/2020-17700/addressing-the-threat-posed-by-wechat-and-taking-additional-steps-to-address-the-national-emergency>.

The White House/US President Trump. (2020). *Executive Order on Addressing the Threat Posed by TikTok*. Available at: <https://trumpwhitehouse.archives.gov/presidential-actions/executive-order-addressing-threat-posed-tiktok/>.

The White House/President Trump. (2018). National Cyber Strategy. Available at: <https://trumpwhitehouse.archives.gov/wp-content/uploads/2018/09/National-Cyber-Strategy.pdf>.

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

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

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

World Telecommunication Standardization Assembly. (2020). Available at: <https://www.itu.int/en/ITU-T/wtsa20/Pages/default.aspx>.

51 Interoperability

Vittorio Bertola and Nicolo Zingales

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 with 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 programs 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

Van Rooijen, A. (2010). *The software interface between copyright and competition law: a legal analysis of interoperability in computer programs.* Kluwer Law International BV.

de Souza, C. R. et al. (2004). Sometimes you need to see through walls: a field study of application programming interfaces. In: *Proceedings of the 2004 ACM conference on Computer supported cooperative* work, 63-71.

IEEE. (1995). *Guide to the POSIX Open System Environment (OSE)*. vol., no., pp.0_3-, 1995, doi: 10.1109/IEEESTD.1995.81544. Available at: <https://ieeexplore.ieee.org/document/552903>.

52 Liability

Nicolo Zingales

See also intermediary 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 (a.k.a. 'direct') and secondary (a.k.a. '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 (e.g., 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 respondent 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), 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* (2013). 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* (2010:1) that its application depends on a number of factors, such as the nature of the relationship between the alleged authorizer 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* (2005), the US Supreme Court found inducement liability for peer-to-peer software offered by Grokster based on 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

Angelopoulos, C. (2020). Harmonising Intermediary Copyright Liability in the EU: A Summary. In: Frosio, G. (2020). *Oxford Handbook of Online Intermediary Liability*. Oxford University Press.

Arnold, R., Davies, P. S. (2017). Accessory liability for intellectual property infringement: the case of authorisation. *The Law Quarterly Review*, 133(Jul), 442-468.

Catty, H. (1999). Joint tort feasance and assistance liability. *Legal Studies*. 19(4). 489-514.

Davies, P. S. (2015). *Accessory Liability* (vol. 13). Bloomsbury Publishing.

Dinwoodie, G. B. (2017). A comparative analysis of the secondary liability of online service providers. In: *Secondary Liability of Internet Service Providers*. 1-72. Springer, Cham.

Jaani, Riordan. (2020). A Taxonomy of Intermediary Liability. In: Frosio, G. (2020). *Oxford Handbook of Online Intermediary Liability*. Oxford University Press.

Husovec, M. (2016). Accountable, not liable: Injunctions against intermediaries. Available at: <https://ssrn.com/abstract=2773768> or <http://dx.doi.org/10.2139/ssrn.2773768>.

Case Law:

Dramatico Entertainment Ltd v. British Sky Broadcasting, 3 CLMR 14 (EWHC 268 (Ch) 2012).

Fish & Fish Ltd v Sea Shepherd, 3 All ER 867 (EWCA Civ 544 2013).

Metro-Goldwyn-Mayer Studios Inc. v. Grokster, 545 (U.S. 913 2005).

53 Marketplace

Enguerrand Marique

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 service, 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 narrowest 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 licensing), 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 license requirements). The success of this assertion varies largely from country to country.

References

Calo, R., Rosenblat, A. (2017). The taking economy: Uber, information, and power. *Colum. L. Rev*.

Chander, A. (2013). How Law Made Silicon Valley.*Emory LJ*, *63*, 639.

Communication from the commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. *A European agenda for the collaborative economy*. COM/2016/0356 final. Available at: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2016%3A356%3AFIN>.

Edelman, B., Stemler, A. (2019). From the digital to the physical: Federal limitations on regulating online marketplaces. *Harv. J. on Legis. 56*, 141.

Goldman, E. (2018). An overview of the United States' section 230 internet immunity.

Goldman, E. (2018). The complicated story of FOSTA and section 230. *First Amend. L. Rev.*, *17*, 279.

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.

Hoppner, T., Westerhoff, P. (2018). The EU's competition investigation into Amazon's Marketplace.

Organization for Economic Co-operation and Development – OECD. (2011). *The Role of Internet Intermediaries in Advancing Public Policy Objectives*. OECD Publishing. Paris.

Stemler, A. (2017). Feedback loop failure: Implications for the self-regulation of the sharing economy. *Minn. JL Sci. & Tech.*,18, 673.

54 Media Pluralism

Michael J. Oghia

At its core, media pluralism refers 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 organizations 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 marginalized 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. Recognizing how vital media plurality is and ultimately working to safeguard it is a critical endeavor going forward.

References

Council of Europe. (2018). Recommendation CM/Rec (2018)1[1] of the Committee of Ministers to Member States on Media Pluralism and Transparency of Media Ownership. Available at: <https://rm.coe.int/1680790e13>.

Frontieres, R. S. (2016). Contribution to the EU public consultation on media pluralism and democracy.

Global Forum for Media Development. (2020). *Joint Emergency Appeal for Journalism and Media Support*. Available at: <https://www.hirondelle.org/de/blog/1172-joint-emergency-appeal-for-journalism-and-media-support>.

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

Helberger, N., Kleinen-von Königslöw, K., Van Der Noll, R. (2015). *Regulating the new information intermediaries as gatekeepers of information diversity*.

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

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

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

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

Wikipedia contributors. (2021). Concentration of media ownership. In: *Wikipedia, The Free Encyclopedia*.

Websites:

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

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

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

Centre for Media Pluralism and Freedom – CMFP. Media Pluralism Monitor. <https://cmpf.eui.eu/media-pluralism-monitor.

GFMD. Internet Governance. Available at: <https://gfmd.info/internet-governance/>.

Media Diversity Institute. Available at: <https://www.media-diversity.org/>.

UNC. (2021). *Do You Live in a News Desert? The Expanding News Desert*. <https://www.usnewsdeserts.com/>.

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

55 Microtargeting

Paddy Leerssen

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 Dobber, Ó Fathaigh and Zuiderveen Borgesius, (2019) 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, Ó Fathaigh, Zuiderveen Borgesius, 2019). In this reading, targeting an entire neighborhood with a single message constitutes regular targeting, whereas tailoring different messages to different users within the neighborhood, based on their personal data profiles, constitutes microtargeting. Overall, the available literature suggests a sliding scale between general and micro-targeting, rather than a strict binary.

'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, Ó Fathaigh, 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 (2020), 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.

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 – due to its role in the spread of disinformation and manipulative content –, 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 IVIR, 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, T., Ó Fathaigh, R., Zuiderveen Borgesius, F. (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>.

Google. (2019). An update on our political ads policy. *Google Official Blog*. <https://blog.google/technology/ads/update-our-political-ads-policy/>.

Gorton, W. A. (2016). Manipulating citizens: How political campaigns' use of behavioral social science harms democracy. *New Political Science*, *38*(1), 61-80.

Leerssen, P. et al. (2018). Platform ad archives: Promises and pitfalls. *Internet Policy Review*, *8*(4). Available at: <https://policyreview.info/articles/analysis/platform-ad-archives-promises-and-pitfalls>.

Rubinstein, I. S. (2014). Voter privacy in the age of big data. *Wis. L. Rev.*, 861.

Tang, Paul. (2020). *European Parliament wants to forbid personalised advertisements*. Available at: <https://www.paultang.nl/en/forbid-personalised-ads/>.

Van Hoboken, J., Appelman, N. Ó Fathaigh, R., Leerssen, P., McGonagle, T., van Ejick, N., Helberger, N. (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. Available at: <https://www.ivir.nl/publicaties/download/Report_Disinformation_Dec2019-1.pdf>.

56 Moderation

Giovanni de Gregorio

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 behavior online, and to minimize aggression and anti-social behavior (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 (Elkin-Koren, 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 (Elkin-Koren, 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 prioritization, 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 recognized 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 minimizing 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, J. M. (2013). Old-school/new-school speech regulation. *Harv. L. Rev.*, 127, 2296.

Belli, L., Zingales, N. (2017). *Platform regulations: how platforms are regulated and how they regulate us.* Leeds. Available at: <https://bibliotecadigital.fgv.br/dspace/handle/10438/19402>.

Bridy, A. (2018). Remediating Social Media: A Layer-Conscious Approach. *BUJ Sci. & Tech. L.*, *24*, 193.

Bloch-Wehba, H. (2020). Automation in moderation. *Cornell Int'l LJ*, 53, 41. <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3521619>.

De Gregorio, G. (2020). Democratising online content moderation: A constitutional framework. *Computer Law & Security Review*, *36*.

De Gregorio, G. (2018). From constitutional freedoms to the power of the platforms: protecting fundamental rights online in the algorithmic society. *Eur. J. Legal Stud.*,*11*, 65.

Elkin-Koren, N., Perel, M. (2019a). Algorithmic Governance by Online Intermediaries. In Brousseau, E., et al. (eds). *Oxford Handbook of Institutions of International Economic Governance and Market Regulation.* (2019). Oxford University Press.

Elkin-Koren, N., & Perel, M. (2019b). Separation of functions for AI: Restraining speech regulation by online platforms. *Lewis & Clark L. Rev.*, *24*, 857. Available at: <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3439261.>

Flew, T., Martin, F., Suzor, N. (2019). Internet regulation as media policy: Rethinking the question of digital communication platform governance. *Journal of Digital Media & Policy*, 10(1), 33-50.

Gorwa, R., Binns, R., Katzenbach, C. (2020). Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society*, *7*(1).

Grimmelmann, J. (2015). The virtues of moderation. *Yale JL & Tech.*, 17, 42.

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

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

Keller, D. (2018). *Internet Platforms: Observations on Speech, Danger, and Money*. Hoover Institution. <https://www.hoover.org/sites/default/files/research/docs/keller_webreadypdf_final.pdf>.

Klonick, K. (2017). The new governors: The people, rules, and processes governing online speech. *Harv. L. Rev.*, *131*, 1598.

Langvardt, K. (2017). Regulating online content moderation. *Geo. LJ*, *106*, 1353.

Pasquale, F. (2016). Platform neutrality: Enhancing freedom of expression in spheres of private power. *Theoretical Inquiries in Law*, *17*(2), 487-513.

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

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

Tushnet, R. (2007). Power without responsibility: Intermediaries and the First Amendment. *Geo. Wash. L. Rev.*, *76*, 986.

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

57 Must-carry

Luã Fergus and Laila Lorenzon

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. 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, L. (2017). Net Neutrality, Zero rating and the Minitelisation of the Internet. *Journal of Cyber Policy*, 2(1), 96-122.

Pieranti, O., Festner, S. (2008). *Estudo comparativo de regras de must carry na TV por assinatura.* Agência Nacional de Telecomunicações, ANATEL.

Valente, J. (2018). *Regulação democrática dos meios de comunicação*. São Paulo: Fundação Perseu Abramo.

Patrick, A., Scharphorn, E. (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>.

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

58 Non-discrimination

Richard Wingfield

The term 'non-discrimination' has a very specific definition and understanding under international law (and international human rights law specifically), 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.

'Non-discrimination' under international (Human Rights) law

The principle of non-discrimination – and, specifically, the prohibition of discrimination – has been translated into several 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, color, 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, color, 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 prohibition of 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.

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 as:

> 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.

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. (1989). UN Human Rights Committee. General Comment No. 18: *Non-discrimination*.

European Commission. *Shaping Europe's digital future*. Glossary. Available at: <https://ec.europa.eu/digital-single-market/en/glossary>.

59 Notice

Luã Fergus and Laila Lorenzon

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 is 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* (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*, 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, as highlighted in *Ozormoor v. T- Mobile USA Inc*. (2008).

The appearance and placement of the notice also are 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* (2007). On the other hand, in Briceno Sprint Spectrum (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 (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 several provisions, including an arbitration clause. When Gateway moved to dismiss a class-action lawsuit considering 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* (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* (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 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* (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* (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 of 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:

- The legal basis for the assertion that the content is unlawful.
- The Internet identifier and description of the allegedly unlawful content.
- The consideration provided to limitations, exceptions, and defences available to the user content provider.
- Contact details of the issuing party or their agent, unless this is prohibited by law.

- Evidence sufficient to document legal standing to issue the request.
- A declaration of good faith that the information provided is accurate.

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

- The reasons why the content at issue is in breach of the intermediary's content restriction policies.
- The Internet identifier and description of the alleged violation of the content restriction policies.
- 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 complaints).

- The identity of individual flaggers should generally not be revealed, however, content flagged by the 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 a durable form accessible even if a user's account is suspended or terminated, and the presentation to users who flag the 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, 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. (2016). Online Contracts: We May Modify These Terms at Any Time, Right? *American Bar Association.* Available at: <https://www.americanbar.org/groups/business_law/publications/blt/2016/05/07_moringiello/>.

Electronic Frontier Foundation – EFF. *EFF Seeks Public Comment About Expanding and Improving Santa Clara Principles.* Available at: <https://www.eff.org/press/releases/eff-seeks-public-comment-about-expanding-and-improving-santa-clara-principles>.

Suzor, N. P., West, S. M., Quodling, A., York, J. (2019). What do we mean when we talk about transparency? Toward meaningful transparency in commercial content moderation. *International Journal of Communication*, *13*, 18.

Case Law:

Badie v. Bank of America, 67 Cal.App.4th 779, 79 Cal. Rptr. 2d 273 (Cal. Ct. App. 1998).

Briceño v. Sprint Spectrum, L.P., 911 So. 2d 176 (Fla. Dist. Ct. App. 2005).

Campbell v. General Dynamics, 407 F.3d 546 (1st Cir. 2005).

Directv v. Mattingly, 376 Md. 302, 829 A.2d 626 (Md. 2003).

Douglas v. U.S. District Court, 495 F.3d 1062 (9th Cir. 2007).

Klocek v. Gateway, Inc., 104 F. Supp. 2d 1332 (D. Kan. 2000).

Knutson v. Sirius XM Radio Inc., 771 F.3d 559 (9th Cir. 2014).

Manasher v. Telecom, No. 06-10749 (E.D. Mich. Sep. 18, 2007).

Ozormoor v. T- Mobile USA Inc., U.S. Dist. LEXIS 58725 (E.D. Mich. June 19, 2008).

Rodman v. Safeway, Inc., 2015 U.S. Dist. LEXIS 17523 (N.D. Cal. 2015).

Websites:

Manila Principle of Intermediary Liability. <https://www.manilaprinciples.org/principles>.

60 Notice-and-notice

Nicolo Zingales

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, A. Schirru, L. (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/>.

61 Notice-and-takedown

Luã Fergus and Laila Lorenzon

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 (Section § 512, DMCA). 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:

- offer serious risks to freedom of expression online, encouraging the arbitrary removal of content
- 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 the posted contents comply with copyright rules is an artificial intelligence tool (e.g., YouTube's Content ID) or not. 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.

de Souza, A., Schirru, L. (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>.

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/>.

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/>.

Websites:

Electronic Frontier Foundation – EFF. *Takedown Hall of Shame*. Available at: <https://www.eff.org/pt-br/takedowns>.

Graduated Response. About Graduated Response. Available at: <http://graduatedresponse.org/new/?page_id=5>.

62 Notice-and-staydown

Nicolo Zingales

'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). In the 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 rightsholders 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 rightsholders 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

Kuczerawy, A. (2020). From ‘Notice and Take Down’ to ‘Notice and Stay Down’: Risks and Safeguards for Freedom of Expression. In: Frosio, G. (2020). *The Oxford Handbook of Online Intermediary Liability*. Oxford University Press.

Husovec, M. (2018). The Promises of Algorithmic Copyright Enforcement: Takedown or Staydown: Which is Superior and why. *Colum. JL & Arts*, 42, 53.

Angelopoulos, C. (2020). Harmonising Intermediary Copyright Liability in the EU: A Summary. In: Frosio, G. (2020). *The Oxford Handbook of Online Intermediary Liability*. Oxford University Press. 315-334.

Frosio, G. F. (2017). Reforming intermediary liability in the platform economy: A European digital single market strategy*. Nw. UL Rev. Online*, 112, 18.

Case Law:

Eva Glawischnig-Piesczek v. Facebook (2019), Case C-18/18 (CJEU 2019).

63 Nudging

Nicolo Zingales

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*' (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 behavioral insights team) within the UK government in order to generate and apply behavioral 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 six different types of design interventions:

1. i[N]centives: 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.
2. 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).
3. 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.
4. Give feedback: sometimes simply informing whether something is going wrong (or well) helps people redirect.

5. 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.
6. 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 (2015) 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 endeavored 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 behavior 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 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 behavioral 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, 2017)
- 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 (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

Barnhill, A. What is Manipulation? (2014). In: Coons, C., Weber, M. (2014) *Manipulation: theory and practice.* Oxford University Press.

Baldwin, R. (2014). From regulation to behaviour change: giving nudge the third degree. *The Modern Law Review*, 77(6), 831-857.

Faden, R., Beauchamp. (2017). A History of Informed Consent (OUP, 1986). In: Yeung, K. (2017). 'Hypernudge': Big Data as a mode of regulation by design. *Information, Communication & Society*, 20(1), 118-136.

Hansen, P. G., Jespersen, A. M. (2013). 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, 4(1), 3-28.

Noggle, R. (1996). Manipulative actions: A conceptual and moral analysis. American Philosophical Quarterly, 33(1), 43-55.

Raz, J. The Morality of Freedom. Oxford: Oxford University Press 1986. *Canadian Journal of Philosophy*, 19(3), 477-490.

Sunstein, C. R. (2015). The Ethics of Nudging, 32, *Yale J. on Reg*., 413, 414. Available at: <http://digitalcommons.law.yale.edu/yjreg/vol32/iss2/6.

Sunstein, C. R. (2016). *The ethics of influence: Government in the age of behavioral science.* Cambridge University Press.

Thaler, R. H. (2015). The power of nudges, for good and bad. T*he New York Times*, 31, 2015. Available at <https://www.nytimes.com/2015/11/01/upshot/the-power-of-nudges-for-good-and-bad.html.

Thaler, R. H., Sunstein, C. R. *Nudge: Improving decisions about health, wealth, and happiness*.

Wilkinson, T. M. (2013). Nudging and manipulation. *Political Studies*, 61(2), 341-355.

64 Online Advertising

Catalina Goanta

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 'content creators/influencers').

References

Goldfarb, A., Tucker, C. (2011). Chapter 6-online advertising. *Advances in Computers*, *81*, 289-315.

Information Commissioner's Office. (2019). Update report into AdTech and real time bidding. 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>.

65 Open Identity

Vittorio Bertola

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:

I. *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;

II. *Authorisation*, i.e., the request and release of permission for an authenticated identity to access a specific resource or service;

III. *Signing*, i.e., the creation of cryptographic attestations of a certain assertion by the owner of the identity;

IV. *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

European Commission. (2014). 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. Available at: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv%3AOJ.L_.2014.257.01.0073.01.ENG>.

66 Open Standard

Vittorio Bertola

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:

I. *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; Weitzner, 2004);*

II. *Definition that requires 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

Dardailler, D. (2007). Definition of open standards. *World Wide Web Consortium*, 29. <https://www.w3.org/2005/09/dd-osd.html>.

Weitzner, D. J. (2004). Standards, patents and the dynamics of innovation on the world wide web. *Working paper MIT Computer Science and Artificial Intelligence Laboratory. <https://www.w3.org/Consortium/Patent-Policy-20170801/>*.

ITU. (2005). IPR Ad Hoc Group. *Definition of Open Standards*. <https://www.itu.int/en/ITU-T/ipr/Pages/open.aspx>.

Bradner, S. (2005). *Intellectual Property Rights in IETF Technology.* BCP 79, RFC 3979, March. <https://tools.ietf.org/html/rfc8179>.

European Commission. (2012). Regulation (EU) N° 1025/2012 of the European Parliament and of the Council of 25 October 2012. On European standardisation, amending Council Directives 89/686/EEC and 93/15/EEC and directives 94/9/EC, 94/25/EC, 95/16/EC, 97/23/EC, 98/34/EC, 2004/22/EC 2007/23/EC, 2009/23/EC and 2009/105/EC of the European parliament and of the council and repealing council decision 87/95/EEC and decision no 1673/2006/EC of the European Parliament and of the Council. Available at: <https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2012:316:0012:0033:EN:pdf>.

European Comission. (2017). Communication from the Commission to the European Parliament, the Council and the European Economic and Social Committee. Available at: <https://ec.europa.eu/docsroom/documents/26583/attachments/1/translations/en/renditions/native>.

67 Optimization

Courtney Radsch

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 e.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

Amodei et al. (2016). Available at: <https://arxiv.org/abs/1606.06565>.

Kulynich et al. (2018). Available at: <https://arxiv.org/abs/1806.02711>.

68 Platform

Luca Belli

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 characterization of the platform as a structure on top of which something – be it a product or a service – may be built and operated. Hence, platforms can be seen as the technical and governance structures (Belli, 2021) that facilitate relationships and exchange of value between different categories of users.

When a platform is qualified as ‘digital’ or ‘online’, it may refer to a vast array of software applications that are frequently loosely defined. Such platforms provide a governance structure, via their private ordering, and a technical architecture, via a wide range of standards, protocols, and algorithms, that facilitate user interactions and exchange of value at scale, thus unleashing network effects. As pointed out by the European Commission (2016b), the term is frequently utilized 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.)

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; Venturini, 2016; Belli; Sappa 2017; Belli; Zingales 2017).

References

Belli L. (2021). Structural Power as a Critical Element of Social Media Platforms' Private Sovereignty. In: Edoardo Celeste, Amélie Heldt and Clara Iglesias Keller (eds). *Constitutionalising Social Media*. Hart. (forthcoming)

Belli L., Venturini J. (2016). Private ordering and the rise of terms of service as cyber-regulation. *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. (2017). The Intermediary Conundrum: Cyber-regulators, Cyber-police or both? *JIPITEC*. Available at: <http://www.jipitec.eu/issues/jipitec-8-3-2017/4620>.

Belli, L., Zingales, N. (2017). *Platform regulations: how platforms are regulated and how they regulate us*. Leeds. Available at: <https://eprints.whiterose.ac.uk/150002/>.

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. (2016a). 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.

European Commission. (2016b). Online Platforms Accompanying the document Communication on Online Platforms and the Digital Single Market. Commission Staff Working Document {COM(2016) 288 final.

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

Tiwana, A. (2014). *Platform Ecosystems Aligning Architecture, Governance, and Strategy*.

IVIR (2015). Digital platforms: an analytical framework for identifying and evaluating policy options. Final report. Available at: <https://www.ivir.nl/publicaties/download/1703.pdf>.

69 Platform Governance

Paddy Leerssen

The concept of Governance refers to a ‘decentered’ 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, 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 behavior, 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 recognizable 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, Ranchordás, 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 behavior 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 criticize 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 Gorwa highlights, these complex multi-stakeholder interactions play out across various geographical scales, with overlapping "local, national, and supranational mechanisms of governance" (2018).

Importantly, while the term 'governance' is considered by most anglophone authors as synonymous with regulation, the equivalence between governance and regulation is not universally accepted beyond English-language literature. As noted by Belli (2016; 2019) the term governance has a procedural connotation, referring to set of processes and the mechanisms that stimulate the interaction and association of different stakeholders with the goal of discussing or elaborating regulation. In this perspective, platform governance should be seen as the set of processes allowing to discuss, chose and, ideally, elaborate the regulatory strategies that will be utilized to regulate platforms.

References

Belli, Luca. (2016). *De la gouvernance à la regulation de l'Internet*. Paris: Berger-Levrault

Belli, Luca. (2019). Internet Governance and Regulation: A Critical Presentation. In: Belli, Luca and Cavalli Olga. *Internet Governance and Regulations in Latin America.* FGV Direito Rio. Available at: <https://www.gobernanzainternet.org/book/book_en.pdf>.

Benkler, Y. (2011). A free irresponsible press: Wikileaks and the battle over the soul of the networked fourth estate. *Harv. CR-CLL Rev.*, 46, 311. Available at: <http://benkler.org/Benkler_Wikileaks_current.pdf>.

Bruns, A. (2019). After the 'APIcalypse': social media platforms and their fight against critical scholarly research. *Information, Communication & Society*, 22(11), 1544-1566 <https://eprints.qut.edu.au/131676/>.

Caplan, R., Gillespie, T. (2020). Tiered governance and demonetization: The shifting terms of labor and compensation in the platform economy. *Social Media+ Society*.

Goanta, C., Ranchordás, S. (2020). *The Regulation of Social Media Influencers*. Edward Elgar Publishing.

Gillespie, T. (2016). Governance of and by platforms. *SAGE handbook of social media*, 254-278. Available at: <https://culturedigitally.org/wp-content/uploads/2016/06/Gillespie-Governance-ofby-Platforms-PREPRINT.pdf>.

Gorwa, R. (2019). What is platform governance? *Information, Communication & Society*, 22(6), 854-871. Available at: <https://gorwa.co.uk/files/platformgovernance.pdf>.

Gorwa, R. (2019). The platform governance triangle: Conceptualising the informal regulation of online content. Internet Policy Review, 8(2), 1-22. Available at: <https://policyreview.info/articles/analysis/platform-governance-triangle-conceptualising-informal-regulation-online-content>.

Kreiss, D., McGregor, S. C. (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), 499-522.

Van Dijck, J., Poell, T., De Waal, M. (2018). *The platform society: Public values in a connective world*. Oxford University Press.

70 Platform Neutrality

Konstantinos Stylianou

‘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).

The concept of ‘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 – or search neutrality – (Frischmann, 2004). If platforms guarantee equal conditions to all complements, then the complements can compete on the merits.

The concept has also 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 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. (2014). Platform Neutrality: Building an Open and Sustainable Digital Environment. Opinion n˚. 2014-2.

Frischmann, B. M. (2004). An economic theory of infrastructure and commons management. *Minn. L. Rev*.,89, 917.

Renda, A. (2015). Antitrust, regulation and the neutrality trap: A plea for a smart, evidence-based internet policy. *CEPS Special Report*. 104.

Yoo, C. S. (2004). Would Mandating Broadband Network Neutrality Help or Hurt Competition-A Comment on the End-to-End Debate. J. *Telecomm. & High Tech. L*., 3, 23.

71 Pornography

Yasmin Curzi

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, 2012:19) 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: 22) with campaigns against pornography.

For MacKinnon (1987), for instance, pornography and the culture that portrays women as sex objects are responsible for the maintenance of sexual violence and sex discrimination – briefly, according to MacKinnon sexuality is expropriated from women by the male-dominated State, in the same way that, for the Marxists, labor is expropriated from workers by the capitalist State. Her work – which 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 social groups (e.g., poor, black and latina women).

Other feminists, however, fear that the fighting 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 movement contemplated liberal concerns about individual rights and choices and inspired 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.

72 Prioritization

Courtney Radsch

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.

73 Proactive Measures

Daphne Keller and Nicolo Zingales

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 that is 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 requires 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 appropriate terrorist material from their services, including by deploying 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

Barata, Barata. (2020). Positive Intent Protections: Incorporating a Good Samaritan principle in the EU Digital Services Act. Center for Democracy and Technology.

Kuczerawy, Aleksandra. (2018). The EU Commission on voluntary monitoring: Good Samaritan 2.0 or Good Samaritan 0.5? Ku Leuven. Available at: <https://www.law.kuleuven.be/citip/blog/the-eu-commission-on-voluntary-monitoring-good-samaritan-2-0-or-good-samaritan-0-5/>.

74 Recommender Systems

Rossana Ducato

'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 personalized 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 personalized, 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 behavior (Jannach et al., 2010). Most recommender systems rely on machine learning techniques, including deep neural networks (Goanta, 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 organization 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 defense 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, 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, organized 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

European Commission. (2020). Consultation Results. Ranking transparency guidelines in the framework of the EU regulation on platform-to-business relations – an explainer. Available at: <https://ec.europa.eu/digital-single-market/en/news/ranking-transparency-guidelines-framework-eu-regulation-platform-business-relations-explainer>.

Aggarwal, C. C. (2016). *Recommender systems* (Vol. 1). Cham: Springer International Publishing.

Busch, C. (2016). Crowdsourcing consumer confidence: How to regulate online rating and review systems in the collaborative economy. *European Contract Law and the Digital Single Market*. Intersentia, Cambridge.

Cobbe, J., Singh, J. (2019). Regulating recommending: motivations, considerations, and principles. *Considerations, and Principles*.

Dellarocas, Chrysanthos. (2003). The Digitisation of Word of Mouth: Promise and Challenges of Online Feedback Mechanisms. *Management Science* 49 (10): 1407–1424.

Ducato, R. (2020). Private Ordering of Online Platforms in Smart Urban Mobility: The Case of Uber's Rating System. In: *Smart Urban Mobility.* Springer, Berlin, Heidelberg. 301-323.

Felfernig, A., Boratto, L., Stettinger, M., Tkalčič, M. (2018). *Group recommender systems: An introduction.* Springer.

Goanta, C., Spanakis, G. (2020). *Influencers and Social Media Recommender Systems: Unfair Commercial Practices in EU and US Law.* Available at: <https://ssrn.com/abstract=3592000>.

Jannach, D. et al. (2010). *Recommender systems: an introduction*. Cambridge University Press.

Ranchordás, S. (2018). Online reputation and the regulation of information asymmetries in the platform economy. Critical Analysis of Law

Smorto, Guido. (2016). Reputazione, Fiducia e Mercati. Europa e diritto privato.

Twigg-Flesner, Christian. (2018). The EU's Proposals for Regulating B2B Relationships on Online Platforms – Transparency, Fairness and Beyond. EuCML.

75 Red Flag Knowledge

Nicolo Zingales

‘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 several 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 provider 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. YouTub*e, 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

Besek, J. M., Keiter, O. W. (2018). Capitol records vimeo: The peculiar case of pre-1972 sound recordings and federal copyright law. *Columbia Journal of Law & the Arts*, 41(4), 559-582.

Toto, Carolyn. (2016). When it comes to the DMCA, a Red Flag becomes harder to Fly. Pillsbury. *Internet and Social Media Law Blog.* Available at <https://www.internetandtechnologylaw.com/dmca-red-flag/>.

Terrica Carrington. (2018). Twenty Years of the DMCA: Notice and Takedown in Hindsight (Part II), Copyright Alliance Blog. Available at: <https://copyrightalliance.org/ca_post/twenty-years-dmca-notice-and-takedown/>.

US Copyright Office. (2020). Section 512 of title 17: A report of the register of copyrights. Available at <https://www.copyright.gov/policy/section512/section-512-full-report.pdf>.

Case Law:

Capitol Records, LLC v. Vimeo, LLC, 826 F.3d 78 (2d Cir. 2016).

UMG Recordings, Inc. v. Shelter Capital Partners LLC, 667 F.3d 1022 (9th Circuit, 2011).

Viacom International v. YouTube, (2nd Circuit Court of Appeals 2012).

76 Regulation

Roxana Radu

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". The promulgation of a binding set of rules designed to influence business or social behavior (Baldwin et al., 2012) is what sets regulation apart from other forms of governmental intervention. Back in the 19th century, John Stuart Mill referred to it as a governmental intervention "in the affairs of society" (1848) and that understanding has also prevailed in relation to the digital world, which brought new regulatory concerns.

The objective of regulation is to foster equilibrium and ensure the proper functioning of complex systems which may or may not originate within the state. In the digital ecosystem, the most utilized and effective forms of regulation are those of a private nature, such as contractual agreements (terms of service, privacy policies, etc.) and the *lex informatica* (Reidenberg, 1998) composed of software and hardware that define the architecture of the Internet (Lessig, 2006). The dominance of such private forms of regulation does not mean that the classic tools of public regulation – such as international conventions, laws, decrees, administrative regulations, and decisions made by national courts and agencies – lose their relevance. It only implies that this normative pluralism must be considered by regulators when choosing the most pertinent regulatory techniques and elaborating the most effective tools, ideally via open and participatory governance processes (Belli, 2016; 2019).

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 rulemaking, 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, Radu, 2014). They allowed the growth of Internet services, intermediaries, and platforms in the absence of strong public regulation frameworks.

An important tension in digital regulation has been the one between hard and soft law instruments, between what is legally codified and what is agreed informally in an attempt to influence behavior 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) at the domestic, regional, and international level.

Regulatory debates have long focused on the relationship between the regulator and the regulated entity. Over time, digital regulation 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 exerting influence over the design of public rules in its interest). At an operational level, regulation happens through sets of practices, and it is thus "collectively mediated and legitimized by the key communities whose buy-in is necessary" (Radu, 2019:25).

Over time, the Internet has seen a regulatory shift, away from the application of general telecommunications rules towards the creation of Internet-specific provisions starting in the 1990s and culminating in harmonized legislation in the European Union. The focus of regulation has also changed in recent years: 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.

Belli, Luca. (2016). *De la gouvernance à la regulation de l'Internet.* Paris: Berger-Levrault

Belli, Luca. (2019). Internet Governance and Regulation: A Critical Presentation. In: Belli, Luca and Cavalli Olga. *Internet Governance and Regulations in Latin America.* FGV Direito Rio. Available at: <https://www.gobernanzainternet.org/book/book_en.pdf>.

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. *The Independence of the Media and Its Regulatory Agencies: Shedding New Light on Formal and Actual Independence Against the National Context*. Polity Press. 15-54.

Kettemann, M. (2020). *The Normative Order of the Internet: A Theory of Rule and Regulation Online.* Oxford: Oxford University Press.

Lessig, L. (2006). *Code and Other Laws of Cyberspace. Version 2.0*. New York: Basic Books.

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.

Reidenberg, J. R. (1998). Lex Informatica: The Formulation of Information Policy Rules Through Technology. *Texas Law Review*. Vol. 76. N° 3.

77 Remedy

Chris Wiersma

A simple definition is given by legal dictionaries, emphasizing the element of 'recovery' or 'repair', thus referring to the end-result of a process described as an "effective grievance mechanism" (Le Docte, 2011). The term 'remedy' is formally embedded in many international treaties, e.g., 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 e.g., 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 recognized 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 favorable 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" (a.k.a. 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

Am Zehnhoff, H. W., Timmermans, H., Salmon, Y., Schmatz, E. (2011). *Le docte: viertalig juridisch woordenboek; Dictionnaire de termes juridiques en quatre langues; Rechtswörterbuch in vier Sprachen;* Legal dictionary in four languages. 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. (2019). DCPR 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>.

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.

UN, 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>.

UN. (2011). Guiding Principles on Business and Human Rights: United Nations “Protect, Respect and Remedy” Framework. HR/PUB/11/04. Available at: <https://www.ohchr.org/documents/publications/guidingprinciplesbusinesshr_en.pdf>.

78 Repeat Infringer

Luca Belli

The concept of 'Repeat infringement' has been established by the Digital Millennium Copyright Act (DMCA), 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, § 512 grants so-called safe harbor 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.

A well-known version of the repeat infringer mechanisms has been established by France with its "three strikes" or "graduated response" system. This system, created in 2009 by the HADOPI Law, is implemented by an administrative authority HADOPI (*Haute Autorité pour la diffusion des* œuvres *et la protection des droits sur Internet*), which acts on reports of suspected infringement from rightsholder groups. Based on the reports, HADOPI can contact, warn, and sanction any French Internet users. The system is based on the understanding that this graduated approach may persuade the infringers to change their behaviour.

Importantly, a study by Arnold et al. (2014) examining empirical data on first five years of implementation of the French three-strikes rule has found that the HADOPI system has not deterred individuals from engaging in digital piracy and that it did not reduce the intensity of illegal activity of those who did engage in piracy.

References

Arnold, Michael A. and Darmon, Eric and Dejean, Sylvain and Pénard, Thierry, Graduated Response Policy and the Behavior of Digital Pirates: Evidence from the French Three-Strike (Hadopi) Law (May 28, 2014). Available at: <http://dx.doi.org/10.2139/ssrn.2380522>.

Sawicki, A. (2006). Repeat Infringement in the Digital Millennium Copyright Act. University of Chicago Law Review: Vol. 73: Iss. 4, Article 7. Available at: <https://chicagounbound.uchicago.edu/cgi/viewcontent.cgi?article=5386&context=uclrev>.

79 Responsibility

Nicolo Zingales

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 (a.k.a. '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

Belli, L., Zingales, N. (2017). Online Platforms' Roles and Responsibilities: A Call for Action. in: Belli, L., Zingales, N. (2017). *Platform regulations: how platforms are regulated and how they regulate us*. Leeds, 21-32.

Center for Law & Democracy. (2016). Recommendations for Responsible Tech. Available at: <http://responsible-tech.org/wp-content/uploads/2016/06/Final-Recommendations.pdf>.

Council of Europe. (2017). Recommendation CM/Rec 2017 of the Committee of Ministers to member states on the roles and responsibilities of internet intermediaries. Available at: <https://rm.coe.int/recommendation-cm-rec-2017x-xx-of-the-committee-of-ministers-to-member/1680731980>.

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>.

UN/OHCHR. (2011). Guiding Principles on Business and Human Rights: United Nations "Protect, Respect and Remedy" Framework. HR/PUB/11/04. Available at: <https://www.ohchr.org/documents/publications/guidingprinciplesbusinesshr_en.pdf>.

UN/OHCHR. (2016). A/HRC/32/38. Report of the Special Rapporteur to the Human Rights Council on Freedom of expression, states, and the private sector in the digital age. Available at: <https://www.ohchr.org/en/issues/freedomopinion/pages/privatesectorinthedigitalage.aspx>.

Websites:

Stop Hate for Profit. Available at: <https://www.stophateforprofit.org/>.

80 Revenge Pornography/ Non-consensual Intimate Images

Richard Wingfield

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 appropriately describe the phenomenon and this term is 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 videos of an individual without their consent. Unlike 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). The literature on the subject (Franks; Waldman, 2018) has also drawn attention to the role of deep fake technologies on the manipulation of images of women and girls, with the end of creating 'deep nudes', i.e., images altered to 'involuntary fake porn'.

Legal definitions of the phenomenon, particularly those that criminalize 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

Several states and subnational jurisdictions governments have prohibited distributing non-consensual intimate images through the creation of criminal offenses. Databases of criminal laws in place can be found at, e.g., the Center for Internet and Society and the InternetLab (2018) websites. 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 criminalization, 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 authorization 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

Franks, M. A., Waldman, A. E. (2018). Sex, lies, and videotape: Deep fakes and free speech delusions. *Md. L. Rev.,* 78, 892.

Magaldi, J. A., Sales, J. S., Paul, J. (2020). Revenge Porn: The Name Doesn't Do Nonconsensual Pornography Justice and the Remedies Don't Offer the Victims Enough Justice. *Or. L. Rev.*, *98*, 197.

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. (2018). *How do countries fight the non-consensual dissemination of intimate images?* Available at: <https://www.internetlab.org.br/en/inequalities-and-identities/how-do-countries-fight-the-non-consensual-dissemination-of-intimate-images/>.

81 Right to Explanation

Nicolo Zingales

See also appeal.

The concept of ‘right to explanation’ refers to the informational duties owed by a data controller 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”. 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 analyze or predict aspects concerning that natural person’s performance at work, economic situation, health, personal preferences, interests, reliability, behavior, 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 were 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 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 explanation, 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 data, 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 based on 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 the minimum safeguards that should be offered when such decisions are made.

References

Edwards, L., Veale, M. (2017). Slave to the algorithm: Why a right to an explanation is probably not the remedy you are looking for. *Duke L. & Tech. Rev.*, *16*, 18.

Goodman, B., Flaxman, S. (2017). European Union regulations on algorithmic decision-making and a “right to explanation”. *AI magazine*, *38*(3), 50-57.

Malgieri, G., Comandé, G. (2017). Why a right to legibility of automated decision-making exists in the general data protection regulation. *International Data Privacy Law*.

Malgieri, G. (2019). Automated decision-making in the EU Member States: The right to explanation and other “suitable safeguards” in the national legislations. *Computer law & security review*, *35*(5), 105327.

Mendoza, I., Bygrave, L. A. (2017). The right not to be subject to automated decisions based on profiling. In *EU Internet Law*. Springer, Cham. 77-98.

Selbst, A., Powles, J. (2018, January). “Meaningful Information” and the Right to Explanation. In *Conference on Fairness, Accountability and Transparency*. PMLR. 48-48.

Wachter, S., Mittelstadt, B., Floridi, L. (2017). Why a right to explanation of automated decision-making does not exist in the general data protection regulation. *International Data Privacy Law*, *7*(2), 76-99.

82 Right To Be Forgotten

Stefan Kulk

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, 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

Ausloos, J. (2020). *The Right to Erasure in EU Data Protection Law: From Individual Rights to Effective Protection*. Oxford University Press.

Mayer-Schönberger, V. (2011). *Delete: The virtue of forgetting in the digital age*. Princeton University Press.

Kulk, S., Borgesius, F. Z. (2018). Privacy, freedom of expression, and the right to be forgotten in Europe. *Cambridge Handbook of Consumer Privacy*, 301.

Case Law:

Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and *Mario Costeja González, C-131/12* (CJEU 2014). Available at: <http://curia.europa.eu/juris/liste.jsf?num=C-131/12>.

83 Safe Harbor

Nicolo Zingales

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. In the context of data privacy law, the term "Safe Harbor" refers to a framework, in place since the year 2000, whereby US companies could certify that they respect basic principles of the EU data protection directive, which in turn enabled such companies to transfer data from the EU to the US. In 2015, the framework was invalidated by a ruling of the European Court of Justice in the *Schrems I* case (Case C-362/14).

The scope, limits, and conditions of safe harbors for digital intermediaries, are 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 defense 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.

84 Self-injury

Luã Fergus and Laila Lorenzon

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 in 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 amount 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., et al. (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. Available at: <https://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. (2017). Cyber-bullying and it its impact on mental health (2017). *European Journal of Public Health*. 27, 3. 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-atrás-da-inteligência-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.

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

85 Self-preferencing

Konstantinos Stylianou

'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, 2018).

References

Ahlborn, Christian. Leslie, Will. O'Reilly, Eoin. (2020). *Self-Preferencing: Between a Rock and a Hard Place*. CPI Antitrust Chronicle.

Crémer, J. et al. (2019). *Competition Policy for the Digital Era. Final Report for the European Commission*.

Google, Case AT.39740. (2017). *Google Search (Shopping)*.

Graef, Inge. (2019) Differentiated Treatment in Platform-to-Business Relations: EU Competition Law and Economic Dependence. *Yearbook of European Law.* 38, 448-499.

Zingales, Nicolo. (2018). *Google Shopping: Beware of 'Self-favoring' in a World of Algorithmic Nudging*. CPI Europe.

86 Self-regulation

Rolf H. Weber

'Self-regulation' refers to rules that are autonomously developed and implemented by the thereof concerned persons, groups or enterprises (such as users or users' groups, intermediaries, platform or technology providers, etc.), independently from any structured form of rulemaking (often called 'soft law'). 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). In such a case, the governance consists in a decentralized regulation by private actors.

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, 2017, mainly on page 41).

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 commonsense 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, Marsden, 2003: 2; Tambini, Leonardi, 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 remaining 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, (i.e., if the level of regulation by private actors does not seem to be adequate), new models have been developed and are available 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 rulemaking. In this field self-regulation is usually based on Codes of Conduct and Terms of Use (mostly phrased by the providers of the platforms). Therefore, attention must be paid to a potential disequilibrium between the rights and obligations of the contract parties. The European Law Institute (ELI) has published "ELI Model Rules on Online Platforms" (March 2020), which envisage to provide a fair allocation of rights and obligations (available at: <https://www.europeanlawinstitute.eu/>; Busch, Dannemann, Schulte-Nölke, Wiewiorowska-Domagalska, Zoll, 2020:61-67).

References

Abbott Kenneth W. and Duncan Snidal (2000). 'Hard and Soft Law in International Governance', in *International Organization* 54, 421-456. Available at <https://www.cambridge.org/core/journals/international-organisation>.

Belli, L., Zingales, N. (2017). Platform Regulations. How platforms are regulated and how they regulate us. Leeds.

Black, Julia. (1996). Constitutionalizing Self-Regulation, *The Modern Law Review* 59. 24-55. Available at: <https://www.modernlawreview.co.uk>.

Brown, Ian and Marsden, Christopher T. (2013*). Regulating Code: Good Governance and Better Regulation in the Information Age*, Cambridge MA/London.

Busch, C., Zoll, F. (2020). An introduction to the ELI model rules on online platforms. Journal of European Consumer and Market Law, 9(2).

Gibbons, L. J. (1996). No regulation, government regulation, or self-regulation: social enforcement or social contracting for governance in cyberspace. Cornell JL & Pub. Pol'y, 6, 475. Available at: <https://scholarship.law.cornell.edu/cjlpp>.

Guzman, A. T., Meyer, T. L. (2010). International soft law. Journal of Legal Analysis, 2(1), 171-225. Available at: <https://academic.oup.com/jla>.

Tambini, D., Leonardi, D., Marsden, C. (2008). *Clarifying Cyberspace: Communications self-regulation in the age of Internet convergence*, London.

Weber Rolf H. (2002). Regulatory Models for the Online World, Zurich.

Weber Rolf H. (2012), Overcoming the Hard Law/Soft Law Dichotomy in Times of (Financial) Crisis, *Journal of Governance and Regulation* 1, 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.

87 Shadowban/Shadow Banning

Courtney Radsch

‘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 the platform 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 ‘shadowbanned’ 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 (2015) used 'shadowbanning' 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

Bokhari, Allum. (2018). Facebook Admits to Shadowbanning News It Considers 'Fake" (BreitBart). Available at: <https://www.breitbart.com/tech/2018/07/13/facebook-admits-to-shadowbanning-news-it-considers-fake/>.

Grothaus, Michael. (2019). Here's how to shadow ban your Instagram bullies with the new 'Restrict' feature'. *Fast Company.* Available at: <https://www.fastcompany.com/90412178/instagram-rolls-out-restrict-feature-allowing-users-to-shadow-ban-bullies>.

MILO. (2016). Twitter Shadowbanning is real and happening every day says inside source. *BreitBart*. Availble at: <https://www.breitbart.com/tech/2016/02/16/exclusive-twitter-shadowbanning-is-real-say-inside-sources/>.

Molina, Brett. (2018). Shadow banning: What is it, and why is Trump talking about in on Twitter. *USA Today*. <https://www.usatoday.com/story/tech/nation-now/2018/07/26/shadow-banning-what-and-why-trump-talking-twitter/842368002/>.

Reddit. *On shadowbans.* Available at: <https://www.reddit.com/r/self/comments/3ey0fv/on_shadowbans/>.

Shu, Catherine. (2015). Reddit Replaces its Confusing Shadowban System with Account Suspensions. *Tech Crunch.* Available at: <https://techcrunch.com/2015/11/11/reddit-account-suspensions/>.

Stack, Liam. (2018). What Is a 'Shadow Ban', and Is Twitter Doing It to Republican Accounts? *The New York Times.* <https://www.nytimes.com/2018/07/26/us/politics/twitter-shadowbanning.html>.

Thompson, Alex. (2018). Twitter appears to have fixed "shadow ban" of prominent Republicans like the RNC chair and Trump Jr.'s spokesman. *Vice.* Available at: <https://www.vice.com/en/article/43paqq/twitter-is-shadow-banning-prominent-republicans-like-the-rnc-chair-and-trump-jrs-spokesman>.

88 Sharing Economy

Enguerrand Marique and Yasmin Curzi

See also marketplace.

Sharing economy is defined by the Oxford Dictionary as "an economic system in which assets or services are shared between private individuals, either for free or for a fee, typically by means of the internet". It is also referred to as 'collaborative', 'platform', or 'gig' economy. In spite of having common aspects, 'sharing economy' and 'gig economy' often refer to different activities: the former has at its core the idea of sharing, acquiring, or providing goods and services through an online platform, while the latter refers to flexible/temporary jobs.

Thus, the concept of sharing economy is based on the idea that certain economical activities can be changed towards a common social bond, especially with the decentralization enabled by internet access and connectivity. As Calo and Rosenblat point out, "instead of hailing taxis or booking hotel rooms, today's consumers can download an app or visit a website to connect with individuals willing to provide access to their private cars or homes. The sharing economy, of course, did not emerge spontaneously" (2017: 1625).

The societal adhesion to these apps is usually fed by a meritocratic idealization of its outcomes for the 'entrepreneurs'. Nevertheless, as Calo and Rosenblat (2017) also underline, inequalities are increased and highlighted by these apps – the individuals who usually have success in the 'sharing economy' already have properties or capital for investments. In this sense, Calo and Rosenblat (2017) propose shifting the concept to 'taking economy'.

Similarly, Finck and Ranchordás (2016) distinguish the so-called sharing economy with "genuine sharing practices", (2016:15) driven by solidarity, in several cities such as Medellín, Colombia. They view four features characterizing the sharing economy: a technology that enables multiple and simultaneous transactions, a sociological or cultural element of a sharing mentality, the existence of trust among participants, and urban dynamics. by do not always generate good outcomes for the consumers or entrepreneurs in the sharing economy.

Even in the presence of these elements, though, does not guarantee good outcomes for the consumers or entrepreneurs in the sharing economy: for instance, Uber drivers and AirBnB hosts are subject to extreme control by these companies, without having labor rights accordingly or any possibility of symmetry in their relationship with the companies. Beyond labor rights, Calo and Rosenblat (2017) also point out that regulators are often challenged with questions about the true nature of these companies and, therefore, their liability as "intermediaries of services". More generally, it is apparent that, due to its innovative nature, the economy of the digital platforms constantly defies traditional regulation (Baptista, Keller, 2016; Ranchordás, 2015; Edelmann; Geradin, 2016).

Reference

Baptista, P., Keller, C. I. (2016). Por que, quando e como regular as novas tecnologias? Os desafios trazidos pelas inovações disruptivas. Revista de Direito Administrativo, 273, 123-163.

Calo, R., Rosenblat, A. (2017). The taking economy: Uber, information, and power. *Colum. L. Rev.*, *117*, 1623.

Benjamin G. Edelman; Damien Geradin (2016), Efficiencies and Regulatory Shortcuts: How should we Regulate companies like Airbnb and Uber?. *Stanford Technology Law Review*.

Finck, M., & Ranchordás, S. (2016). Sharing and the City. *Vand. J. Transnat'l L.*, 49, 1299.

Ranchordás, S. (2015). Does sharing mean caring: Regulating innovation in the sharing economy. *Minn. JL Sci. & Tech.*, *16*, 413

89 Social Network

Enguerrand Marique

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'.

90 Spam

Luca Belli

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, utilizes 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 'email' 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*. 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>.

Internet Society. (2018). *What is spam?* Available at: <https://www.internetsociety.org/wp-content/uploads/2017/08/What20Is20Spam_0.pdf>.

91 Technological Protection Measures

Nicolo Zingales

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 rightsholders 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 rightsholders, including the conclusion and implementation of agreements between rightsholders and other parties.

References

Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonization of certain aspects of copyright and related rights in the information society. OJ L 167. 22/06/2001.

Gasser, U. (2006). Legal frameworks and technological protection of digital content: Moving forward towards a best practice model. *Fordham Intell. Prop. Media & Ent. LJ*. 17, 39.

Gasser, U., Girsberger, M. (2004). *Transposing the Copyright Directive: Legal Protection of Technological Measures in EU-Member States – A Genie Stuck in the Bottle?.* Available at <http://cyber.law.harvard.edu/media/files/eucd.pdf>.

Zingales, N. (2015). Of coffee pods, videogames, and missed interoperability: Reflections for EU governance of the internet of things. *TILEC Discussion Paper.* Available at: <https://ssrn.com/abstract=2707570 or <http://dx.doi.org/10.2139/ssrn.2707570>.

92 Terms of Service

Luca Belli

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 or to specific categories of users (e.g. commercial, non-commercial, or individual users, etc.). 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 – also known as "boilerplate contracts" (Radin 2012) – as the user merely adhere to the contractual clauses.

Indeed, 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 express the one-sided control of a single party, thus representing a form of "quasi-normative" power, expression of the "private sovereignty" of the party that unilaterally defines them. (Belli, 2016 & 2021; Cantero Gamito, 2017) 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 o' '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 (s., e.g., 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' utilized 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.

Platform providers may have ample contractual autonomy to define the contractual provisions that regulate their terms and, importantly, the scope and breath of terms of service's regulatory function can only limited by the existence of laws and regulation striking a balance between the protection of users' rights and the contractual autonomy of the intermediaries. (Belli & Venturini 2017) Terms of service are therefore the preeminent legal instrument of the private ordering that regulates each platform. Hence, in the absence of comprehensive fundamental rights protection and consumer protections, private actors providing any kind of digital service may abuse their central role in the "contract governance" (Grundmann et al. 2015) of their systems, contractually regulate their services in a way that may be seen as the most economically efficient, but may not be the most user-interest-oriented.

In this context, an analysis of the terms of service of 50 amongst the most popular online platforms, conducted by the Center for Technology and Society at FGV in partnership with the Council of Europe, assessed the compatibility of such documents with international human rights standards on to freedom of expression, privacy, and due process, and revealed a considerable imbalance of power between platforms and users. (Venturini et al. 2016) While corroborating the fundamental regulatory role of terms of service in shaping the platform's private ordering, the abovementioned study also demonstrates that platform choices, while aiming at achieving the most efficient outcome for the platform provider, may frequently neglect the full protection of users' rights.

References

Belli, L. (2016). De la Gouvernance à la régulation de l'Internet. Berger-Levrault. 202-208.

Belli, L. (2022). Structural Power as a Critical Element of Social Media Platforms' Private Sovereignty. In Celeste, E., Heldt, A., and Iglesias Keller, C. (Eds). Constitutionalising Social Media. Hart. <http://bit.ly/BelliPrivateSovereignty>.

Belli, L., Venturini, J. (2016). Private ordering and the rise of terms of service as cyber-regulation. *Internet Policy Review*, *5*(4). 1-17. Available at: <https://policyreview.info/articles/analysis/private-ordering-and-rise-terms-service-cyber-regulation>.

Belli, L., Sappa, C. (2017). The intermediary conundrum: cyber-regulators, cyber-police or both. *J. Intell. Prop. Info. Tech. & Elec. Com. L.*, 8, 183. Available at: <http://www.jipitec.eu/issues/jipitec-8-3-2017/4620>.

Belli, L., Zingales, N. (2017). *Platform regulations: how platforms are regulated and how they regulate us*. FGV Direito Rio.

Cantero Gamito, M. (2017). Regulation.com: self-regulation and contract governance in the platform economy: a research agenda. European journal of legal studies, Vol. 9, No. 2, pp. 53-68. <https://cadmus.eui.eu/handle/1814/46068>.

Ghosh, S. (2014). Against contractual authoritarianism. *Sw. L. Rev.*, *44*, 239.

Grundmann, S., Möslein F., and Riesenhuber K. (eds). (2015). Contract Governance: Dimensions in Law and Interdisciplinary Research. OUP.

Kim, N. S. (2013). *Wrap contracts: Foundations and ramifications*. Oxford University Press.

Prausnitz, O., Chorley, R. S. T. C. B. (1937). *The standardization of commercial contracts in English and continental law*. Sweet & Maxwell, Limited.

Radin, M. J. (2012). *Boilerplate*. Princeton University Press.

Venturini, J., Louzada, L., Ferreira Maciel M., Zingales, N., Stylianou, K., Belli, L. (2016). Terms of service and human rights: An analysis of online platform contracts. FGV Direito Rio – Revan Editora – Council of Europe. <http://bibliotecadigital.fgv.br/dspace/handle/10438/18231>.

93 Terrorist Content/Cyberterrorism

Ivar Hartmann

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 radicalization language (Fernandez, Alani, 2021). 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, K. (2008). Hacktivism and politically motivated computer crime. Encurve LLC–Building Intentional Security. 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, M., Khawaja, M., Lakhani, S., Reffin, J., Robertson, A., & Weir, D. (2019). Disrupting Daesh: Measuring takedown of online terrorist material and its impacts. *Studies in Conflict & Terrorism*, *42*(1-2), 141-160.

Fernandez, M., Alani, H. (2021). Artificial intelligence and online extremism: Challenges and opportunities. In: McDaniel, J., Pease, K. (2021). Predictive Policing and Artificial Intelligence. Routledge.

Ganesh, B. (2018). The ungovernability of digital hate culture. *Journal of International Affairs*, *71*(2), 30-49.

Goodman, A. E. J. (2018). When you give a terrorist a twitter: holding social media companies liable for their support of terrorism. *Pepp. L. Rev.*, 46, 147.

Sandoval-Almazan, R., Gil-Garcia, J. R. (2014). Towards cyberactivism 2.0? Understanding the use of social media and other information technologies for political activism and social movements. *Government Information Quarterly*, 31(3), 365-378.

Veerasamy, N. (2020). Cyberterrorism–the spectre that is the convergence of the physical and virtual worlds. In *Emerging Cyber Threats and Cognitive Vulnerabilities*. Academic Press. 27-52.

Weimann, G. (2010). Terror on Facebook, twitter, and YouTube. *The Brown Journal of World Affairs*, *16*(2), 45-54

94 Transparency

Richard Wingfield

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, several 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.

95 User

Luca Belli

A platform user is any individual or entity utilizing the services provided by a specific platform. 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 to define the platform governance, either via its terms of service or via its technical architecture. (Lessig 2006; Belli, 2016 & 2022)

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 utilizing or acquiring goods and services for business purposes. (UNCTAD, 2017:6)

Business users are business enterprises that utilize 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

Belli, L. (2016). De la Gouvernance à la régulation de l'Internet. Berger-Levrault. 202-208.

Belli, L. (2022). Structural Power as a Critical Element of Social Media Platforms' Private Sovereignty. In Celeste, E., Heldt, A., and Iglesias Keller, C. (Eds). Constitutionalising Social Media. Hart. <http://bit.ly/BelliPrivateSovereignty>.

European Commission. (2018). Proposal for a Regulation of the European Parliament and of the Council on promoting fairness and transparency for business users of online intermediation services, COM(2018) 238 final, Brussels, 26.4.2018. Available at: <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52018PC0238&from=en>.

Crémer, J., de Montjoye, Y. A., & Schweitzer, H. (2019). Competition Policy for the digital era. European Commission. *Publications Office of the European Union*.

Lessig L. (2006).Code: And Other Laws of Cyberspace. Version 2.0. Basic Books

United Nations Conference on Trade and Development – UNCTAD. (2017). *Manual on Consumer Protection.* Available at: <https://unctad.org/en/PublicationsLibrary/ditccplp2017d1_en.pdf>.

96 User-generated Content

Cynthia Khoo

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

Krum et al. (2008). *User-Generated Content*. IEE CS Guest Editors Introduction.

97 Utility

Luã Fergus and Laila Lorenzon

According to McGregor Jr. et al. (1982), the term ‘utility’ refers to any service provided to people, either directly (by public sector institutions) or indirectly (by public or private companies). The term is associated with a social consensus, usually expressed in the legislation, that certain services must be available to everyone, regardless of income, physical ability, and other individual characteristics. Even when public utilities are neither provided nor funded by the public sector – for social-political reasons – they are, in general, subject to regulation because of the relevance of the service for the public wellbeing or interest. When done in the public interest and motivations, public policies can also provide public services (Anderfuhren-Biget et al., 2014).

A public utility can sometimes have the characteristics of a public good (being non-rival and non-excludable). Still, most are services that can (according to current social norms) be sub-supplied by the market. In most cases, public utilities are services that do not involve the manufacture of goods. They can be provided by local or national monopolies, especially natural monopolies.

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. Regarding the Internet-related debates, most of them focus on whether broadband is a public utility, something that emerged from discussions on the 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, deserves to be subject to stricter regulation. Under this classification, net neutrality rules, for example, could be enforced more effectively (Mosendz, 2014). On the other hand, some find it problematic to classify broadband as a public utility since it is a service with a competitive market, and stronger regulations could damage innovation and create a barrier to entry (Downes, 2016).

There is also a debate about the possibility of seeing platforms as public utilities, considering their infrastructural implications on the informational flow of communications (Rahman, 2018). Hoffmann and Rieder (2020), for example, suggest that platform regulation should consider a broader concept of 'public interest' as a normative reference for evaluating the behavior of digital platforms and, therefore, its regulation. On the other hand, there are authors pointing the dilemma of social protection and innovation/competition with such regulation (e.g., Thierer, 2012).

References

Anderfuhren-Biget, S., Varone, F., Giauque, D. (2014). Policy Environment and Public Service Motivation. London. *Public Administration*. 92(4): 807-825.

Downes, Larry (2016). Why treating the Internet as a public utility is bad for consumers? *Washington Post.* 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, E. B., Campbell, A. K., Macy Jr, J. W., Cleveland, H. (1982). Symposium: The public service as institution. *Public Administration Review*, *42*(4), 304.

Mosendz, Polly (2014). Is Broadband Internet a Public Utility? *The Atlantic.* Available at: <https://www.theatlantic.com/technology/archive/2014/05/is-broadband-internet-a-public-utility/362093/>.

Rahman, K. S. (2018). Regulating informational infrastructure: Internet platforms as the new public utilities. *Georgetown Law and Technology Review*, *2*, 2.

Rieder, B., Hofmann, J. (2020). Towards platform observability. *Internet Policy Review*, *9*(4), 1-28.

Thierer, A. (2012). *The perils of classifying social media platforms as public utilities*. CommLaw Conspectus.

98 Violence (Cyber)

Monika Zaineriute

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).

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’ (Šimonović, 2018). 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). Fifthly, information and communication technology enable 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).

References

Avni, M. (2018). Incitement to Terror and Freedom of Speech. In: *Incitement to Terrorism*. Leiden, The Netherlands: Brill | Nijhoff. doi: <https://doi.org/10.1163/9789004359826_005>.

Blaya, Catherine. Kaur, Kirandeep. Sandhu, Damanjit. (2018). Cyberviolence and Cyberbullying in Europe and India – A Literature Review. In: Smith, Peter K., Sundaram, Suresh., Spars, Barbara A., Blaya, C., Sandhu, D. *Bullying, Cyberbullying and Student Well-Being in Schools – Comparing European, Australian and Indian Perspectives.* 83–106. Cambridge University Press.

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. (2018). Working Group on cyberbullying and other forms of online violence, especially against women and children (CBG). *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>.

Council of Europe. (2019). Cyberviolence. *Cybercrime.* Available at: <https://www.coe.int/en/web/cybercrime/cyberviolence>.

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>.

Šimonović, Dubravka. (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>.

99 User Warning

Luca Belli

In the context of platform governance, 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 utilized to convey an alert and afford a user the possibility to form an informed choice before proceeding to an activity that may have unwanted and potentially negative consequences. While warnings can be used to raise awareness and propose options to one or more users, they are generally deployed to alert them of upcoming events that may lead to dangers or other unpleasant situations within a particular context.

A well-known but non-Internet-related example of warning that illustrate tellingly the purpose of a notice is provided by the so-called Miranda warnings, also frequently referred to as 'Miranda rights', following the 1966 Miranda v. Arizona case. Such workings are a type of notification given by police to criminal suspects in police custody advising them that they have the right to refuse to answer questions or provide information to law enforcement as 'anything you say can be used against you in court.'

Warnings, in the form of contractual notices, are an interesting regulatory mechanisms due to their very limited cost of implementation although they have limits in terms of effectiveness (Ben-Shahar & Schneider, 2011) and may even create additional costs for users, without achieving the regulatory goal, as illustrated by the implementation of cookie banners. (Privacy International, 2019) In this perspective, Calo (2012) notes that regulators, who face limited resources, perceive notice as an attractive regulatory strategy, as it is particularly cheap to implement and easy to enforce; by providing information on their rights and on the service characteristics to users, providers avoid that the regulatory authority 'overregulate' on legitimate business interests. As such, the goal of the warning in the form of a notice is to preserve the conditions for innovation and competition, avoiding an excess of rigid regulatory restrictions. (Calo, 2012)

References

Calo, R. (2012). Against Notice Skepticism in Privacy (And Elsewhere). 87 Notre Dame Law Review 1027. <https://ssrn.com/abstract=1790144>.

Ben-Shahar, O., and Schneider, C. E. (2011). The Failure of Mandated Disclosure, 159 U. PA. L. REV. 647, 658–64.

Privacy International. (21 may 2019). Most cookie banners are annoying and deceptive. This is not consent. <https://privacyinternational.org/explainer/2975/most-cookie-banners-are-annoying-and-deceptive-not-consent>.

100 Wilful Blindness

Nicolo Zingales

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 (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. (2020). *Section 512 of title 17: A report of the register of copyrights*. Available at: <https://www.copyright.gov/policy/section512/section-512-full-report.pdf>.

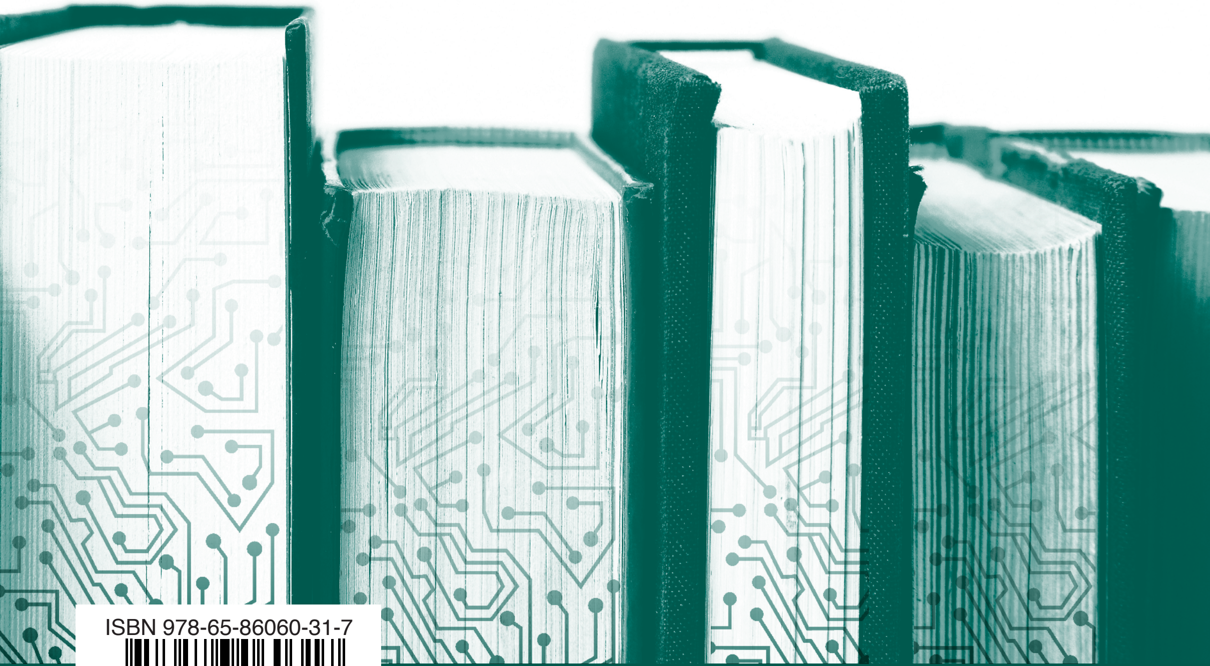
Case Law:

Viacom Int'l, Inc. v. YouTube, Inc., 676 F.3d 19 (2d Cir. 2012).

This Glossary was prepared by the "Glossary Working Group" of the IGF Coalition on Platform Responsibility, a multistakeholder group under the auspices of the UN Internet Governance Forum. The Glossary is a living document and the opinions expressed in the chapters of this volume are the sole responsibility of the authors and do not represent the position of the institutions that support this publication. The members of the working group are (in alphabetical order): Luca Belli, Vittorio Bertola, Yasmin Curzi de Mendonça, Giovanni De Gregorio, Rossana Ducato, Luã Fergus Oliveira da Cruz, Catalina Goanta, Tamara Gojkovic, Terri Harel, Cynthia Khoo, Stefan Kulk, Paddy Leerssen, Laila Neves Lorenzon, Chris Marsden, Enguerrand Marique, Michael Oghia, Milica Pesic, Courtney Radsch, Roxana Radu, Konstantinos Stylianou, Rolf H. Weber, Chris Wiersma, Richard Wingfield, Monika Zalnieriute and Nicolo Zingales.

The need for a **Glossary of Platform Law and Policy Terms** emerged at the 2019 meeditng of the IGF Coalition on Platform Responsibility, to foster a common language for academics, regulators and policymakers when discussing issues of platform responsibility. In this perspective, the elaboration of this Glossary aims at providing the conceptual basis on which legal interoperability between different systems framing platform governance can be built. Through this Glossary, we aim at offering guidance on what specific platform-related concepts mean, so that different stakeholders and, particularly, policymakers may have a better understanding of such a complex set of issues, thus elaborating well-informed and, ideally, good-quality and compatible platform policies and regulations.

Importantly, this Glossary does not aim at being prescriptive, but rather at recognizing that a specific concept may have various meanings and such differences and nuances should be acknowledged and highlighted to allow stakeholders to have a more complete understanding of each issue. This Glossary was presented the 2021 United Nations Internet Governance Forum but should be considered as a "living document." As such, the Glossary can be updated over time, including new contributions from a heterogeneous range of disciplines, stakeholder perspectives and vocabularies.



ISBN 978-65-86060-31-7
9 786586 060317