CELE

Platform Oversight: A Neglected Link in Internet's Regulatory Futures

October 2022

Facultad de Derecho
Centro de Estudios en Libertad
de Expresión y Acceso a la Información

UP
Universidad
de Palermo

Platform Oversight: A Neglected Link in Internet's Regulatory Futures

Agustina Del Campo*

CELE

1. Introduction

Since the mid-2000s, there have been multiple and wide-ranging proposals for traditional regulation[1], co-regulation,[2] and self-regulation[3] of digital platforms and Internet companies. The proliferation of regulatory proposals, framework agreements between states and companies, and unilateral commitments by multiple companies have generated a system of rules of a different legal nature that impact the functioning and conduct of the Internet intermediary ecosystem. [4]

Still, while substantive standards and transparency mandates have been at the center stage in regulatory debates, oversight has been largely ignored by both

* Agustina Del Campo directs the Center for Studies on Freedom of Expression and Access to Information CELE at Universidad de Palermo and serves as vice chair to the Board of the Global Network Initiative since September 2021. This document is the product of the research that started in 2020 at CELE and concluded in July 2021. The author would like to acknowledge Belen Gracia Andia and Emiliano Vitaliani, research assistants in this project. I would also like to thank colleagues and friends Ramiro Alvarez Ugarte, Amalia Toledo, Carolina Botero, Maria Paz Canales, Eduardo Bertoni, Jessica Field, Molly Land, Jason Pielemier, and David Kaye for their invaluable comments and feedback on previous drafts and for nurturing my thinking about these issues over time.

1 Referring to state regulation, characterized by the use of legal rules backed by the jurisdictional power of the state. See: Black, Julia. Decentring regulation: Understanding the role of regulation and self-regulation in a 'post-regulatory' world, Current legal problems 54.1 (2001): 103-146 and Black, Julia, Understanding the Role of Regulation and Self Regulation in a "Post-regulatory World, 54 (2001): 103-46. Marsden, Meyer & Brown explain the distinction between regulation, co-regulation, and self-regulation. We follow this distinction in this paper. See: Platform values and democratic elections: How can the law regulate digital disinformation, Computer Law and Security Review (2020). Article 19 defined these terms: regulation: established by law and characterized by more state interference; co-regulation: defined as regulated self-regulation or self-regulatory mechanisms endorsed or promoted by law; and self-regulation: a model that depends entirely on the will of the parties. Self-regulation and hate speech on social media platforms, page 9.

2 Understood as the regulatory phenomenon whereby the regulatory regime is made up of a complex interaction between general legislation and self-regulatory bodies; thus, for example, legislation can entrust the objectives defined by the legislative authority to recognized experts, such as economic operators, non-governmental organizations, etc. See: C. Marsden, Internet Co-Regulation, Cambridge University Press (2011) 46; Finck, M., Digital Regulation: Designing a Supranational Legal Framework for the Platform Economy, LSE Law, Society and Economy Working Papers 15/2017 (2017) pp. 15-29.

3 "Referring to the situation of a group of people or organisms acting together, performing a regulatory function concerning themselves and others who accept their authority. Self-regulation can be imposed by public authorities or voluntarily adopted by economic operators." See: J. Black, 'Constitutionalising Self-Regulation' (1996) 59 Modern Law Review 24, 27.

4 See Gorwa, R. (2019). The platform governance triangle: conceptualising the informal regulation of online content. Internet Policy Review, 8(2).

governments and commentators. Supervising the compliance or implementation of the proposed rules remains challenging and is usually an afterthought within regulatory proposals. This is an issue to be resolved after agreeing upon the substantive standards.[5] I argue that oversight should be at the center stage of these debates. My reasons are threefold: 1) it forces us to think and clearly state the objectives of the regulation (what we want to see happening and why); 2) it allows us to test the means to our ends; 3) it helps clarify the trade-offs that the substantive regulation proposes.

The desirable way to oversee the effective compliance of human rights by Internet companies remains unclear for civil society, academia, and states. While various documents establish that companies should comply with human rights standards,[6] it is still not evident who should verify such compliance or what the sanctions or responsibilities (if any) should there be for non-compliance, or who should implement them.[7] UN Special Rapporteur D. Kaye's 2018 report, for example, establishes that states have obligations to respect and guarantee human rights, including freedom of expression, even involving companies.[8] In other words, states should refrain from violating human rights online and investigate and punish them when those occur. States should also take preventive actions.[9] However, there seems to be more consensuses on what the responsibility to respect entails than what the obligations to guarantee would look like in an interjurisdictional ecosystem with competing legal frameworks and human rights interpretations at play.[10] For example, there seems to be a consensus that states should refrain from imposing civil liability on companies for third-party content or that they should refrain from forcing companies to block content on

---

5 Special Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Report on Content Regulation, Human Rights Council, UN, 38th session, (2018): Parr. 1: "Despite taking steps to illuminate their rules and government interactions, the companies remain enigmatic regulators, establishing a kind of "platform law" in which clarity, consistency, accountability, and remedy are elusive."

6 Special Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Report on Content Regulation, Human Rights Council, UN, 38th session, (2018).

7 See CELE, Human Rights Impact Assessments: Trends, Challenges, and Opportunities for ICT Sector Adoption (2020) and CELE, ICT and Human Rights: Towards a Conceptual Framework of Human Rights Impact Assessments (2020).

8 Special Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Report on Content Regulation, Human Rights Council, UN, 38th session, (2018). See also Bertoni, Responsabilidad de actores no estatales por violaciones a la libertad de expresión [Liability of non-state actors for violations of freedom of expression], Observacom (2021).

9 Id.

10 I&J Policy Network, Toolkit for Cross border content moderation, retrieved from: https://www.Internetjurisdiction.net/content/toolkit.

vague, ambiguous or abusive terms.[11] But the path for the state to effectively guarantee human rights on the Internet does not seem so clear. What measures should be taken and what should indicators look like? What oversight mechanisms should be created? Who should be in charge of them? What should their scope and subject matter competence be? What tools could be used or would be desirable to exercise such oversight?

In recent years, there have been different answers to the questions above. Broader and narrower oversight structures, with different degrees of complexity, private, public and mixed, multi-sector, and single-sector structures have been designed. The variety may be enriching. As Gorwa explains, "[d]ifferent governance stakeholders have differing levels of regulatory capacity that they bring to the table: as Abbott and Snidal (2009b) argue, each type of actor has different competencies that are required at different phases of the regulatory process, from the initial agenda-setting and negotiations to the eventual implementation, monitoring, and enforcement of governance arrangements."[12] But the current scheme is also complex, burdensome and so far, perceived as inadequate.

Little has been written about the nature and effectiveness of the oversight mechanisms used or suggested so far or their ability to achieve their intended ends.[13] This is where this document wants to contribute, building on Professor Gorwa's paper on "*The Governance Triangle: Conceptualising the informal regulation of online content,*" analyzing the data and practices that arise from the implemented mechanisms to oversee, follow up, supervise or monitor compliance with current regulations and standards that regulate Internet companies to date.

---

[11] By *measures to guarantee* I mean those measures that according to the American Convention on Human Rights (for example) States are to adopt to prevent the violation of human rights, by the states and by third parties. According to Article 2 of the ACHR, states are to adopt legislation and regulation in pursuance of the protection of the rights established in the Convention. Lack of state legislation or regulation could be a source of international liability for failure to prevent and/or guarantee the rights set forth in the ACHR. See IACHR, Compendium on the Obligation of States to Adapt their domestic legislation to the Inter-American Standards of Human Rights, 2021, retrieved from: https://www.oas.org/en/iachr/reports/pdfs/CompediooobligacionesEstados-en.pdf. Say, for example, that the state knew or should have known that companies were imposing discriminating terms and conditions and did nothing to remedy the situation; the state could be held liable. See Bertoni E, Responsabilidad de actores no estatales por violaciones a la libertad de expresión (2021) retrieved from: https://www.observacom.org/responsabilidad-de-actores-no-estatales-por-violaciones-a-la-libertad-de-expresion/.

[12] Gorwa, R. The Platform Governance Triangle: Conceptualising the informal regulation of online content, Internet Policy Review (2019), 8(2), retrieved from https://policyreview.info/articles/analysis/platform-governance-triangle-conceptualising-informal-regulation-online-content.

[13] Many of the initiatives are recent and have been addressed individually, such as NetzDG or the European Codes of Conduct as seen below. But there have not been a lot of studies that compare and analyze the initiatives from the point of view of the supervisory body or mechanism. See German Network Enforcement Act (NetzDG) (2017) retrieved from: https://perma.cc/7UCW-AA3A.

This document does not intend to make a substantive analysis of the limitations or restrictions that the commitments imply for freedom of expression or other human rights, although there are many. The analysis of the compatibility of substantive regulated or self-generated measures with local, regional, and international human rights standards has been addressed elsewhere by other authors. [14] Rather, this article intends to survey and create a typology of enforcement mechanisms in new digital rights regulation.

The first section of this article frames the discussion of oversight in the legal theory of governance; then it analyzes different initiatives of regulation, co-regulation, and self-regulation centering on a few aspects of the mechanisms that impact their independence, impartiality, competence, and effectiveness: who heads and imposes the supervision, the legal nature of the obligations that are supervised; the material and geographical competence of the body; the supervision mechanisms and the tools to implement said supervision. Finally, this article concludes that there are critical challenges to how current supervision is conceived. Strictly self-regulatory mechanisms (i.e. ToS or transparency reports), although positive, lack legitimacy; and the mechanisms that are being designed by the states, as will be developed here, require substantial inputs and amendments to guarantee their independence, their compatibility with human rights, and above all, their effectiveness in achieving their intended purpose. In this framework, this essay suggests that co-regulation with expert multistakeholder oversight could be a plausible and even desirable model for the supervision of cross-jurisdictional behaviors and services and evaluates the best practices and the challenges that this model poses for its adoption in other areas.

[14] See: Special Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Report on Content Regulation, Human Rights Council, UN, 38th session, (2018); Schatsky, Del Campo, Lara & Hernandez, Mirando AlSur: Hacia nuevos consensos regionales en materia de responsabilidad de intermediarios y moderación de contenidos en Internet. [Looking Southward: Towards New Regional Consensus on Intermediary Liability and Content Moderation on the Internet] AlSur (2020); Bukovská, B., The European Commission's Code of Conduct for Countering Illegal Hate Speech Online An analysis of freedom of expression implications," Working Paper of the Transatlantic High Level Working Group on Content Moderation Online and Freedom of Expression (TWG) (2019); Chase, P. H., "The EU Code of Practice on Disinformation: The Difficulty of Regulating a Nebulous Problem," TWG, (2019); Tworek, H. & Leerssen, P., "An Analysis of Germany's NetzDG Law," TWG (2019); European Partnership for Democracy "Virtual Insanity? The need to guarantee transparency in digital political advertising," (2020); Nosák David (2021), "Overview of Transparency Obligations for Digital Services in the DSA," Center for Democracy & Technology (2021).

2. Governance and Oversight

Platforms, broadly, allow people to interact in various degrees of coordination in a virtual space. This coordination would not be possible without what the literature calls platform governance[15]. This concept is associated with a way of structuring social relations that allows them to be coordinated outside of state activity.[16] Coordination does not happen spontaneously but is determined by the policies of each company and the regulations that govern them.[17] Since there are different forms of coordination, each platform governance design will affect the behavior of its users in different ways and towards different ends.[18] The openness of the platform governance system, combined with their growing presence in our daily lives and the increasingly evident need for mechanisms to protect the human rights of users, has led the specialized literature to become increasingly concerned by how this governance is structured and justified and by its oversight structures.[19]

The proliferation of commentary, bills and proposals to regulate and "hold tech companies accountable" worldwide, but particularly in Europe, suggests there is some consensus around the need for independent control bodies over Internet companies to reduce the impact of possible human rights violations in the management and governance of these private networks and platforms.[20] Some concerns are focused on creating regulatory rules in increasingly public spaces[21], concentrating, for example, on the scant democratic nature of self-reg-

---

[15] Gorwa, R. "What is Platform Governance?" Information, Communication & Society 22.6 (2019): 854-871, retrieved from:10.1080/1369118X.2019.1573914

[16] Id.

[17] Id.

[18] Id.

[19] Klonick, Kate. "The new governors: The people, rules, and processes governing online speech." Harv. L. Rev. 131 (2017): 1598 retrieved from https://harvardlawreview.org/2018/04/the-new-governors-the-people-rules-and-processes-governing-online-speech/; Redeker, Dennis, L. Gill, and U. Gasser. "Towards digital constitutionalism? Mapping attempts to craft an Internet Bill of Rights." International Communication Gazette 80.4 (2018): 302-319 retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2687120; Celeste, Edoardo. "Digital constitutionalism: a new systematic theorisation" International Review of Law, Computers & Technology 33.1 (2019): 76-99 retrieved from https://www.tandfonline.com/doi/abs/10.1080/13600869.2019.1562604.

[20] See, for example, positions on the DSA: Article 19, At a glance: Does the EU Digital Services Act protect freedom of expression?, retrieved from https://www.article19.org/resources/does-the-digital-services-act-protect-freedom-of-expression/, Euractiv, Don't throw out the Digital Services Act's key accountability tools (June 2021) retrieved from https://www.euractiv.com/section/digital/opinion/dont-throw-out-the-digital-services-acts-key-accountability-tools/ or European Parliament News, Digital Services Act: regulating platforms for a safer online space for users (Jan 2022), retrieved from https://www.europarl.europa.eu/news/en/press-room/20220114IPR21017/digital-services-act-regulating-platforms-for-a-safer-online-space-for-users.

[21] On social media as public spaces or forums see: USSC, Packingham v. North Carolina, 137 S. Ct. 1730 (2017) retrieved from: https://www.supremecourt.gov/opinions/16pdf/15-1194_08l1.pdf.

ulation of platforms[22]; other concerns center on the design of implementation mechanisms and others on the implementation and operation of said mechanisms[23]. These different dimensions of governance influence its operation and the experience of its users.

Furthermore, there are disagreements about the scope of the concept of oversight or supervision. Oversight can be understood in a broad sense as being subjected to substantive regulations[24] (to regulate or not regulate)[25]. There are also references to oversight as a synonym for transparency.[26] Transparency in these cases is understood as an essential element for consumer protection, societal control, and state control over the impact of different intermediaries on human rights.[27] This type of transparency is what civil society requested with the Manila Principles[28], or more recently in the Santa Clara Principles[29]. UNESCO also refers to this transparency in a recent report where it identifies it as one of the central elements for the good governance of platforms.[30] Those who demand transparency maintain that added information about the practices and actions of each intermediary is essential to understand and contextualize their activity, their capacity for action or reaction, determine adequate standards of diligence, establish expectations, identify necessary and proportionate public policies to govern said activity, etc.[31] These transparency requirements would probably need

---

[22] Yablon, R. "Political Advertising, Digital Platforms, and the Deficiencies of Self-Regulation" Minn. L. Rev. Headnotes 104 (2020): 13 retrieved from https://minnesotalawreview.org/wp-content/uploads/2020/02/Yablon_Final.pdf.

[23] Gorwa, R, and Garton Ash, T. "Democratic transparency in the platform society," Social Media and Democracy: The State of the Field, Prospects for Reform (2020): 286 retrieved from https://www.researchgate.net/publication/343996372_Democratic_Transparency_in_the_Platform_Society.

[0] Gunashekar, S., et al, Oversight of emerging science and technology. Learning from past and present efforts around the world, RAND Europe (2019), retrieved from: www.rand.org/t/RR2921.

[25] See, for example, Balkin, Jack M. "How to regulate (and not regulate) social media." Knight Institute Occasional Paper Series 1 (2020) retrieved from https://knightcolumbia.org/content/how-to-regulate-and-not-regulate-social-media. For a general analysis of self-regulation applied to corporate criminal liability, see: Nieto, A. (2008), Responsabilidad social, gobierno corporativo y autorregulación: sus influencias en el derecho penal de la empresa, [Social liability, corporate governance, and self-regulation: their influences on corporate criminal law] Polít. crim., No. 5, 2008, A3-5, pp.1-18, retrieved from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2712397

[26] UNESCO, "Letting the sun shine in. Transparency and accountability in the digital age," 2021 retrieved from https://unesdoc.unesco.org/ark:/48223/pf0000377231. "Improved transparency by the companies would provide more information to users as well as help to give an evidence-base to the wider public debate about the impact of the companies on democracy, free expression and privacy."

[27] Id.

[28] Manila Principles, retrieved from https://manilaprinciples.org/index.html.

[29] EFF, Santa Clara Principles, retrieved from:https://santaclaraprinciples.org.

[30] UNESCO, "Letting the sun shine in. Transparency and accountability in the digital age," 2021.

[31] See M. Karanicolas, "A FOIA for Facebook, Meaningful transparency for online platforms" (2021) https://papers.ssrn.com/

to be subjected to some sort of verification, control or oversight. Finally, a third definition, and the one I am using in this paper, is oversight in a stricter sense, referring to the control mechanisms by which companies provide information on compliance with existing regulatory/self-regulated initiatives and how that compliance is verified — either by the regulator or by a non-state, external and independent actor.

For the purposes of this article, oversight or supervision mechanisms are the institutions designed to control the use of power.[32] These institutions may focus on the processes, for example, those aimed at monitoring platform management of certain types of content broadly, such as NetzDG[33] or the European Codes of Conduct[34]; or specific case resolution, for example, the Facebook Oversight Board[35].

A central question for any formal or informal institutional system is who should carry out the monitoring activity.[36] The answer to this question has a direct impact on the design of institutions and processes for government and their control.[37] Both the personal and institutional origin of the actors and the power in question, as well as the conditions for their permanence, have a significant influence on their activity.[38] The presence of more or fewer actors with veto power within an ecosystem, as well as the set of acceptable solutions to possible controversies, will have a direct impact on the products — outputs — to which these institutional

sol3/papers.cfm?abstract_id=3964235 where he argues that companies, to the extent that they have adopted a content curation role, particularly around the disinformation debate, should be subject to a state-style access-to-information regime, where the governing principle is openness by default and confidentiality as the exception.

[32] Madison, El Federalista N. 51: Si los hombres fueran ángeles el gobierno no sería necesario. Si los ángeles gobernaran a los hombres, no serían necesarias las contralorías externas ni las internas del gobierno." [If men were angels, no government would be necessary. If angels were to govern men, neither external nor internal controls on government would be necessary] cited in Spanish by Gargarella R. El Derecho como una Conversación entre Iguales [The Law as a Conversation Between Equals], Siglo XXI, (2021).

[33] German Network Enforcement Act (NetzDG) (2017) retrieved from: https://perma.cc/7UCW-AA3A.

[34] https://ec.europa.eu/info/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/eu-code-conduct-countering-illegal-hate-speech-online_en

[35] See https://oversightboard.com. This body of experts was hired by the company to carry out a subsequent control of content moderation in individual cases and under the terms and conditions of service, the principles that the company identified as relevant and guides for its action, and the international standards of human rights.

[36] Kinyondo, Abel, Riccardo Pelizzo, and A. Umar. "A functionalist theory of oversight." African Politics & Policy 1.5 (2015): 1-25 retrieved from https://www.researchgate.net/publication/292227902_A_Functionalist_Theory_of_Oversight.

[37] Gargarella R. El Derecho como una Conversación entre Iguales [The Law as a Conversation Between Equals], Siglo XXI, (2021).

[38] Madison, James. "The federalist no. 10." November 22, 1787 (1787): 1787-88. Gargarella R. El Derecho como una Conversación entre Iguales, Siglo XXI, (2021).

processes may give rise[39]. The design of the oversight mechanism, therefore, impacts the entire regulated ecosystem. Knowing and systematizing the oversight mechanisms, their powers, effectiveness, limits, and benefits is essential to understanding their operation and the expectations we can have for them.

The concern over the institutional mechanisms of supervision is not new and has been key to modern constitutionalism. Alexander Hamilton offered the most classic answer in The Federalist No. 78.[40] He stated that while the excesses that give rise to rights violations usually come from powerful actors, the body in charge of reviewing their actions must have institutional safeguards that provide it with tools to resist the attacks of these actors[41]. Thus, for example, the institutional independence of the control body would function as a guarantee of impartiality in a trial, insofar as those who judge would not be under the control of the interested parties.[42]

Although Hamilton thought about the relationship between rights, independence, and impartiality in the context of state institutions, this logic applies to the discussion around platform governance. This is the principle, for example, used to question and resolve the case of the European Data Retention Directive[43] (EDRD), which required platforms to keep the information they obtained about their users for a certain amount of time so states could use it for future research.[44] One of the main aspects discussed was the absence of an independent body to control the proper treatment of the data collected. According to Van Eijk, the principle that inspired the discussion was that such great power required special guarantees such as an independent oversight system.[45] In resolving the case, the European Court of Justice noted that the existence of an independent control body was a necessary element for processing personal data.[46] As there was no

[39] Tsebelis, George. "Veto players and institutional analysis." Governance 13.4 (2000): 441-474 retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1111/0952-1895.00141.

[40] Hamilton, Alexander, "The federalist no. 78." November 22, 1787 (1787).

[41] Id.

[42] Id.

[43] Directive 2006/24/EC of the European Parliament and the Council (2006) retrieved from: https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2006:105:0054:0063:EN:PDF

[44] ECJ, Joined Cases C-293/12 and C-594/12, Digital Rights Ireland Ltd v Minister for Communications, Marine and Natural Resources and Others and Kärntner Landesregierung and Others (2014), Para. 62. retrieved from https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:62012CJ0293

[45] van Eijk, N. "Standards for Independent Oversight: The European Perspective." Bulk Collection (2017): 382. retrieved from https://dare.uva.nl/search?identifier=afb51c71-09e9-4df7-abf1-132472f20261.

[46] ECJ, Joined Cases C-293/12 and C-594/12, Digital Rights Ireland Ltd v Minister for Communications, Marine and Natural

independent control body, the Court concluded that the EDRD violated art. 8.3 of the Charter of Fundamental Rights of the European Union.[47]

The need for independent control bodies vis-à-vis content and platform governance was also highlighted by civil society in different instances and concerning multiple oversight initiatives, including regulatory initiatives for different types of activities (personal data, content moderation, etc.). Article 19, for example, drafted a proposal in 2018 to create a supervisory body for social media replicating the media councils model and called it social media councils.[48] The main feature of this mechanism would be that they should be independent of the government, companies, and other special interests.[49] Multiple actors and organizations spoke in the same vein when dealing with Facebook's Oversight Board, for instance, highlighting this aspect as essential for the success of the initiative.[50]

The independence of a supervisory body is not characterized only by being able to decide without undue pressure, but also by the control that the different actors may have over the procedure.[51] Oversight mechanisms necessarily highlight some aspects of the supervised activity to the detriment of others. Any oversight process revolves around the prioritization of certain data or aspects, and any prioritization implies a cut and a decision in terms of institutional agenda.[52]

A recent study by the Centre on Regulation in Europe (CERRE) suggests a series of principles for good supervision of regulatory initiatives for digital platforms.[53] Some points refer to the efficacy of the regulation itself (according to criteria of necessity, proportionality, and legality). CERRE also mentions some aspects that would contribute to the creation of an effective supervisory ecosystem: (i) the establishment of an initial general guide by the regulator regarding what is

---

Resources and Others and Kärntner Landesregierung and Others (2014), Para. 62. retrieved from https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:62012CJ0293,

47 Id. Para 68.

48 Article 19, Self-regulation and hate speech on social media platforms (2018).

49 Article 19, Self-regulation and hate speech on social media platforms (2018). They also highlight that the mechanism must be established through a consultative and inclusive process, be democratic and transparent, ensure broad representation of different social sectors including NGOs, adopt a code of ethics; have a strong complaint mechanism and a transparent procedure for dealing with them and work in the public interest. See p. 12.

50 Facebook, Global Feedback & Input on the Facebook Oversight Board for Content Decisions, p. 20-21.

51 Article19, Facebook oversight board: Recommendations for human rights-focused oversight (2019)

52 T. Gillespy, "Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media" (2018).

53 de Streel, A. & Ledger, M., "New Ways of Oversight for the Digital Economy," CERRE, Issue Paper, (2021), p. 6. Retrieved from https://cerre.eu/publications/new-ways-of-oversight-digital-economy/.

expected of the platforms, and individualized guidelines for some of them, as is the case of start-ups; (ii) adopting codes of conduct related to the matter; (iii) the existence of internal compliance mechanisms of the digital platforms, that is, methods developed by the platforms to ensure respect for the rules in force, such as risk assessments, and the appointment of compliance officers; (iv) the empowerment of platform users so that they can make decisions; (v) have the support of independent auditors or trustees.

Some initiatives can be tested against the recommendations of CERRE, while others cannot. Some of the existing legal initiatives, for example, do not establish any specific monitoring mechanisms. This is the case of the DMCA, for example, which does not set a mechanism for monitoring how the delegated powers that the law affords to companies to uphold copyright protections work.[54] The GDPR, did not create a specific monitoring mechanism but rather entrusted supervision to the various personal data protection authorities of the different member states of the Union, which garnered criticism due to the lack of unified criteria for said oversight.[55] While the GDPR deals with data privacy rather than content, the example is valid from an oversight design point of view. And the Christchurch Call -a pledge developed by companies and states immediately after the terrorist attack in Christchurch, New Zealand- although it contains a series of specific commitments from both companies and states, it does not establish a follow-up mechanism for said commitments.[56] It establishes a body of "advisors" that includes distinguished members of international civil society, although it is not clear from the public information available what their concrete role is.[57] Along the same lines, GIFCT, an NGO that emerged from the initiative of several companies in 2017 and became formalized as an organization following up on the

[54] This may explain why academic or civil society initiatives that verify the operation of the law-ordered mechanism are de facto doing said monitoring. See D. Keller & P. Lerssen, "10 - Facts and Where to Find Them: Empirical Research on Internet Platforms and Content Moderation", Cambridge Univ Press (2020) retrieved from https://www.cambridge.org/core/books/social-media-and-democracy/facts-and-where-to-find-them-empirical-research-on-Internet-platforms-and-content-moderation/78DE9202F2D00F2967EFC5CBDCE2CAF0. The DMCA is one of the laws with the biggest global impact since US law establishes a safe harbor system ordering companies to generate a system to receive and adjudicate copyright claims. In the case of large content platforms, there is also an adjudication system that determines the admissibility of the claim and blocks the content, and an appeals system against that adjudication that each company administers. https://www.copyright.gov/legislation/dmca.pdf.

[55] N. Lomas, "GDPRs two year review flags lack of rigorous enforcement", Techcrunch (2020), retrieved from https://techcrunch.com/2020/06/24/gdprs-two-year-review-flags-lack-of-vigorous-enforcement/.

[56] Christchurch Call, retrieved from https://www.christchurchcall.com.

[57] Id. There is a group of civil society organizations that are part of the advisory board working on the creation of a set of criteria and conducting surveys against it to measure implementation of the call by both States and companies.

Christchurch commitments[58], does not establish a monitoring mechanism either. Like the Christchurch Call, it has an independent Advisory Committee made up of representatives of some states and civil society.[59] It is also not entirely clear the role of this Council in following up and monitoring the implementation of actions agreed upon within the GIFCT. While these initiatives mandate companies with specific duties, they do not concern themselves with how that mandate is carried out.

Furthering our knowledge of how the existing oversight processes are structured will contribute to our ability to understand the answers that these systems offer and may allow us to extract best and worst practices as the demand for rules and oversight over platforms and their activity intensifies.

3. A colorful and diverse ecosystem

This section classifies the initiatives analyzed according to different criteria, including who leads the supervision; whether the supervision is binding; the jurisdictional scope of the oversight initiative; the material competence of the supervisory body; the type of oversight envisaged; the consolidation of mechanisms of supervision; and uniformity of criteria and indicators. The initiatives analyzed are not the only ones; rather the criteria we followed was selecting initiatives that we perceive as having a greater potential impact in legislative debates and discussions in the Latin American region.[60]

a. Who leads the oversight?

Different actors are part of the current supervision ecosystem (where they exist) including public, private, and mixed actors. The Governance Triangle of Abbott and Snidal organizes regulatory institutions in terms of the roles that public and private actors play in their creation and governance[61]. The triangle is defined by three groups of actors: the State (in a broad sense, including individual gov-

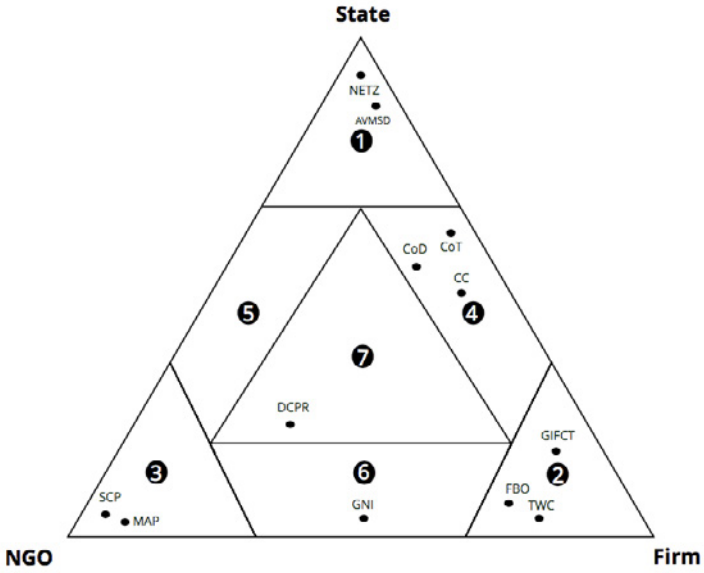
[58] Its founding members were Facebook, YouTube, Microsoft, and Twitter and it was founded in 2017.

[59] Global Internet Forum to Counter Terrorism (GIFCT), retrieved from https://gifct.org/governance.

[60] The relevance of the initiatives in the region is given by the attention paid to these efforts by civil society, academia, and policymakers in formal and informal communications and initiatives.

[61] K. W. Abbott & D. Snidal, The Governance Triangle: Regulatory Standards Institutions and the Shadow of the State, in The Politics of Global Regulation 44, 49-53. Retrieved from https://www.degruyter.com/document/doi/10.1515/9781400830732.44/html.

ernments, as well as groups of states or international organizations), companies (which includes private companies, groups of companies, and business associations), and civil society organizations (CSO). R Gorwa already placed Internet regulatory initiatives within the triangle as follows:



Source: Gorwa: 'Platform Governance Triangle' depicting the EU content regulation landscape, author's own formulation.

Regulatory initiatives are distributed in the triangle considering the role played by each of the three groups of actors in supervision and control.[62] The greater the participation of each group in the supervision and control of the initiative, the closer the initiative gets placed to the vertex that represents that group.[63]

Traditional regulatory initiatives such as Netz-DG[64] (or the Digital Millennium Copyright Act (DMCA) or the European Directive on the Protection of Personal Data (GDPR)) were included in the upper vertex since supervision is carried out by a public authority (although as shown below, oversight depends to a large extent on company reporting). The new RCT regulation and the DSA and DMA regulation proposals also fit within this vertex, since the state regulator will be the overseeing authority (again, from the drafts, with heavy dependency on the companies themselves to report their own activity). Notwithstanding, in some

[62] Id. The location should not be interpreted as a fixed category since several factors can affect the supervision of initiatives and this tool is only heuristic, to carry out a theoretical analysis of the control mechanisms.

[63] Id.

[64] German Network Enforcement Act (NetzDG) (2017) retrieved from: https://perma.cc/7UCW-AA3A.

cases, the law mandates that the company, through its own mechanisms, verify the implementation of the rule and report to the state. This is the case, for example, of NetzDG, which mandates companies to include the aggregated data related to the implementation of the law in their transparency reports.[65]

The Code of Conduct on countering illegal hate speech online (CDO) is at the center because its supervision consists of a collaborative mechanism between civil society organizations and the State (the European Commission). The mechanism is not technically a supervision mechanism but rather a testing tool: a system of "Trusted Flaggers" where some "trusted" organizations are set to identify content that violates the Code of Conduct, report violations and record the reactions and responses of companies to then inform the European Commission.[66] The Code of Conduct on Disinformation (CCD) was also included in the collaborative area, towards the right margin because supervision is a collaborative mechanism that involves companies, the European Commission, and also provides for an independent auditor appointed by member companies.

Regarding the initiatives at the base of the triangle, those in which the dominant role is that of private companies, GNI was placed in the center since its oversight mechanism (called assessment) involves the companies that are part of the initiative and representatives of civil society, academia and investors, while GIFCT was placed on the right vertex because supervision is headed by the signatory companies[67].

b. Legal nature of the obligations: binding/voluntary

The triangle helps to infer the distinction between legally binding initiatives and voluntary ones. The initiatives closest to the upper vertex of the triangle and therefore led by the State have generated legal obligations for companies and their non-compliance entails legal liability.[68]

The initiatives furthest away from the upper vertex of the triangle are voluntary.

---

[65] See Article 19, Germany: The Act to improve Enforcement of the Law in Social Networks (2017).

[66] EU Code of Conduct on Countering Illegal Hate Speech (2016), retrieved from: https://ec.europa.eu/info/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/eu-code-conduct-countering-illegal-hate-speech-online_en

[67] See https://gifct.org. This initiative was created and is directed by companies. However, within this framework, it establishes an advisory council made up of representatives of civil society, governments, and international organizations. See https://gifct.org/governance/.

[68] For example, NetzDG establishes fines for non-compliance, among other types of sanctions. See Article 19, Germany: The Act to improve Enforcement of the Law in Social Networks (2017).

They arise from agreements voluntarily signed by companies to adhere to certain standards, practices, or principles in their activity. While some authors argue, for example, for the applicability of estoppel — or binding unilateral promise — under which what was promised explicitly and freely by the company would constitute a binding legal commitment and therefore be enforceable,[69] this does not seem to be the prevailing position.

c. Jurisdictional scope

The different oversight initiatives vary according to the jurisdictional nature of the regulation, co-regulation, or self-regulation in question. NetzDG, for example, is an eminently national, German initiative, while the Codes of Conduct on Disinformation or Hate Speech are European regional initiatives, where those who intervene as formal authorities are the institutions of the European Community. In the case of GNI, for example, the initiative is global, with companies, scholars, civil society, and investors from different nationalities and regions taking part and acting based on agreed global principles (The GNI Principles on Freedom of Expression and Privacy).[70] The geographic scope of the initiatives will presumably gain relevance as oversight initiatives and mandated transparency initiatives multiply across the globe, imposing different standards and mechanisms to the same actors whose business depends on the interjurisdictional nature of their platforms.[71]

d. Material competence of the supervisory body

What is at stake when we think about oversight mechanisms is the rule of law, the rights of users, and the community as a whole. This rule of law and these individual and collective rights can be violated both by other users in the exercise

[69] See, among others, Barnes v. Yahoo!, 570 F.3d 1096 (9th Cir. 2009) and Yue v. Mao. Also, E. Goldman, Breach of Contract/Promissory Estoppel Claims Bypass Section 230 But Fail Anyways—Yue v. Miao Tech & Marketing Law Blog (2019) & Goldman, Speth on Barnes v. Yahoo and 230 as an Affirmative Defense, Goldman's blog (2009).

[70] The GNI Principles are available on the initiative page. At the time, the negotiation of these principles was controversial and problematic. See, for example, L. Downes, Forbes, Why no one will join the Global Network Initiative (2011). For a short history and description of the GNI see M. Samway, The Global Network Initiative: how can companies in the information and communications technology industry respect human rights? In Business and Human Rights, 1st ed. Routledge (2016).

[71] The Future of Speech Online: Can Transparency Be Mandated by Law?, Center for Democracy & Technology (CDT) (videoconference) (2021).

of their online activities and by the intermediaries in the management of their services.[72] What will each supervisory body or oversight initiative supervise?

The distinction between acts of third parties and acts of the companies is not always easy to make.[73] Two examples: A first one from the GDPR: while the GDPR regulates the treatment of personal data, by regulating the right to be forgotten it is deputizing on companies the enforcement of the rule over third parties.[74] A second example from NetzDG: On the one hand, it establishes obligations of procedure and transparency of practices for companies; on the other hand, it imposes substantive regulations on the speech of users (third parties) that will be subject to the processes and practices of the companies, and it delegates decision-making and implementation of corrective or sanctioning measures to intermediaries.[75]

When monitoring the correct implementation of the rule, the criteria to evaluate the existence and efficiency of protocols and processes to deal with content should be different from the criteria used to determine whether or not those protocols should have been activated (i.e., whether the content should indeed be classified as legal/illegal).[76] Experts in NetzDG explain that the methodology would not be entirely adequate to verify either of the two, as will be analyzed later.[77] The DSA, for its part, seems to aim to regulate both aspects of the equation, establishing obligations for the company vis-à-vis its own conduct and that of its users.[78]

On the other hand, GIFCT, for example, aims to create common standards and practices in the industry to deal with illicit or harmful actions by third parties.[79] And GNI looks at companies in their own activity and how they react and interact with state demands for data and expression.[80]

---

[72] This first distinction already generates problems when analyzing existing initiatives.

[73] See, A Del Campo, Teoricamente interesante, facticamente mundano [Theoretically interesting, factually mundane], CELE (2020).

[74] GDPR Art. 17 Van Cleynenbreugel, Pieter, What should be forgotten? Time to make sense of Article 17 GDPR from the point of view of data controllers. (2017).

[75] See German Network Enforcement Act (NetzDG) (2017) retrieved from: https://perma.cc/7UCW-AA3A.

[76] Tworek & Leerssen, op. cit pointing out the inadequacy of the quantitative methodology regarding the number of requests and content removals to evaluate the correct implementation of the law. On the other hand, they criticize that when the same content simultaneously implies a violation of the law and a violation of the company's terms and conditions of service established by self-regulation, these pieces of content will be reported under the latter category and not under the first.

[77] Id.

[78] Zech, Herbert, "General and specific monitoring obligations in the Digital Services Act." Verfassungsblog: On Matters Constitutional (2021).

[79] GIFCT, https://gifct.org.

[80] GNI Principles, GNI Principles on Freedom of Expression and Privacy

e. Supervisory mechanisms

Of the initiatives analyzed, both binding and non-binding, and public, mixed, or private, only six expressly contemplate instances of specific oversight. These mechanisms include a) self-assessments or internal monitoring; b) audits or assessments by third-party actors selected by the signatory companies of the initiatives (GNI for example)[81]; c) external audits carried out by non-state, external and independent actors selected by the state or supranational organization (such as the European Commission), or by the civil society or group of companies that promote the initiative; d) or supervision by state regulators (either an individual government or the European Commission). Finally, the CDO establishes a different mechanism, already explained in previous sections, where the "Trusted Flaggers" are the ones that test the operation of the mechanism and report on said operation over 6-8 weeks.[82] The table does not contemplate this mechanism.

Different initiatives opt for different mechanisms:

Table 1: Oversight or control

Initiative	Existence of supervision	Type of Supervision
ICT Coalition for Children Online	Yes	External selected by the initiative
Code of conduct on countering illegal hate speech online	Yes* (not for reporting but for initiative evaluation)	External selected by the European Commission
Code of Conduct on Disinformation	Yes** (not carried out)	External selected by the signatory companies
Network Enforcement Act (NetzDG)	Yes	External: German regulator
Global Network Initiative (GNI)	Yes	External selected by the signatory companies

* This initiative does not require reporting, but the audit is used to assess compliance with the objectives in monitoring exercises carried out by a network of independent organizations located in different EU countries.

** Although the Code of Conduct on Disinformation provides that the signatory companies will select an independent auditor to evaluate their reports and compliance with the commitments undertaken, to date, the companies have not selected an auditor.

Source: Developed in-house

The requirement of an independent auditor, whether hired by the company or appointed, may be an essential aspect of self-regulated initiatives. As Prof Goldman argues, independent auditing could help the ecosystem overcome some

[81] GNI Assessments https://globalnetworkinitiative.org/company-assessments/

[82] https://ec.europa.eu/info/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/eu-code-conduct-countering-illegal-hate-speech-online_en#monitoringrounds.

hard-to-solve challenges, including the safe disclosure of source data to the public or regulators, and they may "increase the likelihood of accurate disclosures, because the auditor cares about its reputation and will be risk-averse about any possible inaccuracies (…)."[83]

However, most of the voluntary initiatives do not establish independent control mechanisms. Rather, companies conduct a self-assessment of their own compliance (the reports are sometimes called self-assessments), often embedded in voluntarily generated transparency reports[84].

GNI is an exception among the mixed and voluntary initiatives, as it does establish an independent audit and evaluation requirement. With this mechanism, independent auditors, previously vetted by GNI, prepare the audits of companies, and submit their reports to the GNI Board, made up of companies, scholars, civil society, and investors.[85] The information presented by the companies is strictly confidential, therefore only some extracts and general conclusions are publicly available.[86]

The Code of Conduct on Disinformation and the ICT Coalition for Children contemplate the independent evaluation of the reports submitted by companies but have not yet implemented the mechanism.[87]

f. Reporting, auditing, and other tools

A large part of the initiatives analyzed include transparency obligations consolidated through the presentation of reports or monitoring meetings.

Of the initiatives analyzed, six require the presentation of reports containing information on the fulfillment of the commitments undertaken by the company: ICT Coalition for Children, the Codes of Conduct, GNI, GIFCT, and NetzDg. The other initiatives do not establish a control mechanism that requires reporting. The Code of Conduct to combat Illegal hate speech online sets up a control

---

[83] E. Goldman, The Constitutionality of Mandating Editorial Transparency, 73 Hastings L.J. __ (forthcoming 2022).

[84] Access Now reported that, as of July 2021, 88 companies had published transparency reports. https://www.accessnow.org/transparency-reporting-index/.

[85] See https://globalnetworkinitiative.org/about-gni/.

[86] https://globalnetworkinitiative.org/company-assessments/.

[87] Study for the "Assessment of the implementation of the Code of Practice on Disinformation" - Final Report, SMART 2019/0041." The study was commissioned by the Directorate-General for Communications Networks, Content and Technology (DG CNECT) of the European Commission, and was carried out by VVA Economics & Policy with the support of DisinfoLab (2019). See also ICT Coalition for Children online: https://www.ictcoalition.eu/

mechanism that includes regular meetings between companies and a monitoring exercise by organizations located in different EU countries, which assesses compliance with the commitments undertaken in the Code by companies using an agreed methodology — this methodology is not publicly available.[88] For its part, the Disinformation Code also provides for periodic evaluation meetings between the signatory companies, which could include the European Commission.[89]

Reports

In the six initiatives that require companies to present reports on their implementation, the periodicity for their presentation varies: three of them call for annual reports, one calls for reports every six months, and the other two do not specify frequency.

Three initiatives make reports publicly available, easily accessible to the public to consult, and are usually written in English — except for the Twitter reports on compliance with NetzDG which are only available in German —, while GNI and GIFCT keep reports confidential.[90] Under the European Code of Conduct, the Commission issues reports on compliance sharing data compiled by them on the implementation made by companies. In the case of GNI, the company shares the audit report with the Board, where they are discussed and subject to an evaluation of the progressive implementation of the GNI Principles over time.[91] The organization publishes an annual report which includes extracts and summaries of some reports, the main conclusions that emerged from the exercise, and some recommendations.[92]

---

[88] See Code of conduct on Countering Illegal Hate speech online (2019) retrieved from: https://ec.europa.eu/newsroom/just/items/54300/en; & EU Code of Conduct against illegal hate speech online: results remain positive but progress slows down, retrieved from: https://www.pubaffairsbruxelles.eu/eu-code-of-conduct-against-illegal-hate-speech-online-results-remain-positive-but-progress-slows-down-eu-commission-press/.

[89] EU Code of Practice on Disinformation, section IV retrieved from: https://www.hadopi.fr/sites/default/files/sites/default/files/ckeditor_files/1CodeofPracticeonDisinformation.pdf.

[90] On the ICT Coalition for Children Online, see, for example, the Disney report here (https://www.ictcoalition.eu/medias/commitmentcontent/61/file/2019-twdc-ict-coalition-implementation-report.pdf). On GNI, https://globalnetworkinitiative.org/wp-content/uploads/2020/04/2018-2019-PAR.pdf. Code of Conduct reports on disinformation 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. On GIFCT, refer to https://gifct.org/wp-content/uploads/2021/12/GIFCT-Annual-Report-2021-PV.pdf

[91] See, for example, the report prepared for the 2018/2019 period retrieved from: https://globalnetworkinitiative.org/wp-content/uploads/2020/04/2018-2019-PAR.pdf.

[92] Id.

Below is a table that summarizes the information detailed above:

Table 2: Reporting, frequency, and publicity

Initiative	Presenting reports	Frequency	Publicity
ICT Coalition for Children Online	Yes	Not Established	Yes
Code of conduct on countering illegal hate speech online	Yes*	-	-
Code of Conduct on Disinformation	Yes	Annual	Yes
Network Enforcement Act (NetzDG)	Yes	Biannual	Yes
Global Network Initiative (GNI)	Yes	Biannual**	No
Global Internet Forum to Counter Terrorism (GIFCT)	Yes	Annual	No
Christchurch Call	No	-	-
General Data Protection Regulation (GDPR)	No	-	-
Digital Millennium Copyright Act (DMCA), United States of America	No	-	-

* The European Commission issues periodic reports on the Code's implementation that collect the information provided by Trusted Flaggers.

** The GNI conducts their assessments every 2-3 years and it is not clear from their documents what the periodicity should be.

Source: Developed in-house

Overall, whether reports are public or not, there seems to be a high level of compliance with the reporting obligation, particularly where the periodicity is well established. On the other hand, in the case of ICT for Children Online, as the periodicity of the report is not clear, it is difficult to measure the company's level of compliance. Some have submitted two or three reports, and others have only submitted one.

Table 3: Compliance level (reporting)

Initiative	Compliance level (reporting)
ICT Coalition for Children Online	Medium*
Code of Conduct on Disinformation	High
Network Enforcement Act (NetzDG)	High
Global Network Initiative (GNI)	High
Global Internet Forum to Counter Terrorism (GIFCT)	-

* Periodicity not specified in the commitment

Source: Developed in-house

As expressed in the previous paragraphs, only five of the nine initiatives analyzed contemplate the submission of reports by companies as a control mechanism. One additional initiative includes the publication of a report by the oversight mechanism itself (the Code of Conduct). In the text of the commitments, three of these five initiatives also identify, in greater or lesser detail, the information that these reports must include but only the ICT Coalition for Children Online provides templates for these reports. The GNI provides the Assessment toolkit that was amended in 2018 to include specific information and sections that assessors should look to in preparing their assessment reports.[93] The lack of templates may be due to different reasons, but overall, from the specialized literature, lack of agreement and excess ambiguity over terms are among the most concerning causes.[94]

Initiatives that require reporting are listed below:

Table 4: Report Consistency

Initiative	Pre-established information to present	Templates
ICT Coalition for Children Online	Yes	Yes
Code of Conduct on Disinformation	Yes	No
Network Enforcement Act (NetzDG)	Yes	No
Global Network Initiative (GNI)	Yes*	No
Global Internet Forum to Counter Terrorism (GIFCT)	No	-

* GNI publishes an illustrative assessment toolkit detailing what is expected of companies, auditors, and the Board in the evaluation of the audits presented.[95]

Source: Developed in-house

Although there is some discussion as to whether reporting should be standardized[96], the required information in most cases is not specified, and there are no report templates, so it is difficult to compare the information presented by the companies.[97] For example, in the case of the Code of Conduct on Disinfor-

[93] GNI Assessment Toolkit, https://globalnetworkinitiative.org/wp-content/uploads/2021/11/AT2021.pdf.

[94] MacCarthy, M. "Transparency Requirements for Digital Social Media Platforms: Recommendations for Policy Makers." Algorithms (2020),

[95] GNI, https://globalnetworkinitiative.org/wp-content/uploads/2021/11/AT2021.pdf

[96] D. Keller, Some Humility About Transparency, Ctr. for Internet & Society (Mar. 19, 2021).

[97] See notes No. 21, 70, and 75.

mation, the information presented by Facebook, Twitter, Google and Microsoft regarding "fake or false accounts" (in the framework of the integrity of services commitment) varies as follows:

- They all include the measures/actions/tools that the platforms/services use to prevent and combat fake/false accounts.
- Only Facebook and Twitter provide information on the total number of accounts that were deactivated, and Twitter also includes the number of reports of fake accounts and spam they received from users.
- None of the companies include disaggregated data by region on this topic.

The exceptions are the ICT Coalition for Children Online and the GNI. The ICT CCO does provide a report template with predetermined information for companies that are part of the initiative. However, this does not prevent some companies from providing more information than others. For example, both Facebook and Google, in addition to the explanation of the information requested, include examples with images and links with additional information in their reports, while other companies, such as Disney, do not include detailed information or images.[98] But even when these reports include templates, sometimes there are discrepancies as to how different companies count the instances that they are reporting.[99]

GNI also uses pre-set questions for independent evaluators to collect information from companies. Due to the methodology and the objectives of the initiative, the information is qualitative rather than quantitative, which allows for divergence between the information provided by companies. The complete reports are only accessible to the Board but not to other members, state regulators, or the general public.[100]

Monitoring meetings

The Code of Conduct on Countering Illegal Hate Speech Online provides for monitoring meetings between the signatory companies, the European Commis-

[98] See https://www.ictcoalition.eu/news/news.

[99] D. Keller, Some Humility About Transparency, Ctr. for Internet & Society (Mar. 19, 2021) retrieved from https://cyberlaw.stanford.edu/blog/2021/03/some-humility-about-transparency.

[100] The limited information publicly available about the GNI process and its evaluation may undermine the credibility of its assessments.

sion, and civil society organizations, as a control mechanism for the initiative.[101] The Code of Conduct on Disinformation, for its part, establishes that the signatories will meet periodically to assess developments within the framework of the Code, but there is no public information on those meetings.

In the framework of GNI, the assessment of companies is carried out every two to three years, by companies accredited by the GNI Board, and are subject to review by the Board during assessment review meetings. The standard for review is whether the company is making good faith efforts to implement the GNI Principles and its "Implementation Guidelines" over time, and the decision of whether a participating company meets this standard is adopted within the assessment review meeting after the Board has reviewed the report and asked questions of the assessors and the company.[102]

g. Indicators for monitoring

The European Union Commission in its agenda for better regulation of the European Union (2015) ("Better regulation for better results —An EU agenda") gives an account, among other things, of the importance of having specific tools to supervise the correct and effective implementation of laws and evaluate their validity and adequacy in time.[103] The existence of objective indicators to verify the accurate implementation of the regulation by the regulator is fundamental in this aspect[104].

---

[101] EU Code of Conduct on Countering Illegal Hate Speech Online, retrieved from https://ec.europa.eu/commission/press-corner/detail/es/IP_18_261

[102] GNI Implementation Guidelines, retrieved from https://globalnetworkinitiative.org/implementation-guidelines/.

[103] European Commission, "Better Regulation for Better Results—An EU agenda." COM 2015 (2015). In particular, see point 3.3. Retrieved from https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52015DC0215. The Commission recognizes that regulation has an impact on member countries and users, companies, etc. in multiple sectors and that the effects of regulation are long-lasting. Id. 1.1 "Legislation should do what it is intended to do, it should be easy to implement, provide certainty and predictability and it should avoid any unnecessary burden. Sensible, realistic rules, properly implemented and enforced across the EU. Rules that do their job to meet our common objectives - no more, no less… EU policies should also be reviewed regularly: we should be transparent and accountable about whether we are meeting our policy objectives, about what has worked well and what needs to change." Hence, they also undertake to review the effectiveness of their regulations and evaluate their impact or their modifications over time and regulate in the least harmful and intrusive manner under the principles of necessity and proportionality. Id. 3.1: New, integrated Guidelines on Better Regulation will apply to the Commission's work from today 6. They will ensure that economic, social, and environmental impacts continue to be considered alongside each other in all of the Commission's analytical work together with fundamental rights. They re-commit the Commission to use the best available evidence and science and reinforce the commitment to put in place clear monitoring and implementation plans before measures are adopted. They will also ensure that keeping the EU competitive and the EU's development sustainable remains a priority in all we do.

[104] 6th monitoring round of the Code of Conduct, retrieved from https://digital-strategy.ec.europa.eu/en/library/assessment-code-practice-disinformation-achievements-and-areas-further-improvement.

The types of indicators could vary depending on the initiative or process to be implemented or measured. Indicators can be quantitative (for example, the number of affected pieces of content, implemented processes, appeals received, affected users, time, etc.); or qualitative (usability of processes or systems, user experience, satisfaction evaluation, etc.); and they can be a mixture of both. The absence of indicators hinders the substantive evaluation of the initiative and its supervision, particularly when the latter is given by audits, self-assessments, or reports generated by the companies.[105] The same could be said of the consensus and clarity around the concepts that the initiative uses in each of its mandates.[106]

Regarding the Code of Conduct on Disinformation, the Transatlantic Working Group,[107] a commission that brings together media reps, scholars, civil society, and fact checkers of the Multistakeholder Forum on online disinformation in response to the European Commission's query, reported that the initiative was deficient, to a large extent due to the lack of transparency and clear indicators (Key Performance Indicators).[108] The absence of measurable objectives and performance indicators was decisive in evaluating the effectiveness of the mechanism.[109] In response to the criticism, the European Commission published an action plan ordering companies to detail the measures taken in compliance with the Code and detailed the quantitative indicators that it would look at to evaluate the initiative:[110] number of accounts removed for violating advertising policies; of websites blocked for data scraping; of political advertisements removed due to lack of transparency; of entries in the repository; and of accounts identified as fake.[111] Despite the initial effort, the Commission in charge of evaluating the initiative for the European Commission concluded, more than a year later, that the

[105] See Tworek, H. & Leerssen, P., "An Analysis of Germany's NetzDG Law," TWG (2019) retrieved from https://www.ivir.nl/publicaties/download/NetzDG_Tworek_Leerssen_April_2019.pdf.

[106] Id. One of the Committee's conclusions is that the initiative would benefit from better-defined terms. The lack of consensus around key definitions (such as advertising, organic content, etc.) undermines the implementation and evaluation of the initiative.

[107] Chase, P. H. (2019), "The EU Code of Practice on Disinformation: The Difficulty of Regulating a Nebulous Problem," Working Paper (TWG), p. 9, retrieved from: https://www.ivir.nl/publicaties/download/EU_Code_Practice_Disinformation_Aug_2019.pdf

[108] https://wayback.archive-it.org/12090/20210728072832/https://digital-strategy.ec.europa.eu/en/policies/code-practice-disinformation.

[109] Ibid. "As outlined by the Sounding Board's previous written feedback and comments, the "Code of practice" as presented by the working group contains no common approach, no clear and meaningful commitments, no measurable objectives or KPIs, hence no possibility to monitor the process, and no compliance or enforcement tool: it is by no means self-regulation, and therefore the Platforms, despite their efforts, have not delivered a Code of Practice."

[110] https://wayback.archive-it.org/12090/20210621080449/https://digital-strategy.ec.europa.eu/en/library/report-implementation-communication-tackling-online-disinformation-european-approach.

[111] Chase, P. H. (2019), "The EU Code of Practice on Disinformation: The Difficulty of Regulating a Nebulous Problem," (TWG), p. 9, retrieved from: https://www.ivir.nl/publicaties/download/EU_Code_Practice_Disinformation_Aug_2019.pdf

insufficient indicators combined with the ambiguous and unclear terminology of the agreement made its implementation deficient and its control, impossible.[112]

On the Code of Conduct on Countering Illegal Hate Speech Online, Bukovska pointed out that the sole criterion adopted by the European Commission seems to be speed and number of content removal actions.[113] This methodology, she adds, can hardly be considered an indicator of success. The variation in the numbers, in her opinion, could be explained in different ways, including the growing consensus among companies and cooperating organizations on what types of speech should be removed, attributable to different variables: the flaggers improved understanding of what types of content are better received by companies and promote those resources over others; that the companies have yielded in a show of good faith and dispute less the nature of the content; or that the ToS have changed over time.[114] Consequently, the author states that the only qualitative conclusion that could be drawn from the four monitoring reports was that "there has been a steady increase of removals of "hate speech" content – as vaguely and broadly defined in the Code of Conduct – within the specific time-period based on requests from specific organizations based on unspecified methodology."[115]

Since the launch of the Code, the European Commission has published six reports on its application.[116] These reports present the results of the monitoring exercises carried out each year (over a period of six to seven weeks in each case), in which organizations based in different EU Member States apply a common methodology

[112] Code of Conduct on Disinformation Assessment – (Oct 2021) retrieved from https://digital-strategy.ec.europa.eu/en/library/assessment-code-practice-disinformation-achievements-and-areas-further-improvement.

[113] Bukovska, p. 9.

[114] Bukovska, p. 9. Fourth, the only criterion of "success" presented by the European Commission in the monitoring reports appears to be the speed and number of the removals. As noted above, the most recent 2019 report comments on the improvement of the removal rates and states that only 28.3% of reported content remained online, while "this represents a small increase compared to the 70% one year ago." However, the rate of removals can hardly be considered an indicator of "success;" all it shows is the increase of consensus between the IT companies and the cooperating organizations on what content should be removed. This can be interpreted in several ways. For instance, it might show that the flagging organizations better understand (though the report provides no insight as to how) IT companies' policies and what content might not be acceptable under the respective Community Guidelines. Alternatively, it can indicate that the IT companies simply decided to respond positively to more requests to show goodwill and desire to comply with the exercise. Another possible interpretation is that over time the IT companies changed their content moderation practices and assess the content differently compared to a few years ago.

[115] Bukovska, "All in all, the only qualitative conclusion from the four monitoring reports is that there has been a steady increase of removals of the "hate speech" content – as vaguely and broadly defined in the Code of Conduct – within the specific time based on requests from specific organizations based on unspecified methodology. Overall, the monitoring reports provide very little information on the real effectiveness of the Code of Conduct system and what impact it has in protecting groups at risk of discrimination and hatred and ensuring that the right to freedom of expression is protected."

[116] Consulted reports retrieved from https://ec.europa.eu/info/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/eu-code-conduct-countering-illegal-hate-speech-online_en.

and notify the companies about alleged cases of illegal speech online and record companies' response level and time.[117] The number of participating organizations varies in each exercise.[118] The reports prepared by the Commission present various difficulties when it comes to analyzing the characteristics and efficiency of this control mechanism.[119] On the one hand, reports do not include information on the methodology used by the organizations to report illegal content to companies, nor to evaluate the content moderation carried out by them.[120] Information about the analysis carried out by companies to accept or reject recommendations of content marked as illegal by flaggers is also not included.[121]

On the other hand, the reports only provide information on the answers that the companies give to the requests made by the organizations in the framework of the monitoring exercise in the pre-established period. They do not include information about companies regularly conducting content moderation or how companies have changed their business practices to implement the Code of Conduct, how companies assess content removal requests or how they explain their decision to decline or accept those requests.[122]

Tworek and Leerssen make the same criticism regarding the application of the NetzDG. According to these authors, it is clear from the first two reporting periods that a large part of the implementation of the law has been done by applying the terms and conditions of service and not by applying the law itself.[123] In many cases, companies have established a dual review system in response to these initiatives: they first review the compatibility of the detected or reported content with their own terms and conditions of service; then if the content does not violate the ToS, they apply the analysis of the corresponding law. Only these

[117] See published reports and fact sheets here: https://ec.europa.eu/info/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/eu-code-conduct-countering-illegal-hate-speech-online_en#monitoringrounds

[118] Id.

[119] Assessment of the Code of Practice on Disinformation SWD (2020)180 Final - "[t]he SWD highlights that, in order to ensure a complete and consistent application across stakeholders and Member States, the Code should be further improved in several areas by providing commonly-shared definitions, clearer procedures, more precise and more comprehensive commitments, as well as transparent key performance indicators (KPIs) and appropriate monitoring. Participation should be broadened to include other relevant stakeholders, in particular from the advertising sector."

[120] Bukovska, p. 8. It is not possible to identify from the reports if the flagged content results from the organizations' evaluation of compliance with local criminal legislation or with the Community Guidelines of the companies, nor is it revealed if the flaggers receive some training on the applicable standards on freedom of expression and the need to balance content removal with this right.

[121] Jourová, Věra, How the Code of Conduct helped countering illegal hate speech online (2019).

[122] See Bukovska, pp. 7-10, op. cit.

[123] Tworek & Leerseen, op. cit.

cases are reported as cases of implementation of the law.[124]

Finally, the only criteria for measuring the success of the initiative are the companies' speed in removing illegal content and the number of pieces of content removed. Most of the criticisms and concerns about the NetzDG are in line with this and relate to the risk of over-removals or over-blocking of content, affecting the right to freedom of expression.[125] The German government argues against this claiming the fines for non-compliance with the law only apply to cases of failure to block or remove content when it is a systemic failure to comply on the part of the platform, and not for individual cases.[126] However it is unclear how the authorities determine what qualifies as systemic.[127]

For its part, the DSA establishes, for instance, risk management and due diligence obligations, but as several organizations pointed out, it does not set any guidelines for the organization, procedure, or content of said processes.[128] Nor do they set clear criteria for the adoption of risk mitigation, which the text of the initiative contemplates.[129]

GNI seems different from the previous cases in that the methodology and evaluation indicators are carried out exclusively around qualitative, not quantitative data and process rather than outcomes.[130] As a first step, membership in GNI entails the adoption of the GNI Principles on Freedom of Expression and Privacy.[131] These Principles are based on the major human rights documents, including the United Nations Guiding Principles on Business and Human Rights. In addition to the Principles, GNI has defined guidelines for their implementation agreed

[124] See Tworek & Leerseen, op. cit. See also DSA Human Rights alliance, Statement, October (2021). https://www.accessnow.org/cms/assets/uploads/2021/10/Digital_Services_Act_Human_Rights_Alliance_Statement.pdf.

[125] See note No. 4. https://www.dw.com/en/germany-implements-new-Internet-hate-speech-crackdown/a-41991590

[126] Unterrichtungdurch die Bundesregierung, Bericht der Bundesregierung zur Evaluierung des Netzwerkdurchsetzungsgesetzes (2020).

[127] William Echikson and Olivia Knodt, Germany's NetzDG: A key test for combatting online hate, No. 2018/09, November 2018.

[128] DSA Human Rights Alliance Statement: "Platforms must also put in place reasonable, proportionate, and effective risk mitigation measures. There is no requirement in the text that such measures must be "necessary" and there are no standards in place for such risk assessments. It's up to platforms to decide which mitigation measures to take. The relevant provisions do not require a human rights impact assessment and encourage the set-up of codes of conduct. The role of civil society in all of this is unclear."

[129] See, for example, https://www.accessnow.org/where-the-dsa-is-heading/; Stiftung Neue Verantwartung, The DSA Draft: Ambitious Rules, Weak Enforcement Mechanisms, Policy brief (2021).

[130] https://globalnetworkinitiative.org/wp-content/uploads/2021/11/AT2021.pdf.

[131] https://globalnetworkinitiative.org/gni-principles/.

upon by all its members.[132] These guidelines serve to instruct companies in the implementation and establish a series of concrete actions that member companies commit to doing in the framework of compliance.[133] The assessment toolkit details the process that assessors have to follow and the questions that they have to answer in all assessments.[134] None of these questions pursue quantitative data but rather qualitative and in general terms referring to processes, and the relevance of the GNI principles and the rights to privacy and freedom of expression in the governance of the company. Thus, some of the questions aim to understand who implements the GNI principles within the company and how; the relevance of this team, and how it communicates and interacts with the various product and regional teams, as well as its interaction with the executive management and the Board of shareholders of each company; how the entire organization is trained in compliance with the Principles; the processes regarding due diligence and human rights impact assessments; how these processes are articulated, etc.[135] In addition to the evaluation around governance and processes, companies must include in the assessment a series of cases that show the operation of said processes, the good practices established, and the challenges they faced.[136] The confidential nature of the evaluation process makes these exchanges particularly valuable when they include not only success stories but also examples of failure or problems. The evaluation guided by the toolkit allows for uniform and comparable criteria equally applicable to the different companies and processes that coexist within the GNI: companies that offer a range of products and services that are subject to assessment, including telecommunications equipment, mobile network services, social media platforms, search engines, and content delivery networks).

The assessor's role in these evaluations is particularly relevant and hence, in the GNI framework, the assessors accredited to carry out the assessments must be approved by the GNI Board based on demonstrated capacity and human rights expertise.[137]

---

[132] https://globalnetworkinitiative.org/implementation-guidelines/.

[133] Id. 1.2 These Implementation Guidelines provide further details on how participating companies will put the Principles into practice. The purpose of this document is to:

Describe a set of actions by which a company would demonstrate that it is implementing the Principles with improvements over time.

Provide companies with direction and guidance on how to implement the Principles.

[134] GNI Assessment Toolkit, https://globalnetworkinitiative.org/wp-content/uploads/2021/11/AT2021.pdf, Annex I.

[135] https://globalnetworkinitiative.org/wp-content/uploads/2021/11/AT2021.pdf. Annex I.

[136] https://globalnetworkinitiative.org/wp-content/uploads/2021/11/AT2021.pdf. See the case evaluation framework in Annex II.

[137] GNI, https://globalnetworkinitiative.org/company-assessments/

Finally, the GNI evaluation criteria provide a radically different starting point from that of other regulatory or supervisory initiatives.[138] GNI's assessment is conducted to determine whether the company is making good faith efforts towards continued and improved implementation of the GNI Principles over time.[139] There are several aspects to highlight here, but perhaps the main one is the progressive nature of the standard of "improvement over time." Whether or not the company always complies with the principles is not the objective of assessment; in many cases, the Principles do not ask or point to concrete results, but rather create procedural mandates: due diligence measures or implementing human rights impact studies when the situation warrants according to the Principles.[140] What GNI as an organization assesses is whether companies can show good faith efforts to implement the GNI Principles over time.

Obligations of due diligence and impact assessments have been embedded in GNI's Principles from the beginning. Companies are expected to point out what kinds of procedures they have in place to conduct these due diligence exercises, what processes they have adopted, under what circumstances they conduct these assessments, and what processes are in place to implement recommendations or lessons learned from such exercises. Concrete cases are not evaluated except through "case studies" selected by the company and in some instances suggested by the non-company constituency members, to show how those processes and practices actually work.[141]

4. A proposal for a co-regulatory model

The difficulty of establishing efficient and effective supervision mechanisms is evidenced in multiple ways through this document. The majority of the mechanisms outlined so far to align processes is unclear and lack structure, standards, and indicators to assess their implementation and enforcement. Most of them focus almost solely on quantitative data rather than qualitative data, lack independent verification, and are complex methodologically to provide appropriate

[138] GNI, https://globalnetworkinitiative.org/company-assessments/ "The purpose of the assessment is to enable the GNI Board to determine whether each member company is "making good-faith efforts to implement the GNI Principles with improvement over time."

[139] Id.

[140] GNI Principles retrieved from: https://globalnetworkinitiative.org/gni-principles/.

[141] https://globalnetworkinitiative.org/wp-content/uploads/2021/11/AT2021.pdf. See the case evaluation framework in Annex II.

oversight of company compliance with substantive rules or commitments and freedom of expression standards. Evaluating compliance is hard and burdensome, and, for the most part, ineffective.

In this framework, the Global Network Initiative seems to stand out as a virtuous model, though exclusively within the self-regulation realm and while narrow in its current scope. Among the virtues of the model there is the existence of expert, independent review; the existence of criteria and indicators, mainly qualitative, that allow processes to be evaluated and compared across a wide range of companies that offer radically different services and products; a focus on company processes rather than outcomes and a progressive standard of compliance over time; and the global, non-state, multistakeholder nature of the initiative (that may also be listed as a con). Still, the initiative is exclusively self-regulatory, voluntary, and based on the GNI Principles, which were negotiated by the companies to focus solely on government demands for data or content blocking. The progressive nature of standards can be perceived as soft and too low a bar to address current concerns over tech companies' governance and conduct. The initiative is built on the premises of trust between the different constituencies that are part of the GNI and strict confidentiality, which has hindered the initiative's ability to share concrete insights as to the findings and nuances within each assessment, thereby impacting its legitimacy as an oversight mechanism within the broader ecosystem. Finally, the costs associated with this initiative, particularly for members of civil society and academia (who volunteer their time), but also for companies and investors are significant, especially as reporting obligations are climbing.

While the models that we have seen so far are faulty, for different reasons, there are virtues within each and every one of them. The binding nature of legally mandated oversight can be hardly construed within a self-regulatory mechanism. And the trust developed among the members of a self-regulatory and voluntary oversight mechanism can hardly prosper within a state-led mechanism, where confidentiality usually clashes in one way or another with state principles of access to information and accountability.

Additionally, as Gorwa argues, "Multistakeholder regulatory standards setting schemes have proliferated for other industries because they often are the best out of a slew of bad options (Gasser, Budish, & Myers West, 2015). While they can be easy to dismiss outright as "soft" or "non-binding", the political science literature on transnational governance shows that such schemes are a vital part of the

corporate regulatory toolbox, and can help us understand complex relationships of contestation and bargaining across different actor preferences."[142]

A co-regulatory approach that binds companies to a legitimate, expert, independent oversight of their processes and practices, of their choice (although under threat of state oversight by default), focusing on qualitative rather than quantitative indicators, could serve multiple valuable purposes:

1. Develop the United Nations Guiding Principles on Business and Human Rights and their implementation further: As professors Creamer and Simmons argue, "The report-and-review process is a dialogue—not an exam and certainly not a trial. It is an opportunity for states to learn about best implementation practices or more efficient and effective methods to improve treaty outcomes. When review and dialogue accompany self-reporting, opportunities arise for state and international elites to "learn more about one another's position and perspectives, desires and constraints." Learning exactly how to implement one's treaty obligations is thus a highly plausible explanation for the correlation between cumulative reporting activity and rights improvements."[143] Although their work focused on self-reporting obligations under international human rights treaty systems, the observation may very well apply here. In fact, upon introducing the subject, they cite the experience of U.S. Section 1502 of the Dodd-Frank Act, "which required companies to disclose annually whether they had obtained certain minerals from mines controlled by armed groups in the Congo."[144] As the authors explain, "yet, this much-maligned reporting requirement spawned potentially interesting dynamics. Some companies began to investigate the provenance of their mineral inputs, which incentivized the implementation of firm governance models that increased the feasibility of tracing mineral sourcing. Models of supply chain due diligence were adopted to replace individual firms' uninformative go-it-alone reports. Third-party auditors developed a capacity to certify certain sources as "conflict-free" (…). More importantly, firms were innovating even as they began — often reluctantly — the task of self-repor-

142 Gorwa, R. (2019). The platform governance triangle: conceptualising the informal regulation of online content. Internet Policy Review, 8(2). https://doi.org/10.14763/2019.2.1407.

143 Creamer & Simmons, "The Proof is in the Process: Self-Reporting Under International Human Rights Treaties" Faculty Scholarship at Penn Law, 2145 (2020) retrieved from https://scholarship.law.upenn.edu/faculty_scholarship/2145

144 Id. Citing Dodd-Frank Wall Street Reform and Consumer Protection Act, Pub. L. No. 111-203, 124, § 929-Z, 124 Stat. 1376, 1871 (2010) (codified at 15 U.S.C. § 78o).

ting."[145] A co-regulatory approach that mandates self-reporting and assessment by technical, independent stakeholders could eventually lead to the adoption of better, stronger policies and practices directed at protecting and promoting human rights within tech companies beyond the bare minimums that any goal-oriented legislation could impose.[146]

2. Contribute to and complement transparency efforts and demands. Transparency demands and current obligations have focused mostly, as this paper shows, on qualitative data, based on actual outcomes. The processes leading up to measurable outcomes are not currently looked at for transparency purposes even though the main contribution from the UNGP to the human rights landscape, vis-à-vis businesses, focused on due diligence processes rather than outcomes.[147]

3. Unify accountability standards for transnational services: Unified rules are vital for developing transnational services. This is why we have a Convention for the international trade of goods or an agreement over international private law and rules to determine when courts have jurisdiction and which law they should apply. Global mechanisms capable of developing coherent, solid, unified technical standards and goals for Internet companies in their implementation of human rights standards are imperative.

4. Level the playing field among states with different degrees of geopolitical power to regulate: Different states have different powers to impose obligations or jurisdiction over transnational businesses. Asserting jurisdiction over transnational corporations is not easy and is not always possible. Different legal norms determine the ability of a state to claim jurisdiction over a certain case or person. And afterward, there needs to be an ability to implement or execute whatever norms or judgments that may arise from domestic courts or Congresses. An international, independent, technical oversight mechanism could contribute to leveling the playing field for Sates with less relevant geopolitical regulation powers.

5. Democratic accountability has been sometimes placed against international mechanisms. However, voluntary state (democratically achieved) delegation

---

[145] Id.

[146] Id.

[147] de Felice, D, "Business and Human Rights Indicators to Measure the Corporate Responsibility to Respect: Challenges and Opportunities," 37 HUMAN RIGHTS QUARTERLY 2, pp. 511-555 (2015).

Retrieved from: https://www.ohchr.org/sites/default/files/Documents/Issues/Business/ForumSession4/Felice.pdf.

of certain oversight powers to international tribunals and oversight mechanisms is not only not unheard of, but rather the common standard for most countries around the globe, at least in the human rights field.

GNI could serve as a possible model for co-regulatory initiatives, although of course not the only one. In fact, in 2013 a US Senator submitted an initiative to the US Congress to impose upon companies that operate in jurisdictions that restrict the Internet the duty to create and submit reports on due diligence and human rights impacts analysis; and it proposed that companies that could show good standing in initiatives like the GNI be exempted from such new reporting obligations.[148] Internet regulation and concerns over the tech industry are far more complex today than they were in 2013. The scope of the oversight and assessment as well as the requirements or requisites for initiatives to be considered legitimate under state mandated standards should probably be different. Still, the idea of mandated self-regulation could be an interesting alternative that deserves further discussion and due consideration.

Finally, a co-regulatory approach like the one I am proposing would necessarily be complementary to other initiatives and mechanisms that probably need to be created or revised. State mechanisms to deal with illegal content in the digital age, for example, merit significant revisions. There remain serious challenges to access to justice within the different countries around the globe that have only been aggravated in the digital age. Inter-State mechanisms like MLATs merit revisions rather than patches or informal mechanisms to bypass them.

5. Conclusions

There is an international agreement that human rights apply online as they do offline, and there is a growing consensus that Internet companies (particularly) need to be regulated and overseen to guarantee such standards. There are, however, no agreements as to what that oversight should look like.

Amidst growing social demands and ICT-related scandals, some of the main regulatory and co-regulatory regulatory initiatives that were set off to exercise oversight over Internet companies have been ineffective and/or have not been

[148] See https://www.congress.gov/bill/113th-congress/house-bill/491.

properly evaluated. While recognizing and mandating cooperation — in the form of company reports, transparency, etc. — the envisioned mechanisms lack specificity, the ability to adapt to the ever-changing technological advances that impact the overseen parties, or agreed-upon indicators to facilitate a proper evaluation of compliance.

Self-regulatory initiatives so far have not had better luck. Although significant improvements have been made in transparency reporting, both of State requests and company actions, the reports are not audited and trust in these companies is at its lowest. Citing Gorwa, "Part of the challenge for standards-setting initiatives is that they do truly require collaboration: the received wisdom is that initiatives that are dominated by only one class of actors are unlikely to create meaningful reform (Fransen & Kolk, 2007). Single-actor schemes — whether they just involve NGOs or firms — often have limited long-run success, as they do not make compromises between stakeholders or bring on the right mix of competencies to actually make a difference."[149]

Faced with proliferating initiatives to oversee Internet companies in many varied aspects, the proposal for a co-regulatory mechanism, grounded on technical standards, led by independent stakeholders, capable of generating qualitative indicators of progress and focused on processes could render significant benefits and complement other much-needed transparency initiatives. Although this is not (and should not) be the only model to adopt, establishing a model that allows for the evaluation of the governance and procedures of ICT companies regarding the protection and promotion of human rights (not only in the reduced scope of GNI but also in others) would provide a much-needed basis for other initiatives striving for better accountability. The initiatives that may arise around specific practices could be nurtured and should be recognized as complementary so that the model is effectively sustainable.

[149] Gorwa, R, The platform governance triangle: conceptualising the informal regulation of online content, Internet Policy Review, 8(2), (2019). https://doi.org/10.14763/2019.2.1407.