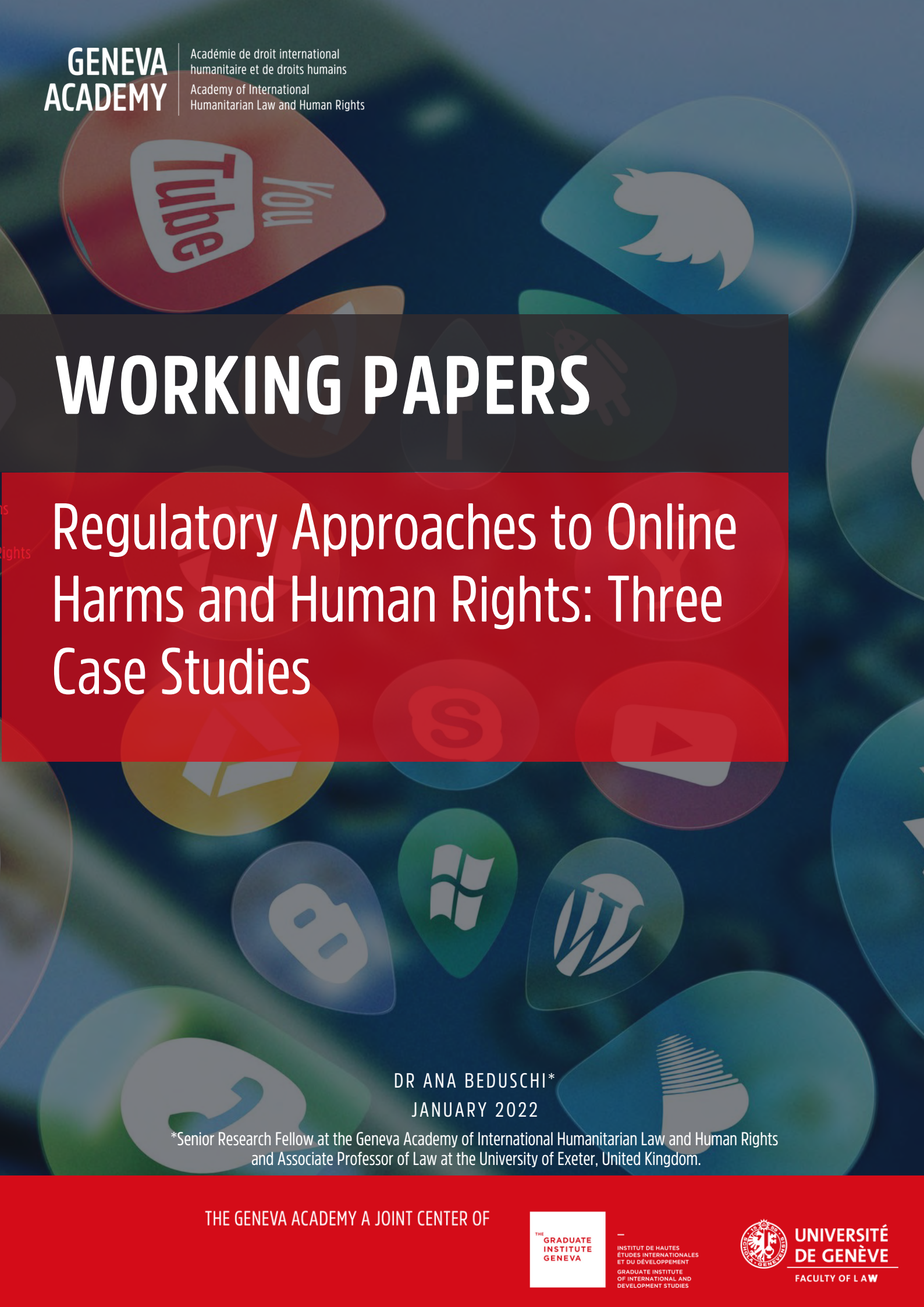
GENEVA ACADEMY
Académie de droit international humanitaire et de droits humains
Academy of International Humanitarian Law and Human Rights



WORKING PAPERS

Regulatory Approaches to Online Harms and Human Rights: Three Case Studies

DR ANA BEDUSCHI*

JANUARY 2022

*Senior Research Fellow at the Geneva Academy of International Humanitarian Law and Human Rights and Associate Professor of Law at the University of Exeter, United Kingdom.

THE GENEVA ACADEMY A JOINT CENTER OF

THE GRADUATE INSTITUTE GENEVA
INSTITUT DE HAUTES ÉTUDES INTERNATIONALES ET DU DÉVELOPPEMENT
GRADUATE INSTITUTE OF INTERNATIONAL AND DEVELOPMENT STUDIES

UNIVERSITÉ DE GENÈVE
FACULTY OF LAW

TABLE OF CONTENTS

I. INTRODUCTION

Online platforms and social media have become an integral part of the daily lives of millions of individuals worldwide. Concerns about the dissemination of illegal content via these platforms and disinformation and misinformation on social media have prompted States and International Organisations to seek to strengthen the regulation of online content.

Due to the significant impact that digital technologies and technology companies can have on the protection of human rights, these should be at the heart of any regulatory and policy frameworks concerning the design, development and deployment of these technologies.[1] The UN Secretary-General has prominently reinforced this message as he called on States 'to place human rights at the centre of regulatory frameworks and legislation on the development and use of digital technologies.'[2]

In particular, the respect of human rights by technology companies, such as online platforms and social media companies, is crucial for leveraging and fostering a rights-respecting technology ecosystem. In this regard, the UN Guiding Principles on Business and Human Rights (UNGPs) can serve as the basis for ensuring that advances in digital technologies are firmly anchored in respect for human rights.

In order to achieve a level playing field of rights-respecting conduct by technology companies, the role of the State in requiring companies to act responsibly is essential. The UNGPs call on States to adopt a "smart mix of measures" of national and international, mandatory and voluntary nature, to support and further implement the UNGPs.[3] This entails requiring businesses headquartered in their jurisdiction to respect human rights. That may be achieved through measures that incentivize companies to fulfil certain human rights obligations or via regulatory efforts requiring companies to disclose specific processes. It is the duty of the State to ensure that business implements appropriate measures to identify, address and mitigate adverse impacts stemming from, or being linked to their business activities, including in the technology sector[4].

Against this backdrop, this paper aims to address the overarching question of how human rights protection can be articulated within regulatory and legislative processes. To this end, the paper discusses the regulatory approaches to online harms, particularly regarding their impact on the right to freedom of expression. The paper examines the regulation of online harms in three jurisdictions – Brazil, the EU and the UK. It builds on a multidisciplinary review of legal, social science, humanities and technology-facing academic and professional literature and the analysis of primary and secondary sources of law. The paper is part of the project *Disruptive Technologies and Rights-based Resilience* – funded by the Geneva Science-Policy Interface – carried out by the Geneva Academy of International Humanitarian Law and Human Rights in partnership with the Office of the United Nations (UN) High Commissioner for Human Rights (OHCHR) B-Tech Project.[5]

The analysis is structured into three parts. First, the paper briefly introduces the human rights framework (section 2). Subsequently, the paper examines three case studies concerning proposals for regulation of online harms in Brazil, the EU and the UK (section 3). Finally, the paper draws conclusions on how to best place

[1] See for a comprehensive analysis of this area: Jonathan Andrew and Frédéric Bernard, Human Rights Responsibilities in the Digital Age: States, Companies and Individuals (Hart Publishing 2021).

[2] UN Secretary General, 'Report of the Secretary-General Roadmap for Digital Cooperation' (2020) < https://www.un.org/en/content/digital-cooperation-roadmap/assets/pdf/Roadmap_for_Digital_Cooperation_EN.pdf> accessed 29 November 2021.

[3] UN Guiding Principle 3, Commentary.

[4] UN OHCHR, B-Tech Project, 'Foundational Paper on the State Duty to Protect' (2021) <https://www.ohchr.org/Documents/Issues/Business/B-Tech/b-tech-foundational-paper-state-duty-to-protect.pdf> accessed 29 November 2021.

[5] See https://www.geneva-academy.ch/research/our-clusters/digitalization-and-new-technologies/detail/82-disruptive-technologies-and-rights-based-resilience accessed 29 November 2021.

human rights, particularly the UNGPs, at the centre of regulatory frameworks and legislation concerning online harms (section 4).

II. THE HUMAN RIGHTS FRAMEWORK

It is generally accepted that International human rights law (IHRL) applies in the digital space and that 'the same rights that people have offline must also be protected online.'[6]

At the UN level, the Universal Declaration on Human Rights and the nine core international human rights treaties form the legal framework of reference for human rights.[7] At the regional level, human rights treaties such as the European Convention on Human Rights, the American Convention on Human Rights and the African Charter on Human and People's Rights establish specific legal regimes.[8]

The UNGPs complement this framework, providing internationally agreed norms applicable to States and businesses as a soft law instrument with significant uptake by business and early process legitimacy through endorsement by key stakeholder groups ranging from academia, civil society, business and governments.[9] The following sub-sections provide a brief analysis of State obligations under IHRL and the UNGPs.

A. STATE OBLIGATIONS UNDER INTERNATIONAL TREATIES ON HUMAN RIGHTS

Under IHRL, State parties to international treaties on human rights owe obligations to the individuals who fall within their jurisdiction. As such, they must respect and ensure the legal rights set forth by these treaties.[10]

A State's jurisdiction is triggered when individuals find themselves in the State's territory or an area outside of the national

[6] UNGA Res 68/167, 21 January 2014, §2; See also Human Rights Council, 'The Promotion, Protection and Enjoyment of Human Rights on the Internet' UN Doc A/HRC/20/L.13, 29 June 2012; Human Rights Council, 'The Promotion, Protection and Enjoyment of Human Rights on the Internet' UN Doc A/HRC/32/L.20, 27 June 2016; M. N. Schmitt (ed), Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations (CUP 2017) 179.

[7] Universal Declaration of Human Rights (adopted 10 December 1948 UNGA Res 217 A(III) (UDHR); International Covenant on Civil and Political Rights (adopted 16 December 1966, entered into force 23 March 1976) 999 UNTS 171 (ICCPR); International Covenant on Economic, Social and Cultural Rights (adopted 16 December 1966, entered into force 3 January 1976) 993 UNTS 3 (ICESCR); International Convention on the Elimination of All Forms of Racial Discrimination (adopted 21 December 1965, entered into force 4 January 1969) 660 UNTS 195 (CERD); Convention on the Elimination of All Forms of Discrimination Against Women (adopted 18 December 1979, entered into force 3 September 1981) 1249 UNTS 13 (CEDAW); Convention Against Torture and Other Cruel, Inhuman or Degrading Treatment or Punishment (adopted 10 December 1984, entered into force 26 June 1987) 1465 UNTS 85 (CAT); Convention on the Rights of the Child (adopted 7 March 1990, entered into force 2 September 1990) E/CN.4/RES/1990/74 (CRC); International Convention on the Protection of the Rights of All Migrant Workers and Members of Their Families (adopted 18 December 1990, entered into force 1 July 2003) A/RES/45/158 (CMW); International Convention for the Protection of All Persons from Enforced Disappearance (adopted 20 December 2006, entered into force 23 December 2010) A/72/280 (CPED); Convention on the Rights of Persons with Disabilities Disappearance (adopted 13 December 2006, entered into force 3 May 2008) 2515 UNTS 3 (CRPD).

[8] Convention for the Protection of Human Rights and Fundamental Freedoms (European Convention on Human Rights) (adopted 4 November 1950, entered into force 3 September 1953, as amended) (ECHR); American Convention on Human Rights (adopted 22 November 1969, entered into force 18 July 1978) (ACHR); African Charter on Human and Peoples' Rights (adopted 27 June 1981, entered into force 21 October 1986) (African Charter).

[9] UN Human Rights Council, 'Report of The Special Representative of The Secretary-General on The Issue of Human Rights and Transnational Corporations and Other Business Enterprises, John Ruggie, on Guiding Principles on Business and Human Rights: Implementing the United Nations 'Protect, Respect and Remedy' Framework' (21 March 2011) UN Doc A/HRC/17/31 (UNGPs).

[10] UN Human Rights Committee, 'General Comment No. 31: The Nature of the General Legal Obligation Imposed on States Parties to the Covenant' (26 May 2004) UN Doc CCPR/C/21/Rev.1/Add, para 5; UN Committee on Economic Social and Cultural Rights, 'General Comment No. 3 The Nature of States Parties' Obligations (Art. 2, Para. 1 of the Covenant)' (14 December 1990) UN Doc E/1991/23, para 2.

territory but where that State exercises 'effective control.'[11] Additionally, States have jurisdiction over individuals who are within their State agents' authority and control.[12]

State obligations are 'both negative and positive in nature.'[13] That entails that States must not only refrain from violating individuals' rights but should also adopt all measures necessary to safeguard the effective respect of these rights. States have the choice concerning which measures they adopt — these can be legislative, judicial, administrative, or any other appropriate measures to fulfil their positive obligations.[14]

Positive obligations compel States to adopt necessary measures even when harm originates in actions or omissions committed by private persons or entities.[15] For instance, States may breach their positive obligations under IHRL when they fail to take appropriate measures or when they do not exercise due diligence to prevent, punish, investigate or redress the harm caused by third parties.[16]

The State duty to protect human rights against abuse by third parties includes harms caused by businesses. This standard of conduct is reaffirmed by the UNGPs, as analyzed below.

B. UN GUIDING PRINCIPLES ON BUSINESS AND HUMAN RIGHTS

The UNGPs offer a firm basis for developing regulatory and policy responses to AI technologies. They provide a set of internationally agreed norms for preventing,

Figure 1 – UNGPs pillar structure. Source: UN Human Rights B-Tech Project (2020)

[11] UN Human Rights Committee, 'General Comment No. 31 n(10), para 10; Legal Consequences of the Construction of a Wall in the Occupied Palestinian Territory (Advisory Opinion) 2004 ICJ 136, para 111; Loizidou v Turkey (Preliminary Objections) App no 15318/89 (ECtHR, 23 March 1995) para 62. See also Bruno Simma and Andreas T. Müller, 'Exercise and Limits of Jurisdiction' in James Crawford and Martti Koskenniemi (eds.) The Cambridge Companion to International Law (CUP 2012) 134-157.

[12] Al-Skeini v United Kingdom, App no 55721/07) (ECtHR, 7 July 2011) at para. 1130-142; Marko Milanovic, Extraterritorial Application of Human Rights Treaties (OUP 2011); Marko Milanovic, 'Al-Skeini and Al-Jedda in Strasbourg' (2012) 23 European Journal of International Law 121-139.

[13] UN Human Rights Committee, 'General Comment No. 31' n(10), para 6.

[14] UN Human Rights Committee, 'General Comment No. 31' n(10), para 7.

[15] Ibid paras 6-8; Airey v Ireland, App no 6289/73 (ECtHR, 9 October 1979) para 32; Marckx v Belgium, App no 6833/74 (EctHR, 13 June 1979) para 31.

[16] UN Human Rights Committee, 'General Comment No. 31' n(10), para 8.

addressing and remediating human rights violations concerning business operations, including in the technology sector.[17]

Proposed by the Special Representative of the Secretary-General on the Issue of Human Rights and Transnational Corporations and Other Business Enterprises, John Ruggie, the UNGPS were unanimously endorsed by the UN Human Rights Council in 2011.[18] The tripartite 'Protect, Respect and Remedy' framework has since become a landmark in business and human rights.[19]

The UNGPs are structured into three pillars. The first pillar reaffirms the State duty to protect human rights. The second pillar introduces the corporate responsibility to respect human rights. The third pillar sets forth the principles on access to remedy for victims of business-related human rights abuses. These sets of principles apply to all States and all businesses enterprises, small or large, operating in all sectors of activities.[20]

While they are not legally binding and thus differ from international human rights treaties, the UNGPs have an authoritative normative power. In particular, they 'have intrinsic persuasive power, inspire or justify prescribed conduct, engender shared expectations of ends and means.'[21]. In addition, they complement existing treaty-based mechanisms on human rights as they apply to businesses — as international treaty law creates obligations towards States primarily.[22]

State-led intergovernmental negotiations aiming at adopting a legally binding instrument to regulate, in international human rights law, the activities of transnational corporations and other business enterprises are currently underway.[23] This legally binding instrument has a broad scope, applying to businesses of all sizes and sectors of activity, thus including technology companies developing AI technologies.[24]

States are encouraged to lay out how they are implementing the UNGPs at the national level, such as in the format of so-called National Action Plans on Business and Human Rights (NAPs). NAPs are policy documents setting out commitments, priorities, and points of action to be adopted to promote the implementation of the UNGPs.[25] NAPs may encompass matters relating to the technology sector — for instance, ensuring that technology companies respect human rights when designing and developing AI systems. At the time of writing, a small number NAPs adopted by States worldwide refer to the technology sector.[26] Thus, there is considerable room for improving the breadth and depth of engagement with the UNGPs in the technology sector via NAPs, notably as fast-developing technologies such as AI impact a variety of human rights.[27]

---

[17] United Nations Human Rights Office of the High Commissioner (OHCHR), 'The UN Guiding Principles in the Age of Technology. A B-Tech Foundational Paper' (2020) OHCHR ‹ https://www.ohchr.org/Documents/Issues/Business/B-Tech/introduction-ungp-age-technology.pdf› accessed 29 November 2021.

[18] UNGPs n(9).

[19] John Gerard Ruggie, Caroline Rees and Rachel Davis, 'Ten Years After: From UN Guiding Principles to Multi-Fiduciary Obligations' (2021) Business & Human Rights Journal 1-19

[20] UNGPs n(9).

[21] Ruggie, Rees and Davis 'Ten Years After' n(19) at 2.

[22] But see Andrew Clapham, Human Rights Obligations of Non-State Actors (OUP 2006).

[23] UN Human Rights Council, 'Report on the sixth session of the open-ended intergovernmental working group on transnational corporations and other business enterprises with respect to human rights' (14 January 2021) UN Doc A/HRC/46/73.

[24] Legally Binding Instrument to Regulate, in International Human Rights Law, the Activities of Transnational Corporations and Other Business Enterprises, Second Revised Draft (06 August 2020) ‹https://www.ohchr.org/Documents/HRBodies/HRCouncil/WGTransCorp/Session6/OEIGWG_Chair-Rapporteur_second_revised_draft_LBI_on_TNCs_and_OBEs_with_respect_to_Human_Rights.pdf› accessed 29 November 2021, Article 3.

[25] UN Working Group on Business and Human Rights, Guidance on National Action Plans on Business and Human Rights (UN 2016).

[26] The Danish Institute for Human Rights, National Action Plans on Business and Human Rights (2021) ‹https://globalnaps.org/issue/information-communications-technology-ict-electronics/› accessed 29 November 2021.

[27] Richard Wingfield, Ioana Tuta and Tulika Bansal, 'The tech sector and national action plans on business and

When adopting, promoting and supporting the implementation of the UNGPs, States may embrace a variety of measures — a smart mix of measures to foster business respect for human rights, including in the technology sector.

States have a duty to protect against human rights abuses by 'taking appropriate steps to prevent, investigate, punish and redress such abuse through effective policies, legislation, regulations and adjudication.'[28] They should also set out clear expectations about business' respect for human rights.[29] In doing so, States are invited to adopt a smart mix of measures — national and international, mandatory and voluntary, to support and further the implementation of the UNGPs.[30]

Such a diversity of measures is even more important in the context of digital technologies. That is due to the fast pace of technological advances and the considerable impact these technologies can have on human rights.[31]

Often, legislative measures alone are not apt to prevent abuses as they lack robust enforcement and monitoring mechanisms that can keep pace with the advances in technology. At the same time, relying solely on corporate self-regulation without a solid normative framework may lead to feeble respect for human rights.

Accordingly, some States and regional organizations are increasingly leaning towards adopting more mandatory measures, both in the general context of business and human rights and the specific context of the application of certain digital technologies – for instance, regulation to prevent and mitigate online harms. Still, a calibrated mix of mandatory and voluntary measures is necessary. Fine-tuning this mix will help ensure the effectiveness of laws, regulations and policies, and promote corporate uptake and compliance.

Building on the existing human rights framework, the following section analyses three legislative proposals concerning online harms in three different jurisdictions worldwide.

III. CASE STUDIES: AN ANALYSIS OF ONLINE HARMS REGULATION IN BRAZIL, THE EU, AND THE UK

The following case studies seek to evaluate how different legislative proposals take human rights considerations into account, particularly the corporate responsibility to respect human rights embedded in the three pillars of the UNGPs. As discussed in the previous section, the State duty to protect human rights entails that States adopt a "smart mix" of measures requiring technology companies to respect human rights. That is, for instance, the case in the context of the recent online harms regulations in the European Union, the UK and Brazil.

The analysis focuses on the right to freedom of expression online. Throughout these case studies, the overarching question threading is how human rights protection can be articulated within regulatory and legislative processes. More specifically, we investigate:

- Whether legislative proposals put forward substantive or procedural rules (or both), and what are the key challenges and opportunities for each of these approaches.
- Whether there are potential reverse negative impacts for the protection of

human rights. A thematic supplement to the "national action plans on business and human rights toolkit 2017 edition"' (2020) The Danish Institute for Human Rights.

[28] UN Guiding Principle 1.

[29] UN Guiding Principle 2.

[30] UN Guiding Principle 3, Commentary.

[31] Murray, Daragh, 'Using Human Rights Law to Inform States' Decisions to Deploy AI' (2020) 114 AJIL Unbound 158-162; Daragh Murray, Pete Fussey, Lorna McGregor and Maurice Sunkin, 'Effective Oversight of Large-Scale Surveillance Activities: A Human Rights Perspective' (2021) 11 Journal of National Security Law and Policy 1-25.

freedom of expression.

A. BRAZIL: 'FAKE NEWS' BILL AND BAN ON SOCIAL MEDIA CONTENT MODERATION BILL

The Brazilian Internet Freedom, Responsibility, and Transparency Bill, commonly known as the "Fake News" Bill, was proposed in 2020 and since approved by one of the Chambers of Congress (the Senate).32 The bill is still with the lower chamber (the Chamber of Deputies) for approval pending debates at the time of writing. The bill aims to make social media platforms responsible for content moderation with a view of fighting misinformation and disinformation. That relates to the State duty to protect human rights by adopting measures requiring these platforms to respect human rights in the digital space.

The bill is concerned with providing mechanisms to increase the transparency of social media platforms, including private messaging on the Internet. For instance, it establishes an obligation of reporting items, including the number of false accounts removed from the platforms, the number of artificial accounts (bots) present in the platform, the number of complaints received, and the time lapsed between the receiving a complaint by a user and resolving it.[33] The bill also provides that social media platforms should facilitate data access for academic research purposes.[34]

The bill lists a variety of obligations for social media platforms, including specific obligations for private messaging services such as WhatsApp.[35] These are widely used in Brazil and have been a key avenue for disseminating disinformation during the last presidential elections and the current pandemic crisis.[36] While the bill establishes various procedural obligations, such as reporting obligations, it contains several important substantive obligations. These include data retention,[37] the identification of the nature of content as paid or promoted, and the identification of the account paying for the diffusion of the content.[38]

On its face, the bill is concerned with human rights themes such as the right to privacy. [39] However, commentators have highlighted that, paradoxically, the "Fake News" Bill poses important risks to privacy and freedom of speech rights.[40] A key point of criticism of the bill is the lack of definition of false content, disinformation and misinformation that are otherwise central to the bill's implementation.[41] The bill leaves

[32] Projeto de Lei nº 2630/2020. Institui a Lei Brasileira de Liberdade, Responsabilidade e Transparência na Internet PL 2630/2020 ("Fake News" Bill) [Hereinafter Bill 2630/2020].

[33] Article 13 Bill 2630/2020.

[34] Article 13 (6) Bill 2630/2020.

[35] Articles 9-11 Bill 2630/2020.

[36] Luca Belli, 'WhatsApp skewed Brazilian election, showing social media's danger to democracy' (2018) The Conversation available at https://theconversation.com/whatsapp-skewed-brazilian-election-showing-social-medias-danger-to-democracy-106476 (accessed 7 January 2022); Rafael Evangelista and Fernanda Bruno, 'WhatsApp and political instability in Brazil: targeted messages and political radicalisation' (2019) Internet Policy Review 1-23; Cláudia Pereira Galhardi, Neyson Pinheiro Freire, Maria Cecília de Souza Minayo, Maria Clara Marques Fagundes 'Fact or Fake? An analysis of disinformation regarding the Covid-19 pandemic in Brazil' (2020) 25 Ciência & Saúde Coletiva 4201-4210.

[37] Article 10 Bill 2630/2020.

[38] Articles 14-17 Bill 2630/2020.

[39] Mentioned in Articles 14 and 16 of the Bill 2630/2020.

[40] Raphael Tsavkko Garcia, 'Brazil's "fake news" bill won't solve its misinformation problem' (2020) MIT Technology Review available at https://www.technologyreview.com/2020/09/10/1008254/brazil-fake-news-bill-misinformation-opinion/ (accessed 7 January 2022); Namrata Maheshwari and Greg Nojeim, 'Update on Brazil's Fake News Bill: The Draft Approved by the Senate Continues to Jeopardize Users' Rights' (2020) Centre for Democracy & Technology available at https://cdt.org/insights/update-on-brazils-fake-news-bill-the-draft-approved-by-the-senate-continues-to-jeopardize-users-rights/ (accessed 7 January 2022); Veridiana Alimonti, 'Brazil's Fake News Bill: Perils and Flaws of Expanding Existent Data Retention Obligations' (2021) Electronic Frontier Foundation available at https://www.eff.org/deeplinks/2021/11/brazils-fake-news-bill-perils-and-flaws-expanding-existent-data-retention (accessed 7 January 2022).

[41] Article 4 (II) of the initial proposal - Bill 2630/2020 (removed from the version approved by the Senate).

considerable leeway for further regulation to define these concepts – including secondary legislation emanating directly from the executive.

Recently, the Brazilian President signed a provisional measure on the 6th of September 2021 aiming to ban social media platforms from de-platforming users and removing various types of content, including those relating to COVID-19 misinformation.[42] Reportedly, social media accounts linked to current government officials have been central to spreading misinformation and disinformation, for instance, in the context of the Covid-19 pandemic.[43]

The provisional measure had immediate binding legal force, albeit it needed approval by the two chambers of Congress to become permanent law.[44] The President of the two chambers later rejected the provisional measure, contesting the measure's constitutionality.[45] Following that, the government proposed a new Bill no 3227/2021, which fully reproduces the text of the rejected provisional measure.[46]

While the "Fake News" Bill proposed to establish a legal basis for content moderation, it did not fully define illegal content.[47] The bill remains unclear about content moderation concerning misinformation and disinformation. Members of the Brazilian Parliament will have to clarify the concepts and the obligations relating to content moderation, notably as the more recent Bill 3227/2021 seems to disregard and, to a certain degree, jeopardize the obligations set forth by the "Fake News" Bill.

Bill 3227/2021 prescribes a substantive and exhaustive list of content that social media platforms may remove. It considerably limits the cases in which content may be moderated and removed – misinformation is not considered as a valid reason for content removal.[48] In this regard, this bill defies the objective and purpose of the "Fake News" Bill and may encourage the continuous dissemination of false content.

Finally, the "Fake News" Bill proposes an array of sanctions, ranging from warnings to financial penalties, a temporary suspension of social media platforms' activities in the country, and the prohibition of exercising its activities.[49] These are additional to any criminal, civil or administrative liability provided for by any other laws applicable to these technology companies.[50] Judicial authorities would be in charge of imposing the sanctions.[51] The bill does not provide specific details about scrutiny, monitoring and enforcement procedures. That is not the case in the context of the EU's Digital Services Act, as discussed below.

B. EU: DIGITAL SERVICES ACT LEGISLATIVE PROPOSAL

In December 2020, the European Commission proposed a Digital Services Act package

[42] Medida Provisória nº 1.068, 6 September 2021, DOU169-A p. 1.

[43] Felipe Bonow Soares, Raquel Recuero, Taiane Volcan, Giane Fagundes, Giéle Sodré, 'Research note: Bolsonaro's firehose: How Covid-19 disinformation on WhatsApp was used to fight a government political crisis in Brazil' (2021) 2 The Harvard Kennedy School Misinformation Review 1-13; Bryan Harris, 'Spread of fake news adds to Brazil's pandemic crisis' Financial Times (London, 13 July 2020); Vanessa Barbara, 'Miracle Cures and Magnetic People. Brazil's Fake News Is Utterly Bizarre' New York Times (New York City, 5 July 2021); Freedom House, Joint Statement. Brazil: Disinformation Bill Threatens Freedom of Expression and Privacy Online (29 June 2020) available at https://freedomhouse.org/article/brazil-disinformation-bill-threatens-freedom-expression-and-privacy-online (accessed 7 January 2022).

[44] Article 62, Brazilian Federal Constitution.

[45] Agência Senado, Pacheco devolve MP que dificultava retirada de conteúdo da internet (14 September 2021) available at https://www12.senado.leg.br/noticias/materias/2021/09/14/pacheco-devolve-mp-que-dificultava-retirada-de-conteudo-da-internet/#conteudoPrincipal (accessed 7 January 2022).

[46] Projeto de Lei nº 3227/2021, 20 September 2021 [Hereinafter Bill 3227/2021].

[47] Article 12 (2) Bill 2630/2020.

[48] Article 8-C Bill 3227/2021.

[49] Article 28 (I-IV) Bill 2630/2020.

[50] Article 28 Bill 2630/2020.

[51] Ibid.

containing two legislative proposals – the Digital Services Act (DSA) and the Digital Markets Act (DMA).[52] Both legislative proposals have the explicit aim to protect fundamental rights while levelling the playing field for businesses operating in digital services.[53] The DSA seeks to create a common set of rules on intermediary services obligations and accountability. The DMA aims to regulate the behaviour of large online platforms that act as gatekeepers in the digital markets. Both proposals are currently being discussed at the European Parliament and the Council and should be adopted according to the ordinary legislative process.[54] While there are overlaps between the DSA and the DMA concerning the regulation of large online platforms, the analysis in this working paper focuses on the DSA only.

The current version of the proposed DSA applies to intermediary services, classified according to their function and size. These are intermediary services offering network infrastructure, hosting services, online platforms, and very large online platforms (reaching more than 10% of 450 million users in Europe). The DSA applies to intermediary services provided to recipients established or resident in the EU, regardless of the place of establishment of the providers of these services.[55]

A key innovative aspect of the DSA lies in the imposition of layered due diligence obligations on intermediary services[56] according to their function and size. All intermediary services have a set of basic obligations.[57] Additional sets of obligations are cumulatively superposed for hosting services,[58] online platforms[59] and very large online platforms.[60] Such obligations concern, for example, transparency reporting (basic or enhanced according to the type and size of intermediary service),[61] terms and conditions

[52] Proposal for a Regulation of the European Parliament and of the Council on a Single Market for Digital Services (Digital Services Act) and amending Directive 2000/31/EC (COM/2020/825 final) [hereinafter 'DSA']; Proposal for a Regulation of the European Parliament and of the Council on contestable and fair markets in the digital sector (Digital Markets Act) (COM/2020/842 final) [hereinafter 'DMA'].

[53] In the EU, fundamental rights are guaranteed by the Charter of Fundamental Rights of the EU, which provides a catalogue of human rights similar to those provided by the European Convention on Human Rights or the International Covenant on Civil and Political Rights. See Paul Lemmens, 'The Relation between the Charter of Fundamental Rights of the European Union and the European Convention on Human Rights – Substantive Aspects' (2001) 8 Maastricht Journal of European and Comparative Law 49-67; Koen Lenaerts, 'Exploring the Limits of the EU Charter of Fundamental Rights' (2012) 8 European Constitutional Law Review, 375-403.

[54] See notably European Parliament, Opinion of the Committee on Civil Liberties, Justice and Home Affairs for the Committee on the Internal Market and Consumer Protection on the proposal for a regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC, 28 September 2021 (PE692.898v07-00) (2020/0361(COD)); European Parliament, Opinion of the Committee on Legal Affairs for the Committee on the Internal Market and Consumer Protection on the proposal for a regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC, 11 October 2021 (PE694.960v02-00) (2020/0361(COD)); Council of the European Union, Presidency compromise text on Chapters I and III, with respective recitals, 4 June 2021 (9288/21) (2020/0361(COD)); Council of the European Union, Presidency compromise text on Chapters I and II, with respective recitals, 2 September 2021 (11459/21) (2020/0361(COD)).

[55] Article 1 (3) DSA.

[56] In the original DSA proposal, intermediary service is understood as corresponding to one of the following services: "a 'mere conduit' service that consists of the transmission in a communication network of information provided by a recipient of the service, or the provision of access to a communication network; a 'caching' service that consists of the transmission in a communication network of information provided by a recipient of the service, involving the automatic, intermediate and temporary storage of that information, for the sole purpose of making more efficient the information's onward transmission to other recipients upon their request; a 'hosting' service that consists of the storage of information provided by, and at the request of, a recipient of the service." (Article 2 (1) f DSA). The Council proposed to add another category of services to this list, to include online search engines (Council of the European Union, Presidency compromise text on Chapters I and II, n(54) 79).

[57] Articles 10-13 DSA.

[58] Articles 14-15 DSA.

[59] Articles 14-24 DSA.

[60] Articles 25-33 DSA.

[61] Article 13 and 33 DSA.

respecting fundamental rights,[62] complaint handling,[63] out of court dispute settlement,[64] notice and action procedures,[65] risk assessment and mitigation,[66] independent audits.[67]

The DSA is mainly concerned with processes and systems for checks and balances anchored in the respect for fundamental rights as recognized in EU law.[68] For instance, the DSA aims to tackle illegal content online without fully defining it.[69] Instead, it imposes mechanisms for dealing with illegal content such as transparency reporting[70] and statement of the reasons[71] when illegal content is removed or disabled, third party notification,[72] and a system of priority treatment of notices submitted by entities that are granted trusted flagger status.[73]

Such an approach is in line with the principle of subsidiarity and the objective of harmonizing EU law in this area. Defining illegal content is a complex matter that often requires a contextual approach and the balancing of different rights and interests at stake.[74] Removal of illegal content risks leading to censorship in certain cases.[75] There is certainly the possibility of leaving to domestic laws to determine what is considered illegal content (such as child pornography materials). Yet, this approach is not exempt from criticisms. For instance, content such as disinformation may be considered illegal in some Member States and not in others.[76] That can lead to an asymmetry in what is required from online platforms in the EU as they may have to remove content that in some national contexts is considered illegal, whereas that would not be the case in the other Member States.[77]

However, while the respect for fundamental rights underpins the DSA, commentators have highlighted that some of the provisions in the DSA may lead to negative impacts on these rights.[78] For instance, the obligations on very large platforms to undertake a risk assessment

[62] Article 12 DSA.

[63] Article 17 DSA.

[64] Article 18 DSA.

[65] Article 14 DSA.

[66] Articles 26 and 27 DSA.

[67] Article 28 DSA.

[68] Fundamental rights recognised by the Charter of Fundamental Rights of the EU and by the General Principles of EU law. See Charter of Fundamental Rights of the European Union [2000] OJ C 364/01; Takis Tridimas, The General Principles of EU Law (OUP 2007).

[69] Illegal content is defined in the DSA as "any information, which, in itself or by its reference to an activity, including the sale of products or provision of services is not in compliance with Union law or the law of a Member State, irrespective of the precise subject matter or nature of that law" (Article 2 (1) (g) DSA).

[70] Articles 13 and 23 DSA.

[71] Article 15 DSA.

[72] Article 14 DSA.

[73] Article 19 DSA.

[74] Barrie Sander, 'Freedom of Expression in the Age of Online Platforms: The Promise and Pitfalls of a Human Rights-Based Approach to Content Moderation' (2020) 43 Fordham International Law Journal 966-969; Barrie Sander, 'Democratic Disruption in the Age of Social Media: Between Marketized and Structural Conceptions of Human Rights Law' (2021) 32 European Journal of International Law 159-193; Article 19, Internet companies alone can't prevent online harms (2020) available at https://www.article19.org/resources/internet-companies-alone-cant-prevent-online-harms/ (accessed 10 November 2021); Jacob Berntsson and Maygane Janin, 'Online Regulation of Terrorist and Harmful Content' (2021) Lawfare, available at https://www.lawfareblog.com/online-regulation-terrorist-and-harmful-content?s=09 (accessed 7 January 2022).

[75] Tarleton Gillespie, Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media (Yale University Press 2018) 176.

[76] European Regulators Group for Audiovisual Media Services, Notions of Disinformation and Related Concepts (2021) available at https://erga-online.eu/wp-content/uploads/2021/03/ERGA-SG2-Report-2020-Notions-of-disinformation-and-related-concepts-final.pdf (accessed 10 November 2021).

[77] Ronan Ó Fathaigh, Natali Helberger and Naomi Appelman, 'The perils of legally defining disinformation' (2021) 10 Internet Policy Review 4, 1-25.

[78] Johann Laux, Sandra Wachter, Brent Mittelstadt, Taming the few: Platform regulation, independent audits, and the risks of capture created by the DMA and DSA' (2021) 43 Computer Law & Security Review 1-12; Access Now, Joint Statement of the Digital Services Act Human Rights Alliance (21 October 2021) available at https://www.accessnow.org/cms/assets/uploads/2021/10/Digital_Services_Act_Human_Rights_Alliance_Statement.pdf (accessed 10 November 2021); Ilaria Buri and Joris van

and put in place mitigation strategies[79] is deemed problematic as it would require these platforms 'to act as if they were law enforcement authorities.'[80] In other words, that would entail the risk that online platforms would over-remove content that would otherwise be perfectly legal. According to the current version of Article 26 of the DSA, very large online platforms must identify, analyze and assess any significant systemic risks posed within the functioning and use of their services. Such systemic risks include, for example, the dissemination of illegal content.[81] These platforms must also put in place risk mitigation measures, including adapting their content moderation systems.[82] Due to the large-scale operations entailed, very large online platforms may opt for adopting automated content moderation algorithms.[83] These algorithms may remove controversial but not illegal content to ensure compliance with this obligation. Overzealous platforms may also decide to ban provocative individuals from their platforms. In both cases, actions adopted by these platforms may lead to arbitrary restrictions of lawful expression of opinions online.

In any case, the DSA provides for a comprehensive system of general guidelines to be issued by the Commission in cooperation with the Digital Services Coordinators from the different Member States.[84] It is hoped that these guidelines could provide best practices informing, for instance, the ways in which online platforms can comply with their legal obligations without falling for the temptation of over-removing content that would otherwise be the legal expression of opinions and ideas of their users. Such an information-sharing system may become vital notably as the activities of intermediary services can have important consequences for the protection of human rights, such as the right to privacy and freedom of expression online.

Regarding the types of liability resulting from non-compliance and corresponding sanctions, the DSA is mainly concerned with civil liability. The Commission can impose fines on very large online platforms for intentional or negligent non-compliance.[85] Periodic penalty payments may also be imposed to a level of up to 5 % of the average daily turnover in the preceding financial year per day.[86] In addition to these actions that can be taken by the Commission, in case of non-compliance, users can directly bring a complaint to the platforms,[87] use out-of-court dispute settlement bodies,[88] or seek redress before courts.

Finally, the DSA provides for the appointment of a "Digital Services Coordinator" at the Member State level to oversee the enforcement of the regulation.[89] It sets forth that the "European Board for Digital Services", an independent advisory group, would also contribute to the guidance and consistency of the regulation's application.[90] At the level of the Member States, specific laws should be adopted to specify any specific penalties for non-

---

Hoboken, 'The DSA Proposal's Impact on Digital Dominance' (2021) VerfBlog, available at https://verfassungsblog.de/power-dsa-dma-01/ (accessed 10 November 2021); Valentina Golunova and Juncal Montero Regules, The Digital Services Act and freedom of expression: triumph or failure? (2021) Digital Society Blog, available at https://www.hiig.de/en/the-digital-services-act-and-freedom-of-expression-triumph-or-failure/ (accessed 10 November 2021).

[79] Articles 26 and 27 DSA.

[80] Pieter Van Cleynenbreugel, 'The Commission's digital services and markets act proposals: First step towards tougher and more directly enforced EU rules?' (2021) Maastricht Journal of European and Comparative Law 1-20, 9.

[81] Article 26 (1) (a) DSA.

[82] Article 27 (1) (a) DSA.

[83] Robert Gorwa, Reuben Binns, and Christian Katzenbach, 'Algorithmic content moderation: Technical and political challenges in the automation of platform governance' (2020) Big Data & Society 1-15; Tarleton Gillespie, 'Content moderation, AI, and the question of scale' (2020) Big Data & Society 1-5.

[84] Article 27 DSA.

[85] Article 59 DSA.

[86] Article 60 DSA.

[87] Article 17 DSA.

[88] Article 18 DSA.

[89] Articles 38-46 DSA.

[90] Articles 47-49 DSA.

compliance with the legal obligations. That may include the temporary suspension of services of platforms that would consistently refuse to comply with their obligations. The supervision, monitoring and enforcement of obligations concerning very large online platforms follow specific rules set forth in the DSA.[91] The European Commission would play a central role in this regard by having investigatory powers and the possibility to impose fines.

Such a regulatory system bears some similarities with that of the UK's Online Safety Bill, discussed below.

C. UK: DRAFT ONLINE SAFETY BILL

The Draft Online Safety Bill aims to establish a new regime for regulated internet services and provide new powers to the Office of Communications (OFCOM) to act as its main regulator.[92] It imposes a duty of care on providers of regulated services concerning illegal and legal but harmful content.[93]

Regulated services are user-to-user services (e.g. Twitter or Facebook) or search services (e.g. Google or Bing) that have links to the UK due to the service having a significant number of users in the UK or the UK users forming one of the target markets for the service.[94] These links can also be found if the service is capable of being used in the UK and there is a material risk of significant harm to individuals in the UK arising from content present on user-to-user services or encountered in or via search results.[95] Exemptions can be nevertheless found in schedule 1 of the Draft Online Safety Bill and include, for example, services that only enable user-generated content in the form of emails.[96]

As with the DSA, the Draft Online Safety Bill is also process-oriented. It encompasses a variety of duties, including reporting and redress and record-keeping duties.[97] Regulated services also have risk assessment and safety duties.[98] These include risk assessments and safety duties for illegal content such as those related to terrorism and child sexual exploitation and abuse.[99] These duties are in line with the State duty to protect human rights and to ensure that companies, in this case, those in the technology sector, respect human rights as per the UNGPs.

The draft bill further breaks down the obligations according to the type of users – children or adults for user-to-user services and children for search services.[100] Safety duties concerning children include, for example, the obligation of user-to-user services to 'take proportionate steps to mitigate and effectively manage risks of harm' to children identified during the risk assessment process due to harmful content available on the service.[101] In this case, content might be legal but still harmful to children – the draft bill does not exhaustively define such type of content. It is plausible that such can be the case of content on matters relating, for example, to self-harm or eating disorders which have a great potential to harm children in different age groups.[102]

More controversially, the Draft Online Safety Bill proposes to impose risk assessment and

[91] Articles 50-66 DSA.

[92] Draft Online Safety Bill Presented to Parliament by the Minister of State for Digital and Culture by Command of Her Majesty, May 2021, CP 405 [Hereinafter 'Draft Online Safety Bill'].

[93] Part 2, Draft Online Safety Bill.

[94] Section 3 (4) and (5), Draft Online Safety Bill.

[95] Section 3 (6), Draft Online Safety Bill.

[96] Section 1, Schedule 1, Draft Online Safety Bill.

[97] Sections 15 and 6, Draft Online Safety Bill (user-to-user services) and Section 24 and 25, Draft Online Safety Bill (search services).

[98] Sections 7 and 9, Draft Online Safety Bill (user-to-user services) and Section 19 and 21, Draft Online Safety Bill (search services).

[99] Section 7 (8) and Section 9, Draft Online Safety Bill (user-to-user services) and Section 19 (3) and Section 21, Draft Online Safety Bill (search services).

[100] Section 7 (3, 6, 7) and Section 10, Draft Online Safety Bill (user-to-user services) and Section 19 (2) and Section 22, Draft Online Safety Bill (search services).

[101] Section 10 (2), Draft Online Safety Bill.

[102] Rafe Jennings, 'Regulating content on user-to user and search service providers' (2021) UK Human Rights Blog available at https://ukhumanrightsblog.com/2021/08/02/regulating-

safety duties on user-to-user "category 1" services (i.e. largest platforms) for so-called "legal but harmful" content concerning adults.[103] They must identify potential risks and set out the strategies to mitigate such risks in their terms of service.

The draft bill does not fully define this type of content. "Priority" harmful content is to be later designated by the Secretary of State regulations.[104] Other types of legal but harmful content can be identified by a service provider through risk assessment if there are 'reasonable grounds to believe that the nature of the content is such that there is a material risk of the content having, or indirectly having, a significant adverse physical or psychological impact on an adult of ordinary sensibilities'.[105] By doing so, the Draft Online Safety Bill gives a considerable margin of appreciation to service providers to identify and remove such content from their platforms, potentially violating the freedom of expression of their users.

However, as highlighted by the House of Lords in a report from 2021, the Draft Online Safety Bill presents important risks for the protection of freedom of speech online.[106] The report clearly states that the members of the House of Lords 'are not convinced that [the duties imposed on platforms relating to legal but harmful content] are workable or could be implemented without unjustifiable and unprecedented interference in freedom of expression.'[107]

A significant risk relates to censorship by the over-removal of content by online platforms.[108] That can be due to their willingness to comply with their legal obligations, such as the duties of care under the Draft Online Safety Bill. Censorship can also arise from governments' demands to online platforms to remove content that is not in line with their political affinities.[109] While some large platforms have tried to resist such pressures in the past,[110] a significant risk to democracy and the protection of freedom of expression online persists.

Certainly, OFCOM will have greater powers as a regulator and could guide user-to-user services. For instance, OFCOM will issue codes of practice to support the services navigating their different duties.[111] That could be particularly helpful concerning the removal of high priority content such as those related to terrorism. Although these codes of practice will have an advisory nature, regulated services providers would likely use them as guidelines for ensuring compliance with their multiple and sometimes complex legal obligations. Therefore, these codes of practice will likely become an important asset for the regulatory framework in the UK. Even more so as the sanctions for failing to comply with the obligations outlined in the bill range from significant fines to criminal liability for

---

content-on-user-to-user-and-search-service-providers/ (accessed 10 November 2021).

[103] Section 7 (5-7) and Section 11, Draft Online Safety Bill.

[104] Section 46, Draft Online Safety Bill.

[105] Section 46 (3), Draft Online Safety Bill.

[106] House of Lords, Communications and Digital Committee, Free for all? Freedom of expression in the digital age (2021) HL Paper 54.

[107] Ibid at para. 182.

[108] David Kaye, Speech Police. The Global Struggle to Govern the Internet (Columbia Global Reports 2019) at 113; Alexander Brown, Models of Governance of Online Hate Speech (Council of Europe 2020) at 24; Daphne Keller, Empirical evidence of "over-removal" by internet companies under intermediary liability laws (2020) Center for Internet and Society at Stanford Law School, available at http://cyberlaw.stanford.edu/blog/2015/10/empirical-evidence-over-removal-internet-companies-under-intermediary-liability-laws (accessed 7 January 2022);

[109] David Kaye, Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression (2018) U.N. Doc. A/HRC/38/35 at 6; Gregory P. Magarian, 'The Internet and Social Media' in Adrienne Stone and Frederick Schauer (eds) The Oxford Handbook of Freedom of Speech (OUP 2021) 350-368 at 357; Eliza Bechtold and Gavin Phillipson, 'Glorifying Censorship? Anti-Terror Law, Speech, and Online Regulation' in Adrienne Stone and Frederick Schauer (eds) The Oxford Handbook of Freedom of Speech (OUP 2021) 519-541 at 538.

[110] Twitter, Updates on our response to blocking orders from the Indian Government (10 February 2021) available at https://blog.twitter.com/en_in/topics/company/2020/twitters-response-indian-government (accessed 7 January 2022).

[111] Chapter 5, Draft Online Safety Bill.

senior managers in some circumstances[112] and service blocking orders.[113]

Accordingly, due to the impact that regulated services can have on human rights, it is crucial that OFCOM consult with a wide range of stakeholders, including civil society organizations and academia, in a comprehensive and meaningful manner.[114]

IV. CONCLUSION

The analysis of the three case studies showed that legislative measures laying down substantive and process-oriented obligations for online platforms might support legal certainty. Nonetheless, paradoxically, depending on how these legal obligations are set, they may lead to potential violations of the very rights that they seek to protect.

That is, for instance, the case of imposing risk assessment and mitigation obligations to large online platforms, which may lead to over-removal of legal content – thus potentially violating the lawful expression of opinions online. Similarly, the lack of a clear definition about what type of content is considered illegal, together with the establishment of controversial categories of content that are "legal but harmful", may also lead to confusion and potential violation of users' freedom of expression.

International human rights law (IHRL) may provide a general framework for human rights in the context of content moderation obligations, albeit with some limitations.[115] For instance, the European Cout of Human Rights (ECtHR) has consistently held that opinions that disturb, shock or offend part of a population are not necessarily illegal, although the expression of such opinions may be restricted according to the terms of Article 10, paragraph 2 of the European Convention on Human Rights (ECHR).[116] However, a contextual case-by-case analysis is required, limiting the reach and usefulness of this jurisprudence as a guide.[117] In addition, the ECtHR has not always followed the standards it set for different types of speech. For instance, in the context of speech concerning terrorism, the court appears to have difficulties setting clear and consistent principles.[118]

Moreover, direct regulation of online content is even more challenging due to the transnational character of online speech.[119] The sheer scale of online communications also renders content moderation even more difficult. Tighter regulation may encourage large platforms to adopt algorithmic content moderation, which brings a variety of issues, including undue removal of legal content, bias and potential for discrimination of users on the basis, for example, of race, sex or gender.[120]

---

[112] Section 73, Draft Online Safety Bill.

[113] Sections 91-94, Draft Online Safety Bill.

[114] Section 29 (5)-(6), Draft Online Safety Bill.

[115] David Kaye, Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression (2018) U.N. Doc. A/HRC/38/35; Barrie Sander, 'Freedom of Expression in the Age of Online Platforms: The Promise and Pitfalls of a Human Rights-Based Approach to Content Moderation' (2020) 43 Fordham International Law Journal 966-969; Evelyn Douek, 'The Limits of International Law in Content Moderation' (2021) 6 UC Irvine Journal of International, Transnational, and Comparative Law 37-75.

[116] Handyside v UK App no 5493/72 (ECHR, 7 December 1976) at para. 49.

[117] Eliza Bechtold and Gavin Phillipson, 'Glorifying Censorship? Anti-Terror Law, Speech, and Online Regulation' in Adrienne Stone and Frederick Schauer (eds) The Oxford Handbook of Freedom of Speech (OUP 2021) 519-541 at 528-529.

[118] Ibid at 529.

[119] Gregory P. Magarian, 'The Internet and Social Media' in Adrienne Stone and Frederick Schauer (eds) The Oxford Handbook of Freedom of Speech (OUP 2021) 350-368 at 356.

Reuben Binns, Michael Veale, Max Van Kleek and Nigel Shadbolt, 'Like trainer, like bot? Inheritance of bias in algorithmic content moderation' (2017) International conference on social informatics 405-415; Robert Gorwa, Reuben Binns, and Christian Katzenbach, 'Algorithmic content moderation: Technical and political challenges in the automation of platform governance' (2020) Big Data & Society 1-15; Merlyna Lim and Ghadah Alrasheed, 'Beyond a technical bug: Biased algorithms and moderation are censoring activists on social media' (2021) The Conversation available at https://theconversation.com/beyond-a-technical-bug-biased-algorithms-and-moderation-are-censoring-

The study of the three examples of legislative proposals analyzed in this paper demonstrates that many challenges still lie ahead. There is no "one size fits all" approach for online harms regulation. Still, there is a crucial need for legislative proposals to consider the effects they may have on the protection and respect of human rights.

activists-on-social-media-160669 (accessed 7 January 2022).

The Geneva Academy of International Humanitarian Law and Human Rights

The Geneva Academy provides post-graduate education, conducts academic legal research and policy studies, and organizes training courses and expert meetings. We concentrate on branches of international law that relate to situations of armed conflict, protracted violence, and protection of human rights.

The Geneva Academy of International Humanitarian Law and Human Rights

Villa Moynier
Rue de Lausanne 120B
CP 1063 - 1211 Geneva 1 - Switzerland
Phone: +41 (22) 908 44 83
Email: info@geneva-academy.ch
www.geneva-academy.ch