

CEUB

EDUCAÇÃO SUPERIOR

ISSN 2237-1036

REVISTA DE DIREITO INTERNACIONAL
BRAZILIAN JOURNAL OF INTERNATIONAL LAW

**Closing the gap between
UNGPs and content regulation/
moderation practices**

**Fechando a lacuna entre o UNGPs
e as práticas de regulação/
moderação de conteúdo**

Sebastian Smart

Alberto Coddou McManus

VOLUME 19 • N. 2 • 2022
EXTRATERRITORIAL MECHANISMS, INTERNATIONAL
COOPERATION, AND PROTECTION OF VICTIMS
OF HUMAN RIGHTS VIOLATIONS

Sumário

CRÔNICAS.....	11
CRÓNICA REVISIÓN DE LAUDOS ARBITRALES DE INVERSIÓN 2020: 2º ENCUENTRO ANUAL (SANTIAGO DE CHILE, 07-08/06/2021).....	13
Nadia de Araujo, Marcelo De Nardi, Gustavo Ribeiro, Fabrício Polido, Inez Lopes e Matheus Oliveira	
CRÔNICA A RESPEITO DAS NEGOCIAÇÕES DO FUTURO TRATADO SOBRE A CONSERVAÇÃO E O USO SUSTENTÁVEL DA BIODIVERSIDADE MARINHA ALÉM DA JURISDIÇÃO (BBNJ): DESTAQUES DA 5ª ICG E DESAFIOS PARA A SUA CONCLUSÃO	43
Carina Costa de Oliveira, Bárbara Mourão Sachett, Júlia SchützVeiga, Philippe Raposo e Paulo Henrique Reis de Oliveira	
DOSSIÊ.....	50
André de Carvalho Ramos e Manoela Carneiro Roland	
A JURISDIÇÃO DE NECESSIDADE E O TRATADO VINCULANTE: A SAGA DO ACESSO TRANSNACIONAL À JUSTIÇA DAS VÍTIMAS DE ATIVIDADES DE EMPRESAS TRANSNACIONAIS.....	57
André de Carvalho Ramos e Manoela Carneiro Roland	
TRANSTERRITORIALITY AS A THEORY TO HOLD CORPORATIONS ACCOUNTABLE FOR HUMAN RIGHTS VIOLATIONS: THE APPLICATION OF ITS PRINCIPLES IN VEDANTA AND NEVSUN CASES	68
Ana Cláudia Ruy Cardia Atchabahian	
ACCESS TO JUSTICE THROUGH BUSINESS AND HUMAN RIGHTS: THE CHILEAN EXPERIENCE ON TRANSNATIONAL MINING.....	84
Daniel Jacomelli Hudler e Marcelo Benacchio	
MODEL INTERNATIONAL MOBILITY CONVENTION: AN INTER-AMERICAN SYSTEM OF HUMAN RIGHTS REFLECTION ON THE NON-CRIMINALIZATION PRINCIPLE.....	102
Lutiana Valadares Fernandes Barbosa e Ana Luisa Zago de Moraes	

EVOLUÇÃO DA PROTEÇÃO DAS MULHERES VÍTIMAS DE VIOLÊNCIA SEXUAL NA JURISPRUDÊNCIA DA CORTE INTERAMERICANA DE DIREITOS HUMANOS: INCORPORAÇÃO DA PERSPECTIVA DE GÊNERO.....	118
Ana Maria D'Ávila Lopes	
O CONTROLE DE CONVENCIONALIDADE COMO PERSPECTIVA FUTURA PARA A PROTEÇÃO DE DIREITOS DA POPULAÇÃO LGBTQIA+ EM NÍVEL GLOBAL.....	139
Dilermando Aparecido Borges Martins e Melina Girardi Fachin	
TEMAS GERAIS	156
EXTRATIVISMO E (NEO) COLONIZAÇÃO NA AMÉRICA LATINA: A RESPONSABILIDADE SOCIAL EMPRESARIAL NO ÂMBITO GLOBAL E REGIONAL.....	183
Larissa Ramina e Lucas Silva de Souza	
JURISDIÇÃO UNIVERSAL: “CAIXA DE PANDORA” OU UM CAMINHO PARA A REALIZAÇÃO DOS INTERESSES DA HUMANIDADE?	214
Claudia Regina de Oliveira Magalhães da Silva Loureiro	
A AMPLIAÇÃO DA JURISDIÇÃO INTERNACIONAL: O SURGIMENTO DE UMA JURISDIÇÃO INTERNACIONAL EM MATÉRIA PENAL.....	245
Elizabeth Goraieb e Paulo Emilio Vauthier Borges de Macedo	
CLOSING THE GAP BETWEEN UNGPs AND CONTENT REGULATION/MODERATION PRACTICES .	269
Sebastian Smart e Alberto Coddou McManus	
TEACHING AND RESEARCH OF INTERNATIONAL LAW IN AN EXPANDED WORLD: UNDERSTANDING FROM THE INDIAN PERSPECTIVE	295
Shuvro Prosun Sarker e Prakash Sharma	
LEGAL RESPONSE TO PROTECTION OF RIGHT TO COMMUNICATE E APPROPRIATE ADULTS DURING PROCESS OF ARREST OR DETENTION.....	314
Bassim Jameel Almusawi	
IS INVESTMENT FACILITATION A SUBSTITUTE OR SUPPLEMENT? A COMPARATIVE ANALYSIS OF CHINA AND BRAZIL PACTICES.....	326
Dan Wei e Ning Hongling	

AMPLIANDO A PROTEÇÃO SOCIAL AOS MIGRANTES À LUZ DA DIRETIVA DE PROTEÇÃO TEMPORÁRIA DA UNIÃO EUROPEIA: LIÇÕES DA INVASÃO DA UCRÂNIA.....344

Julia Motte-Baumvol, Tarin Cristino Frota Mont'alverne e Gabriel Braga Guimarães

RESENHA362

Lucas Carlos Lima

Closing the gap between UNGPs and content regulation/moderation practices*

Fechando a lacuna entre o UNGPs e as práticas de regulação/moderação de conteúdo

Sebastian Smart**

Alberto Coddou McManus***

Abstract

Both human-based and automated decisions are shaped by national legislation, international regulations as well as on platforms community guidelines or Terms of Service, a set of private rules that control what is allowed and what is not allowed on digital platforms. Such restrictions of what may or may not be deemed acceptable content are taking place in a context of increasing “digital authoritarianism” in which global internet freedom has declined for eleven consecutive years. Moreover, content moderation practices are currently shaped by corporate interests, within increasingly concentrated markets that are influenced by the network effects of the digital economy. Although it is well recognised that the same rights that people have offline must also be protected online, there are increasing examples of legislation and companies Terms of Service which are leading to restrictions on freedom of expression and other human rights. On the other hand, most of national regulations constitute fragmented regulatory attempts and do not touch upon several issues that could propel content moderation practices that could better respect human rights, such as design features or proceedings closer to responsive regulation models. In this paper we argue that, despite its limitations and its formal legal status, the United Nations Guiding Principles on Business and Human Rights (UNGPs) offer a good starting point to address some of these problems and encourage rules and procedures for moderating user-generated online content that put human rights at the very centre of those decisions. In the first section, we briefly describe and analyse the features of UNGPs that we think make them an adequate normative framework to address the human rights challenges that arise in the context of regulating and moderating user-generated content in the online world. We argue that, despite the limitations of an experimental and polycentric approach, the UNGPs have supported some progress, especially when it comes to the recognition of human rights standards in the tech sector. In the second section, we provide the reader with a brief introduction to ways in which content regulation/moderation has negatively affected the human right to freedom of expression. The following section addresses the relationship between UNGPs and the duties or roles that each actor could play in addressing the issue: governments, corporations, civil society organizations and individuals. Finally, we end with some concluding remarks and future research questions. Arguing that go-

* Recebido em 21/03/2022
Aprovado em 04/07/2022

** Doctor por la University College London, actualmente docente de la facultad de Derecho de la Universidad Austral de Chile.
Email: sebastian.smart@alumni.ucl.ac.uk

*** Doctor por la University College of London, actualmente profesor titular de la facultad de Derecho de la Universidad Austral de Chile.
Email: alberto.coddou@uach.cl

vernments have not made too much when it comes to business and human rights in the digital environment. On the other hand, the developments that come from companies have been mostly reactive and not preventive. In this scenario, the current wave of transnational and national regulatory attempts of online content needs to start from a human rights-based approach, and the UNGPs serve as a good starting point.

Keywords: human rights; technology; content moderation; business and human rights.

Resumo

Tanto as decisões baseadas em humanos quanto as automatizadas são moldadas pela legislação nacional, regulamentos internacionais, bem como pelas diretrizes da comunidade de plataformas ou Termos de Serviço, um conjunto de regras privadas que controlam o que é permitido e o que não é permitido em plataformas digitais. Essas restrições do que pode ou não ser considerado conteúdo aceitável estão ocorrendo em um contexto de crescente “autoritarismo digital”, no qual a liberdade global na Internet diminuiu por onze anos consecutivos. Além disso, as práticas de moderação de conteúdo são atualmente moldadas por interesses corporativos, em mercados cada vez mais concentrados e influenciados pelos efeitos de rede da economia digital. Embora seja bem reconhecido que os mesmos direitos que as pessoas têm off-line também devem ser protegidos online, há exemplos crescentes de legislação e termos de serviço de empresas que estão levando a restrições à liberdade de expressão e outros direitos humanos. Por outro lado, a maioria das regulações nacionais constituem tentativas regulatórias fragmentadas e não abordam diversos temas que poderiam impulsionar práticas de moderação de conteúdo que melhor respeitassem os direitos humanos, como características de design ou procedimentos mais próximos de modelos de regulação responsivos.

Neste artigo, argumentamos que, apesar de suas limitações e de seu status legal formal, os Princípios Orientadores sobre Empresas e Direitos Humanos (UNGPs) das Nações Unidas oferecem um bom ponto de partida para abordar alguns desses problemas e encorajar regras e procedimentos para moderar conteúdo online que coloca os direitos humanos no centro dessas decisões. De

fato, os UNGPs fornecem as bases para o desenvolvimento de uma estrutura normativa responsiva que seja consistente além das fronteiras, abordando problemas gerados por tentativas regulatórias fragmentadas e de corporações multinacionais operando em diversos contextos. Em uma área onde não há soluções fáceis ou claras, os UNGPs fornecem salvaguardas processuais para abordar coletivamente os problemas gerados na moderação de conteúdo. Além disso, o quadro normativo fornecido pelos UNGPs constitui a espinha dorsal para julgar os governos que estão adotando políticas públicas deficientes ou para abordar diferentes questões na ausência de regulamentação. Por fim, a maior inovação do UNGP é a criação de um dever independente de respeitar os direitos humanos para atores privados poderosos, como plataformas de mídia social que, em alguns casos, acumulam mais poder e capacidades institucionais do que uma parcela significativa dos governos em todo o mundo. Embora a literatura anterior já tenha adotado uma abordagem de direitos humanos para a moderação de conteúdo, tentamos aqui fornecer uma visão mais detalhada do papel específico que os UNGPs podem desempenhar ao enfrentar os desafios do que foi chamado de “indústria oculta” da moderação de conteúdo dentro “principalmente sistemas submersos de governança tecnológica”.

Na primeira seção, descrevemos e analisamos brevemente as características dos UNGPs que acreditamos torná-los uma estrutura normativa adequada para enfrentar os desafios de direitos humanos que surgem no contexto da regulamentação e moderação do conteúdo gerado pelo usuário no mundo online. Argumentamos que, apesar das limitações de uma abordagem experimental e policêntrica, os UNGPs têm apoiado alguns avanços, especialmente quando se trata do reconhecimento de padrões de direitos humanos no setor de tecnologia. Na segunda seção, fornecemos ao leitor uma breve introdução às maneiras pelas quais a regulamentação/moderação de conteúdo afetou negativamente o direito humano à liberdade de expressão. A seção seguinte aborda a relação entre os UNGPs e os deveres ou papéis que cada ator pode desempenhar ao lidar com o problema: governos, corporações, organizações da sociedade civil e indivíduos. Por fim, encerramos com algumas considerações finais e futuras questões de pesquisa.

Palavras-chave: Direitos Humanos, tecnologia, moderação de conteúdo, comércio e Direitos Humanos

1 Introduction

The internet is the most important global tool to access information, yet it can also be used for surveillance, to discriminate, to limit the right to freedom of expression and impinge on other fundamental rights. During the last 10 years we have seen an exponential growth in the usage of internet and access to content online; at the same time, governments and companies are deciding about the content that we can produce, share and access online. In 2012, journalist Adrian Chen published an article describing how outsourced workers removed content from online platforms that was otherwise considered acceptable, such as images of women breastfeeding or two men kissing¹. In recent years we have seen how platforms that moderate content online have increasingly used artificial intelligence to try to control the escalation in hate speech, disinformation and abuse that comes with the increasing use of internet. However, automated processes are extremely poor at making determinations relating to the nature of content given their inability to determine context, and the difficulties in defining terms such as “bullying” or “insult”². As the Facebook Papers revealed, algorithmic content moderation depends on integrity systems that are not well trained to discern between acceptable and non-acceptable forms of speech, especially in languages other than English³. Moreover, design features in several social media platforms have amplified speech in ways that run contrary to the idea of a digital public sphere: rather than debates that can subject to scrutiny the truth and opinion of a diverse audience, digital pla-

tforms have massified disinformation, reinforced division, impacted on mental health and affected privacy⁴.

Both human-based and automated decisions are shaped by national legislation, international regulations as well as on platforms’ community guidelines or Terms of Service, a set of private rules that control what is allowed and what is not allowed on digital platforms. Such restrictions of what may or may not be deemed acceptable content are taking place in a context of increasing “digital authoritarianism” in which global internet freedom has declined for eleven consecutive years⁵. Moreover, content moderation practices are currently shaped by corporate interests, within increasingly concentrated markets that are influenced by the network effects of the digital economy⁶. Although it is well recognised that the same rights that people have offline must also be protected online,⁷ there are increasing examples of legislation and companies’ Terms of Service which are leading to restrictions on freedom of expression and other human rights⁸. On the other hand, most of national regulations constitute fragmented regulatory attempts and do not touch upon several issues that could propel content moderation practices that could better respect human rights, such as design features or proceedings closer to responsive regulation models.

In this paper we argue that, despite its limitations and its formal legal status, the United Nations Guiding Principles on Business and Human Rights (UNGPs) offer a good starting point to address some of these problems and encourage rules and procedures for moderating

¹ CHEN, A. *Inside Facebook’s outsourced anti-porn and gore brigade, where ‘camel toes’ are more offensive than ‘crushed heads’*. Available at: <http://gawker.com/5885714/inside-facebooks-outsourced-anti-porn-and-gore-brigade-where-camel-toes-are-more-offensive-than-crushed-heads>. Accessed on: 20 Nov. 2021.

² DUARTE, N.; LLANSO, E.; LOUP, A. *Mixed messages? the limits of automated social media content analysis*. 2018. Available at: <https://apo.org.au/sites/default/files/resource-files/2017-11/apo-nid240471.pdf>. Accessed on: 16 Mar. 2022.

³ HAUGEN, F. *Formal meeting (oral evidence session): draft online safety bill*. London: UK Parliament, 2021. Available at: <https://committees.parliament.uk/event/5594>. Accessed at: 28 Oct. 2021.

⁴ UK PARLIAMENT. *Joint Committee on the Draft Online Safety Bill*. London, 2021. Available at: <https://publications.parliament.uk/pa/jt5802/jtselect/jtonlinesafety/129/12902.htm>. Accessed on: 10 Mar. 2022.

⁵ FREEDOM HOUSE. *Freedom on the Net 2021: the global drive to control big tech*. 2021. Available at: <https://freedomhouse.org/report/freedom-net/2021/global-drive-control-big-tech>. Accessed on: 15 Nov. 2021.

⁶ See the challenges address by the DIGITAL MARKETS ACT. EU COMMISSION. *Europe fit for the digital age: new online rules for platforms*. 2021. Available at: https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/digital-services-act-ensuring-safe-and-accountable-online-environment/europe-fit-digital-age-new-online-rules-platforms_en. Accessed on: 15 Mar. 2022.

⁷ UNITED NATIONS. Human Rights Council. *Report n. A/HRC/RES/32/13*. The promotion, protection and enjoyment of human rights on the Internet: resolution. Human Rights Council, 2016.

⁸ UNITED NATIONS. Human Rights Council. *Report n. A/HRC/38/35*. Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, David Kaye. Human Rights Council, 2018.

user-generated online content that put human rights at the very centre of those decisions. Indeed, UNGPs provide the foundations for the development of a responsive normative framework that is consistent across borders, addressing problems generated by fragmented regulatory attempts and of multinational corporations operating in diverse contexts. In an area where there are no easy or clear-cut solutions, UNGPs provide procedural safeguards to collectively address the problems generated in content moderation. Moreover, the normative framework provided by the UNGPs constitute the backbone against which to judge governments that are adopting deficient public policies or to address different issues in the absence of regulation. Lastly, UNGPs greatest innovation is the creation of an independent duty to respect human rights for powerful private actors such as social media platforms that in some cases accumulate more power and institutional capacities than a significant share of governments around the globe. Although previous literature has already adopted a human rights approach to content moderation,⁹ we here attempt to provide a closer look upon the specific role that UNGPs may play in addressing the challenges of what has been called the “hidden industry” of content moderation within “mostly submerged systems of technological governance”¹⁰.

In the first section, we briefly describe and analyse the features of UNGPs that we think make them an adequate normative framework to address the human rights challenges that arise in the context of regulating and moderating user-generated content in the online world. We argue that, despite the limitations of an experimental and polycentric approach, the UNGPs have supported some progress, especially when it comes to the recognition of human rights standards in the tech sector. In the second section, we provide the reader with a brief introduction to ways in which content regulation/moderation has negatively affected the

human right to freedom of expression. The following section addresses the relationship between UNGPs and the duties or roles that each actor could play in addressing the issue: governments, corporations, civil society organizations and individuals. Finally, we end with some concluding remarks and future research questions.

2 The origins and experimentalist governance of UNGPs

When the UNGPs were being drafted, the main sectors that grabbed attention were companies that operate through extended supply chains and extractive companies¹¹. Tech companies were usually considered only narrowly in the context of censorship and surveillance. In fact, during that period, and due to the occupy movement and the Arab Spring, tech companies, and the platforms they provided, were seen as enablers for democratic activism. During the Arab Spring, the internet and mobile technology was used to mobilise, organise, and campaign for political change. Some tech companies developed innovative tools to support dissidents, bloggers, and writers, such as encrypted technologies and ways to bypass restrictions by devising alternative routes, such as virtual private networks (VPNs). Tech companies were mostly seen as enablers of human rights, rather than threats to them¹². This narrative did not only affect the drafting of this international human rights instrument but created an environment where government regulations were seen as negative and counterproductive¹³.

According to John Ruggie, the late entrance of the tech sector within UNGPs’ discussions had to do with the low number of tech companies involved in allegations of adverse human rights impacts at that time. As of 2010, allegations were mainly targeted at the extractive sector, which accounted for 28 per cent of all such allegations. Retail and consumer products were not far behind; 20 per cent of allegations involved concerns around the long and complex supply chains used in the sector. The pharmaceutical and chemical industries’

⁹ ⁹ JØRGENSEN, R. F.; ZULETA, L. Private governance of freedom of expression on social media platforms: EU content regulation through the lens of human rights standards. *Nordicom Review*, Copenhagen, v. 41, n. 1, p. 51-67, 2020. LAIDLAW, E. B. (ed.). Myth or promise? the corporate social responsibilities of online service providers for human rights. In: TADDEO, M.; FLORIDI, L. *The responsibilities of online service providers*. Springer: Cham, 2017.

¹⁰ KLONICK, Kate. Law and technology: content moderation modulation. *Communications of the ACM*, [S.l.], v. 64, n. 1, p. 29-31, January 2021. Available at: <https://cacm.acm.org/magazines/2021/1/249450-content-moderation-modulation/full-text>. Accessed on: 15 Nov. 2021.

¹¹ RUGGIE, J. *Just Business: multinational corporations and human rights*. New York: Norton & Company, 2013.

¹² KHONDKER, H. H. Role of the New Media in the Arab Spring. *Globalizations*, v. 8 n. 5, p. 675-679, 2011.

¹³ COHEN, J. E. *Between truth and power*. [S.l.]: Oxford University Press, 2019.

third-place ranking reflected a combination of access to essential medicines and environmental hazards that impact the right to health. Infrastructure and utility companies, and the food and drinks industry also had a large number of allegations, mainly because of the impact on human rights of heavy water and fertiliser use. The tech sector had a small number of allegations, comprising only 5% of the total allegations studied by Ruggie and his team.¹⁴ The small number of allegations involving the Information and Communication Technology (ICT) sector can, to some extent, explain the lack of involvement of the industry during the discussions of the UNGPs. However, Ruggie also noted that due to user backlash when companies met demands from some governments, including during the Arab spring, to turn over user information or censor their services, the ICT sector started to be under scrutiny¹⁵.

In recent years, however, we have seen how companies and governmental activities in the ICT context have generated an increasingly adverse impact on human rights¹⁶. When it comes to content moderation, we have seen how governments have increasingly asked online platforms to remove content that they consider a threat. In other cases, governments have undertaken direct action to censor speech online, through imprisonment of political opponents or discrimination against minority groups¹⁷. On the opposite side, social media platforms have been negligent concerning their duties of care and contributed to human rights abuses. There are numerous examples of this side-effect on human rights, probably one of the latest and most notorious was the role of Facebook in Myanmar, where the company recognised that they had not done enough to help prevent their platform from being used to foment division and incite offline violence in the country.¹⁸ Ac-

ording to some authors, this lack of control resulted in Facebook contributing to the ethnic cleansing of that country's Muslim Rohingya population, including the forced displacement of over 700,000 people and as many as 25,000 killings.¹⁹

The growing awareness of adverse human rights impacts has resulted in more attention to the tech sector when it comes to business and human rights discussions. It has also led to the acknowledgement by investors, academics and governments of the growing governance gap between ICT practices and human rights standards. Such visibility has resulted in the creation of multi-stakeholder projects such as the Global Network Initiative (GNI), which place human rights at the centre. An international human rights law approach can provide a unifying framework to the diverse challenges that arise for the tech sector across diverse national contexts and a “normative baseline against illegitimate state restrictions” or in the absence of any regulation at all²⁰. While the UNGPs were not released until 2011, the GNI framework was influenced by the work developed by John Ruggie that served as the foundation of the UNGPs²¹. UNGPs are flexible enough to reflect changes in the business landscape, especially in this dynamic and forward-looking industry, which has disrupted traditional ways of “doing business”. Using the “Protect, Respect and Remedy” framework developed by the UNGPs, the GNI Principles recognise in its Preamble that:

the duty of governments to respect, protect, promote and fulfil human rights is the foundation of this human rights framework. That duty includes ensuring that national laws, regulations and policies are consistent with international human rights laws and standards on freedom of expression and privacy. ICT companies have the responsibility to respect

¹⁴ RUGGIE, J. *Just Business: multinational corporations and human rights*. New York: Norton & Company, 2013.

¹⁵ RUGGIE, J. *Just Business: multinational corporations and human rights*. New York: Norton & Company, 2013.

¹⁶ Vigeo Eiris. has identified 272 Software & IT Services controversies over the last four years, making this one of the most controversial sectors in its research universe. VIGEO EIRIS. *Human rights in a globalised world: why do companies need to pay more attention?* 2018. Available at: https://www.vigeo-eiris.com/wp-content/uploads/2018/11/2018_Human-rights-study_VFok.pdf. Accessed on: 10 Nov. 2021.

¹⁷ FREEDOM HOUSE. *Freedom on the Net 2018*. 2018. Available at: https://freedomhouse.org/sites/default/files/FOTN_2018_Final.pdf. Accessed on: 15 Oct. 2021.

¹⁸ DUNSTAN, A. Our Human Rights Impact Assessment of Facebook in Myanmar. November 2018. *Business for Social Responsibility*.

Available at: <https://www.bsr.org/en/our-insights/blog-view/facebook-in-myanmar-human-rights-impact-assessment>. Accessed on: 10 Oct. 2021.

¹⁹ RUGGIE, J. *Facebook in the rest of the world*. 2018. Available at: https://media.business-humanrights.org/media/documents/files/documents/John_Ruggie_Facebook_15_Nov_2018.pdf. Accessed on: 5 Oct. 2021.

²⁰ JØRGENSEN, R. F. *A human rights-based approach to social media platforms*. 2021. Available at: <https://berkeleycenter.georgetown.edu/responses/a-human-rights-based-approach-to-social-media-platforms>. Accessed on: 10 Oct. 2021.

²¹ SAMWAY, M. The Global Network Initiative: how can companies in the information and communications technology industry respect human rights? In: BAUMANN-PAULY, D.; NOLAN, J. (ed.). *Business and human rights: from principles to practice*. London: Routledge, 2016. p. 136-140.

and promote the freedom of expression and privacy rights of their users²².

International human rights law is state-centric in nature in the sense that states – not individuals, not companies – are the primary duty bearers. Part of this obligation, as highlighted by the UNGPs, is a duty upon the state to ensure that private actors do not violate human rights, articulated as a general duty to protect human rights or referred to as the horizontal effect of human rights law. Whereas human rights law is focused on the vertical relation (state obligations to the individual), it recognises the horizontal effect that may arise between private parties²³. The horizontal effect implies a state duty to protect human rights in the realm of private parties, for example, via industry regulation²⁴. Within the tech sector, when social media platforms accumulate more private power and institutional capacities of many governments around the world, and when issues of scale and potential harm become acute, the horizontal effect of human rights highlight the state duty to protect. However, the ways in which the state can perform this duty is far from clear, especially when addressing the behaviour of corporations that may accumulate more information about their own citizens or influence in public affairs as never imagined before.

When it comes to corporate conduct, we can observe at least three types of governance systems that are interrelated: international public law, civil governance and corporate governance²⁵. The UNGPs have made good progress in unifying and aligning the discourses and rationales of these three governance sub-systems. To do so, the process led by John Ruggie started with the identification of a baseline of key issues in the field of business and human rights and tested these issues through

public consultations²⁶. During the six years before the adoption of the UNGPs, the office of the Special Representative for Business and Human Rights conducted forty-seven formal consultations around the world, made numerous visits to key businesses, and spoke with governments, Human Rights Council representatives, and other stakeholders such as local and international non-governmental organisations and academics²⁷. As Cesar Rodriguez-Garavito argues, understanding who participated in the UNGPs' creation, and how they participated are not simply empirical questions but rather sit at the heart of the theory of governance that inspired the UNGPs²⁸. The participation of multinational corporations and international organizations provides support for developing of a normative framework that is consistent across borders and flexible enough to acknowledge different institutional capacities. As opposed to a state-centric approach, the process of debating, creating and drafting these principles developed a gradual and polycentric approach²⁹. Within this approach, for example, it is possible to grasp how the standard of “due diligence” was appropriated by the human rights community and developed a new meaning beyond the corporate jargon.

The field of business and human rights is a clear example of what experimental governance identifies as global legal pluralism. Not just because it aims at regulating different business sectors of different scales in a globalised world, but also because it compounds a variety of human rights related issues that entails all categories of rights, from social to civil and political rights. Moreover, it is developed through national, international and transnational law, a mix of hard and soft law, most of which has been developed without a coordina-

²² GLOBAL NETWORK INITIATIVES. *GNI Principles on freedom of expression and privacy*. 2017. Available at: <https://globalnetworkinitiative.org/gni-principles/>. Accessed 10 Nov. 2021.

²³ VAN DIJK, P. *et al. Theory and practice of the European Convention on Human Rights*. Antwerp: Oxford, 2006.

²⁴ In relation to human rights exercised online, this could refer to state intervention through content regulation, surveillance or law enforcement in order to prevent or punish the infringement of human rights. See for example DEIBERT, R. *et al.* (eds.). *Access Controlled: the shaping of power, rights, and rule in cyberspace*. London: The MIT Press, 2010.; DEIBERT, R. *et al.* (eds.). *Access denied the practice and policy of global internet filtering*. London: The MIT Press, 2008.

²⁵ RODRÍGUEZ-GARAVITO, C. *Business and human rights: beyond the end of the beginning*. Cambridge: Cambridge University Press, 2017.

²⁶ John Ruggie identifies three main issues: (a) ‘prevailing patterns of corporate-related human rights abuse’ (b) ‘existing legal standards and their application to states and business enterprises’ and (c) mapping out ‘the

attributes and rapid expansion of voluntary corporate social responsibility initiatives, pointing out their strengths as well as their shortcomings. RUGGIE, J. *Just Business: multinational corporations and human rights*. New York: Norton & Company, 2013.

²⁷ RUGGIE, J. *Just Business: multinational corporations and human rights*. New York: Norton & Company, 2013.

²⁸ RODRÍGUEZ-GARAVITO, C. *Business and human rights: beyond the end of the beginning*. Cambridge: Cambridge University Press, 2017.

²⁹ DE BÚRCA, G.; KEOHANE, R. O.; SABEL, C. New modes of pluralist global governance. *International Law and Politics*, New York, v. 45, p. 723-739, 2013.

ted node or overarching framework³⁰. In this scenario, the UNGPs attempt to merge the state's general duty to protect human rights and the corporate responsibility to respect human rights into a single normative framework, sometimes dubbed as a principled and pragmatic approach towards finding solutions to human rights problems. The UNGPs elaborate the distinction between the state's duty to protect human rights and the corporate responsibility to respect human rights through three pillars, often called the "Protect, Respect, and Remedy" framework. The first pillar (Protect) focuses on the role of the state in protecting individuals' human rights against abuses committed by non-state actors; the second pillar (Respect) addresses the corporate responsibility to respect human rights; and the third pillar (Remedy) sets out the roles of state and non-state actors in securing access to remedy. As argued by experimental governance proponents, stakeholders' - mainly civil society - participation in these processes is an essential source of pressure for compliance³¹.

As we will address in what follows, current developments in the tech sector illustrate how the experimental and polycentric governance approach of UNGPs has unfolded³². The next sections illustrate the ability of the UNGPs to generate unity in a field that has been historically fragmented by different regulatory systems and approaches. As opposed to a classical state-centred approach, the UNGPs have been recognised and implemented by companies and states through their own awareness as well as by the pressure of other private

actors and civil society. The lack of focus on the tech sector in the initial development of the UNGPs does not mean necessarily that the industry was outside of these discussions. Quite the contrary: the creation of the GNI and of international organizations' initiatives such as the B-Tech Project show that there is an interest in the issue. However, there is a lack of literature analysing to what extent these standards are operationalised in practice by both governments and the tech sector, especially concerning one of the main challenges that arise for this sector: content regulation/moderation.

3 Content regulation/moderation practices and the right to freedom of expression

Whether in the spread of disinformation, or decisions about what content we are able (or most likely) to post and see, the activities of governments and platforms concerning content regulation/moderation are coming under scrutiny as never before. While content regulation entails the role of governments in governing online speech, content moderation refers to the actions or decisions adopted by private actors that allow for the publication of information or opinions in the online world. Although content regulation/moderation can be exercised before a publication is made or afterwards, what is important is that these practices entail the "screening, evaluation categorisation, approval or removal/hiding of online content"³³.

The role of platforms in relation to content has changed in recent years. Traditionally, it was understood that there was a distinction between those platforms which merely hosted content and publishers which make editorial decisions. Such a view is reflected in a number of legal regimes that exclude the liability of those platforms that simply host content³⁴. Yet, it could be argued that online platforms are no longer entirely

³⁰ See for example, the previous instruments on which UNGPs rely on: ORGANIZATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT. *OECD Guidelines for Multinational Enterprises*. 2011 Edition. Available at: <http://mneguidelines.oecd.org/text/>. Accessed on: 5 Nov. 2021; INTERNATIONAL LABOUR ORGANIZATION. *ILO Declaration on Fundamental Principles and Rights at Work*. International Labour Organization, 1998. Available at: www.ilo.org/declaration/lang-en/index.htm. Accessed on: 5 Nov. 2021. See also NOLAN, Justine. Hardening soft law: are the emerging corporate social disclosure laws capable of generating substantive compliance with human rights?. *Revista de Direito Internacional*, Brasília, v. 15, n. 2, p. 64-83, 2018.

³¹ MELISH, T.; MEIDINGER, E. Protect, respect, remedy and participate: "New Governance" lessons for the Ruggie Framework. In: MARES, Radu (ed.). *The UN guiding principles on business and human rights: foundations and implementations*. Leiden: Brill-Nijhoff, 2012. See also CARDIA, Ana Cláudia Ruy. Reparação de vítimas à luz de um tratado sobre empresas e direitos humanos. *Revista de Direito Internacional*, Brasília, v. 15, n. 2, p. 2-11, 2018.

³² RODRÍGUEZ-GARAVITO, C. *Business and human rights: beyond the end of the beginning*. Cambridge: Cambridge University Press, 2017.

³³ FLEW, T.; MARTIN, F.; SUZOR, N. Internet regulation as media policy: rethinking the question of digital communication platform governance. *Journal of Digital Media & Policy*, [S.l.], v. 10, n. 1, p. 33-50, 2019.

³⁴ Art. 14. European Union's Directive on electronic commerce. Report n. 2000/31/EC. Establishes that service providers should not be held liable for content hosted unless (a) they have "actual knowledge" of its illegal nature or (b) upon obtaining such actual knowledge, they fail to act expeditiously to remove or to disable access to the content.

neutral in hosting and making available content online. In a world where participating in the online world has become almost indispensable, online platforms have been transformed into key gatekeepers of the content we can produce, share, and see and make decisions through their own logical architecture and rules³⁵. Companies might not generate original content but determine the manner and order of the available content and use algorithms to decide what content users see, which is a form of curation. Indeed, for Tarleton Gillespie, “moderation is, in many ways, the commodity that platforms offer”, especially in a digital landscape where the abundance of available information makes moderation necessary³⁶. Without governing content, it would be impossible to navigate in the seemingly endless world of online information.

In a world in which digital public spheres are controlled by a few gatekeepers, decisions regarding design features of the platforms have shaped what content is possible, community standards defined what is permissible and non-permissible online speech, and algorithms have generally influenced what kind of content is visible³⁷. Although social media platforms are built upon the premise of an unlimited expansion of speech, of making the marketplace of ideas bigger, corporate decisions may not necessarily coincide with the protection of the right to freedom of expression. As put by De Gregorio,

the interest of platforms is not just focused on facilitating the spread of opinions and ideas across the globe but establishing a digital environment where users feel free to share information and data that can feed commercial networks and channels and, especially, attract profits coming from advertising³⁸.

These decisions, frequently motivated by a logic of accumulation and data extraction, pose a serious risk to freedom of expression as exercised in the online world, which highlights its double dimension, both private

and public. Indeed, content moderation and regulation practices entail risks to both the privacy dimension that is necessary for freedom of expression and to a more public dimension, which entails considering freedom of expression as necessary to “foster a democratic culture”³⁹.

However, apart from the role of platforms, the conditions governing the relationship between online services and their users are defined also by state regulation. National governments are increasingly creating legislative and policy proposals which have led to greater regulation of online content and the imposition of increased liability on platforms⁴⁰. Upon this legal framework, content moderation is usually governed by Terms of Service or “law of the platform”, which is generally applicable in different jurisdictions⁴¹. While there are some examples of multi-stakeholder creation of these Terms of Service, in general, they are unilaterally defined - with the support of expert legal advice focused in preventing and mitigating any case of responsibility- and implemented by service providers. Apart from these explicit forms of public-private governance, “the techniques and organizational structures of content moderation are not neutral but affect how content is reviewed and which values are prioritized”⁴². Therefore, both content regulation and content moderation practices, either explicit or implicit, can limit access to information and condition the participation of users, ultimately limiting the right to freedom of expression and information.

While certain forms of content can be justifiably restricted (such as child sexual abuse imagery or incitement

³⁵ ZITTRAIN, J. A history of online Gatekeeping. *Harvard Journal of Law and Technology*, [S.l.], v. 19, n. 2, p. 253-298, 2006.

³⁶ GILLESPIE, T. Platforms are not Intermediaries. *Georgetown Law Technology Review*, [S.l.], v. 2, n. 2, p. 198-216, 2018.

³⁷ SANDER, B. Freedom of expression in the age of online platforms: the promise and pitfalls of a human-rights based approach to content moderation. *Fordham International Law Journal*, [S.l.], v. 43, n. 4, p. 939-1006, 2020.

³⁸ DE GREGORIO, G. Democratising online content moderation a constitutional framework. *Computer Law & Security Review*, [S.l.], v. 2, p. 105374, 2020.

³⁹ BALKIN, J. Free speech in the algorithmic society: Big Data, Governance, and new school speech regulation. *Davis Law Review*, [S.l.], v. 51, p. 1149-1210, 2020.

⁴⁰ FREEDOM HOUSE. *Freedom on the Net 2018*. 2018. Available at: https://freedomhouse.org/sites/default/files/FOIN_2018-Final.pdf. Accessed on: 15 Oct. 2021. There has also been some action regarding national courts subjecting platforms’ decisions such as social media exclusion. CELESTE, E. Social media bans: what platforms can learn from National Courts? 2021. *Blog British Association of Comparative Law*. Available at: <https://british-association-comparative-law.org/2021/05/21/social-media-bans-what-platforms-can-learn-from-national-courts-by-edoardo-celeste/>. Accessed on: 18 Oct. 2022.

⁴¹ DE FILIPPI, P.; BELLI, L. The Law of the Cloud v the Law of the Land: challenges and opportunities for innovation. *European Journal of Law and Technology*, [S.l.], v. 3, n. 2, p. 23, 2012.

⁴² SANDER, B. Freedom of expression in the age of online platforms: the promise and pitfalls of a human-rights based approach to content moderation. *Fordham International Law Journal*, [S.l.], v. 43, n. 4, p. 939-1006, 2020.

to violence) as in pursuance of legitimate aims (such as the prevention of crime or the protection of the rights of others), human rights organisations are increasingly concerned about risks of content being removed which is in fact protected by the right to freedom of expression⁴³. According to international human rights law, the right to freedom of expression “is applicable not only to “information” or “ideas” that are favourably received or regarded as inoffensive or as a matter of indifference, but also to those that offend, shock or disturb the State or any sector of the population”⁴⁴. Moreover, freedom of expression entails the broad protection of any form of speech and across national boundaries, and thus any restrictions to online content must be legal, proportionate, necessary, and legitimate⁴⁵. International human rights law provides an overall framework for both states and private corporations when addressing the complex issues of balancing the protection of freedom of expression and other fundamental rights. States have a general duty to protect human rights and, therefore, should ensure that any legislation which is of specific application to online platforms does not restrict freedom of expression explicitly or in its effects. Despite its contested legal status, there is a kind of consensus in international human rights law, according to which business enterprises should respect human rights, especially when they operate across borders⁴⁶. This means that platforms – in order to ensure a consistent degree of protection of human rights – have a responsibility not to restrict freedom of expression exercised via their technologies in a way which is inconsistent with international human rights law and standards.

In what follows, we argue that UNGPs provide the groundings for a human-rights based approach to one of the most hotly debated issues for governments, the

tech industry and probably the biggest challenge for social media platforms. In particular, we will base our analysis and examples on the particular harms that content regulation/moderation practices entail for the right to freedom of expression and the possible solutions that UNGPs survey for addressing the complex issues that arise when attempting to govern speech in the online world. In the last sub-section, we will provide the reader with a brief reflection about the role of individuals and civil society in protecting human rights in the context of content regulation/moderation practices.

4 UNGPs and content regulation/moderation

1. The State Duty to Protect Human Rights and Content Regulation

The “foundational” principle under pillar one of the UNGPs says that states must protect against human rights abuses by third parties (including business enterprises) within their territory and/or jurisdiction by taking appropriate steps to prevent, investigate, punish and redress such abuse through effective legislation, policies, regulation and adjudication. Additionally, the duty to protect human rights entails assessing the role of the state as an economic actor, beyond legal and policy measures, such as in public procurement activities or in the role of the State as an end-user of ICT⁴⁷. The state obligation to protect human rights entails both a positive and negative element. It requires the state to refrain from certain conduct, but also to take positive steps to ensure the enjoyment of the right in question. Freedom of expression, for example, requires that the state refrain from engaging in censorship, but also that it enables freedom of expression through diverse forms of regulation⁴⁸. A key question is how to apply these

⁴³ See summary of Special Rapporteur’s Consultations at David Kaye’s report on the issue. UNITED NATIONS. Human Rights Council. *Report n. A/HRC/38/35*. Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, David Kaye. Human Rights Council, 2018.

⁴⁴ EUROPEAN COURT OF HUMAN RIGHTS. *Application No. 5493/7 Handyside v the United Kingdom*. 1976.

⁴⁵ HUMAN RIGHTS COMMITTEE. *Report n. CCPR/C/GC/34*. General Comment 34. United Nations, 2011, paragraph 12. See also: UN. Report n. A/HRC/38/35. Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression - regulation of user-generated online content. 2018, paragraph 7.

⁴⁶ ARNOLD, D. G., Transnational corporations and the duty to respect basic human rights. *Business Ethics Quarterly*, [S.L.], v. 20, n. 3, p. 371-399, 2010.

⁴⁷ B-TECH FOUNDATION PAPER. *Bridging governance gaps in the age of technology*: key characteristics of the State duty to protect. 2020. Available at: <https://www.ohchr.org/Documents/Issues/Business/B-Tech/b-tech-foundational-paper-state-duty-to-protect.pdf>. Accessed on: 10 Nov. 2021.

⁴⁸ The scope and potential limitations of the right to freedom of expression are developed in UN Human Rights Committee (HRC), note 35. At the regional level, in the ruling *Editorial Board of Pravoye Delo and Shtetel v Ukraine*. Application n. 33014/05, 2011. The European Court of Human Rights for the first time acknowledged that Article 10 imposes on states a positive obligation to create an appropriate regulatory framework to ensure effective protection of

regulatory frameworks to tech companies that are often global and not based in the jurisdiction of the government that is trying to prevent human rights abuses. In this regard, the UNGPs provide a normative universal baseline that “increases the chances that technology-oriented State policies and standards can become consistent across geographies, even while the exact form of these will be constructed to meet local realities”⁴⁹.

While the UNGPs say that states are not generally required to regulate the extraterritorial activities of businesses domiciled in their territory and/or jurisdiction, they also recognise that states are not generally prohibited from doing so, providing that there is a recognised jurisdictional basis (UNGP Principle 2)⁵⁰. The UNGPs recognise that there may be strong policy reasons for states to be clear about the expectations and behaviour of businesses abroad. However, considering the “dynamic dimension” of these principles, which “push the development of new norms and practices that go beyond the initial content of the UNGPs and improve companies” compliance with human rights standards⁵¹, there has been recent developments on the extraterritorial obligations of states regarding companies or businesses that may be deemed under their jurisdiction, authority or control⁵².

The main tool that governments have to protect human rights when it comes to impacts in the tech sector is legislation and administrative regulation. When it comes to the tech sector, most of the normative frameworks have been reactions to harms developed by

either states, companies and even users themselves. In the case of content regulation, legislation from government has evolved, as have companies’ concerns over such legislation, facing two main challenges. One the one hand, the pace of technological developments poses different kind of challenges for governments’ bureaucracies that lack the expertise and information that ICT companies have. On the other hand, most of the regulatory attempts are focused on the behaviour of the tech giants, which does not represent the diverse reality of the sector, and are mainly directed at redressing harms rather than promoting better environments⁵³.

At the beginning of the 2000s, the concerns of tech companies - mostly headquartered in the United States - were based on conflicts of law, especially when it comes to conflicts with restrictive laws such as those passed in China. By the mid 2000s China created what has been known as the “Golden Shield”, a technology used to create legal pressure to censor content online⁵⁴. This initiative led to Microsoft’s decision to remove the blog of the Chinese political journalist Michael Anti and Yahoo! to disclose the account data of Shi Tao, which ultimately resulted in his arrest⁵⁵. At the height of these scandals, Google decided to enter the Chinese market announcing that they would limit the display of certain search results based on Chinese censorship requirements, a decision that led to huge criticism from media and civil society⁵⁶. Nowadays, in a number of cases, despite the state’s intention to prevent potential negative consequences of technologies through legislation, they have ended up putting other human rights at risk. For example, the right to erasure, or the “right to be forgotten” that is now formalised in the European Union’s General Data Protection Regulation, has the potential to address legitimate privacy concerns but

journalists’ freedom of expression on the Internet.

⁴⁹ B-TECH FOUNDATION PAPER. *Bridging governance gaps in the age of technology*: key characteristics of the State duty to protect. 2020. Available at: <https://www.ohchr.org/Documents/Issues/Business/B-Tech/b-tech-foundational-paper-state-duty-to-protect.pdf>. Accessed on: 10 Nov. 2021.

⁵⁰ CERQUEIRA, D.; MONTGOMERY, A. Extraterritorial obligations: a missing component of the UN Guiding Principles that should be addressed in a binding treaty on business and human rights. 2018. *Due Process Foundation Blog*. Available at: <https://dplf-blog.com/2018/02/08/extraterritorial-obligations-a-missing-component-of-the-un-guiding-principles-that-should-be-addressed-in-a-binding-treaty-on-business-and-human-rights/>. Accessed on: 20 Mar. 2022.

⁵¹ RODRÍGUEZ-GARAVITO, C. *Business and human rights: beyond the end of the beginning*. Cambridge: Cambridge University Press, 2017.

⁵² COMMITTEE ON ECONOMIC, SOCIAL AND CULTURAL RIGHTS. *Report n. E/C.12/GC/24*. General comment No. 24. On State obligations under the International Covenant on Economic, Social and Cultural Rights in the context of business activities. 2017, paragraphs 25-28.

⁵³ UK PARLIAMENT. *Joint Committee on the Draft Online Safety Bill*. London, 2021. Available at: <https://publications.parliament.uk/pa/jt5802/jtselect/jtonlinesafety/129/12902.htm>. Accessed on: 10 Mar. 2022.

⁵⁴ QIANG, X. How China internet police control speech on the Internet. *Radio Free Asia*, [S.l.], v. 24, 2008. Available at: https://www.rfa.org/english/commentaries/china_internet-11242008134108.html. Accessed on: 29 Jan. 2019.

⁵⁵ SCHOFIELD, J. Microsoft takes down Chinese blogger. *The Guardian*, [S.l.], 2 January 2006. Available at: <https://www.theguardian.com/technology/blog/2006/jan/04/microsofttakes>. Accessed on: 24 Jan. 2019.

⁵⁶ THOMSOM, C. Google’s China problem (and China’s Google problem). *New York Times*, New York, 23 April 2006. Available at: <https://www.nytimes.com/2006/04/23/magazine/23google.html>. Accessed on: 24 Jan. 2019.

also has potential impacts upon the rights to freedom of expression and access to information⁵⁷.

As said before, operations of internet intermediaries are heavily influenced by the legal and policy environments of states⁵⁸. From an international human rights perspective, the state has an obligation to ensure enabling environments for freedom of expression and to protect its exercise⁵⁹. It also has a duty to ensure that private entities do not impair with the freedoms of opinion and expression⁶⁰. In specific cases, when the conditions of legality and necessity are met, the state can require intermediaries and online platforms to restrict content such as child sexual abuse, copyright infringement or incitement to violence⁶¹. In reality, as put by Peggy Hicks, Director of Thematic Engagement for UN Human Rights, “nearly every country that has adopted laws relating to online content has jeopardised human rights in doing so”⁶². Faced with periodical scandals, Governments have rushed to enact regulations without broad and public consultation with diverse stakeholders and experts, providing “simple solutions to complex problems”⁶³. Within this scenario, the role of state actors regarding content regulation has entailed different

forms of previous censorship, currently prohibited under human rights standards, or led to the unjustified restriction of freedom of expression or other fundamental rights.

In addition to ensuring that the broader legal framework, where it touches upon freedom of expression as it is exercised online, is consistent with international human rights law and standards, states should ensure that any legislation which is of specific application to platforms does not restrict freedom of expression explicitly or in its effects. Moreover, as put by article 13 of the American Convention of Human Rights, which recognizes the right to freedom of thought and expression, states must not restrict this right through “indirect methods or means”. Several commentators have argued that this international treaty provides textual grounding for the express prohibition of indirect restrictions such as private law regulations⁶⁴. Recently, there has been heightened scrutiny upon regulation that, with the purpose of protecting freedom of expression in the online world, may end up contributing to more concentrated digital markets, which in itself pose risks to this right⁶⁵. In this scenario, some regulatory frameworks have opted to distinguish online platforms on the grounds of their size, in order to prevent any “unintended impacts they could have on the pluralism of content and providers of consumers services that may be available”⁶⁶.

According to the UNGPs, the general duty to protect human rights entails adopting internal regulations that take into account the overarching human rights standards that may be applicable and not just the single issue or right at stake, which provides an opportunity for governments to justify complex issues where balancing and proportionality between different rights is required. Any regulatory proposal must incorporate human rights standards from the beginning of the policy debates and institutional design choices, anticipating

⁵⁷ FAZLIOGLU, M. Forget me not: the clash of the right to be forgotten and freedom of expression on the Internet. *International Data Privacy Law*, [S.l.], v. 3, n. 3, p. 149-157, 2013.

⁵⁸ MACKINNON, R. *et al. Fostering freedom online: the role of internet intermediaries*. Paris: UNESCO Publishing, 2015.

⁵⁹ UNITED NATIONS. Human Rights Council. *Report n. A/HRC/38/35*. Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, David Kaye. Human Rights Council, 2018.

⁶⁰ HUMAN RIGHTS COMITEE. *Report n. CCPR/C/GC/34*. General Comment 34. United Nations, 2011, paragraph 7.

⁶¹ Freedom of expression is not an absolute right, it may be restricted in accordance with the provisions of article 19, paragraph 3 and article 20 of the International Covenant on Civil and Political Rights (ICCPR). Specifically, restrictions should be provided by law and should set with ‘sufficient precision’ the difference between lawful and unlawful expression. Moreover, restrictions must be necessary and proportionate, in accordance with one of the enumerated ‘legitimate purposes’ in article 19 paragraph 3. UNITED NATIONS. Human Rights Council. *Report n. A/HRC/38/35*. Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, David Kaye. Human Rights Council, 2018. paragraph 7.

⁶² UNITED NATIONS. *Moderating online content: fighting harm or silencing dissent?* 2021. Available at: <https://www.ohchr.org/EN/NewsEvents/Pages/Online-content-regulation.aspx>. Accessed on: 12 Nov. 2021.

⁶³ UNITED NATIONS. *Moderating online content: fighting harm or silencing dissent?* 2021. Available at: <https://www.ohchr.org/EN/NewsEvents/Pages/Online-content-regulation.aspx>. Accessed on: 12 Nov. 2021.

⁶⁴ CENTER FOR STUDIES ON FREEDOM OF EXPRESSION AND ACCESS TO INFORMATION [CELE]. *Submission to the UN SR on the Protection and Promotion of Freedom of Opinion and Expression*. Palermo, 2017.

⁶⁵ RAMÍREZ, I. *Online content regulation and competition policy*. 2020. Available at: <https://orgs.law.harvard.edu/antitrust/2020/12/03/online-content-regulation-and-competition-policy/>. Accessed on: 10 Nov. 2021.

⁶⁶ GLOBAL NETWORKING INITIATIVE. *Content regulation and human rights: analysis and recommendations*. 2020. Available at: <https://globalnetworkinitiative.org/wp-content/uploads/2020/10/GNI-Content-Regulation-HR-Policy-Brief.pdf>. Accessed on: 10 Nov. 2021.

and preventing any potential negative impact (UNGP Principle 2). Additionally, regulatory efforts should be careful enough to apply a single normative framework based on human rights while acknowledging the different sizes and scales of ICT companies. In a digital market characterized by excessive concentration, UNGPs may provide a comprehensive and practical approach to acknowledge the different risks and institutional capacities that are involved in each case, which range from ICT start-ups to multinational companies with different branches. By default, UNGPs may also provide a framework for addressing protection gaps that result from digital technologies that are constantly evolving and changing⁶⁷. Moreover, and following the idea of a “smart mix of measures” (UNGP Principle 3), states may refer to voluntary initiatives that could gain support in cases where the expected behaviour may be implemented without undermining the legitimacy of the overall normative framework.

For the B-Tech Project, and following UNGP Principle 8, “states must ensure that they have the necessary policy coherence – as well as capacity and ability- to effectively protect people against harms involving technology companies”. In this regard, in order to fulfil their obligations under the UNGP Protect pillar, states have begun to unify their regulatory attempts through their national human rights institutions and National Action Plans (NAPs) on business and human rights, which are generally debated, designed and implemented through open consultation processes with experts and diverse stakeholders⁶⁸. In doing so, we strongly believe that NAPs should include digital issues. However, not much has been done so far⁶⁹. Most NAPs’ commitments do not relate to specific forms of protection, responsibility and remedies, but just lay the ground

for potential developments on specific issues related to ICTs. For example, the Swedish government highlights in its NAP that internet freedom and privacy are among the great global issues of the future. It states that it is fundamental for Sweden that the human rights that apply offline also apply online, adding that as a result of a Swedish initiative, the OECD Guidelines for Multinational Enterprises now call on companies to support human rights on the internet⁷⁰. Yet apart from these developments, the Swedish government does not highlight any action point or future development in the field. Similarly to Sweden, the Irish NAP highlights past actions such as providing a fourfold increase in the funding for the work of the Data Protection Commission, and the UK NAP highlights that the government has strengthened international rules relating to digital surveillance, including leading work in the Wassenaar Arrangement to adopt new controls on specific technologies of concern⁷¹. Other countries have developed action points. In particular, the Polish government has committed to draft a regulation to counteract restrictions on the freedom of speech⁷². Also, the Finnish government proposed to create a roundtable discussion on how to ensure the protection of privacy with the authorities, ICT companies and civil society⁷³. In terms of process, the government of the Netherlands undertook a Sector Risk Analysis in 2014 which identified the electronics sector as among those with the greatest risk of adverse human rights impacts⁷⁴. The Dutch government has committed to negotiating voluntary corporate social responsibility agreements that focus on transparency, dialogue with stakeholders, and monitoring of agreements with those sectors. As we have seen here, although NAPs have constituted interesting efforts at

⁶⁷ B-TECH FOUNDATION PAPER. *Bridging governance gaps in the age of technology*: key characteristics of the State duty to protect. 2020. Available at: <https://www.ohchr.org/Documents/Issues/Business/B-Tech/b-tech-foundational-paper-state-duty-to-protect.pdf>. Accessed on: 10 Nov. 2021.

⁶⁸ For the role National Human Rights Institutions in achieving policy coherence in the tech sector, see UTLU, D. Public policy and digital technologies: the role of national human rights institutions in achieving policy coherence. *B-Tech*, [S.l.], May 2021. Available at: <https://www.ohchr.org/Documents/Issues/Business/B-Tech/b-tech-blog-policy-coherence-nhris-tech.pdf>. Accessed on: 10 Nov. 2021.

⁶⁹ DE FELICE, D.; GRAF, A. The potential of National Action Plans to implement human rights norms: an early assessment with respect to the UN Guiding Principles on Business and human rights. *Journal of Human Rights Practice*, [S.l.], v. 7, n. 1, p. 40-71, 2015.

⁷⁰ SWEDEN. Government Offices of Sweden. *Action plan for business and human rights*. Here and in the following footnotes, references to NAPs can be found in the OHCHR website. 2015. <https://www.ohchr.org/en/issues/business/pages/nationalactionplans.aspx>. Accessed on: Nov. 2021.

⁷¹ IRELAND. Government of Ireland. *National Plan on Business and human rights 2017-2020*. 2017. UNITED KINGDOM. Government of United Kingdom. *Good business: implementing the UN guiding principles on business and human rights*. 2016.

⁷² POLAND. Government of Poland. *Polish National Action Plan for the Implementation of the United Nations Guiding Principles on Business and Human Rights 2017-2020*. 2017.

⁷³ FINLAND. Government of Finland. *National Action Plan for the implementation of the UN Guiding Principles on Business and Human Rights*. 2014.

⁷⁴ NETHERLANDS. Government of Netherlands. *National Action Plan on Business and Human Rights*. 2013.

materializing the state's duty to protect human rights, they have not been precise enough to detail how the different stakeholders should handle the complex issues of freedom of expression and other fundamental rights in the online world.

Additionally, of particular concern is legislation which attaches liability to platforms for content that is available on them. Broadly speaking, intermediary liability laws are the “rules to protect intermediaries from liability for the content third parties publish on their platforms”⁷⁵. Yet, there is no universally accepted definition of an intermediary; rather, different entities and reports have defined the term differently, from definitions that expressly exclude content producers to broad definitions that include both content producers and hosts⁷⁶. For the reasons set out in the previous section, in this paper, we take a broad concept of intermediaries referring to platforms that both host content and make editorial decisions.

There are at least three categories of intermediary liability, as seen in Table 1: broad protections, conditional protections, and strict liability⁷⁷. Broad immunity regimes are those that exempt intermediaries from liability for a wide range of third-party content. Conditional protections are those liability mechanisms in which intermediaries are exempt from liability for third-party content if certain conditions are met; intermediaries will usually be liable when they have “knowledge” of infringing content. Finally, strict liability regimes are those where platforms are held liable for content, even if they are not aware of it; “they are most likely to result in overly broad restrictions of freedom of expression, as they require the platform proactively to monitor and remove content, even without notification”⁷⁸. Beyond

this different categories of intermediate liability, there is some consensus on those working in the field of content regulation and human rights that “lawmakers should resist the temptation to shift all legal liability from those generating illegal content to intermediaries”, placing in the state the primary responsibility for complying with human rights standards in this regard⁷⁹.

Table 1: Intermediary Liability Regimes

Apart from national legislation/regulation, states can limit freedom of expression through a series of other activities or actions. First, some states demand extraterritorial removal of links, websites and other content alleged to violate local law⁸⁰. One example is the ruling against Google, made by the Commission Nationale de L'Informatique et des Libertés in March 2016, which required the global takedown of links to search information banned in France under the European Union “right to be forgotten”⁸¹. This sets a precedent of state authorities banning search results not just inside their own jurisdictions, but also asserting that jurisdiction across the globe. On the other hand, state authorities are increasingly seeking content removals outside of legal processes or even through Terms of Service requests⁸². In some cases, government have created specialised offices to request content removal⁸³. These situations open a series of questions about the potential application of UNGPs - which openly does not regulate extraterritoriality - to address the increasingly sophisticated forms by which governments' actions may end up generating adverse human rights impacts, specifically in the issue of freedom of expression. In

⁷⁵ UNITED NATIONS. Human Rights Council. *Report n. A/HRC/38/35*. Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, David Kaye. Human Rights Council, 2018. paragraph 14.

⁷⁶ MACKINNON, R. *et al. Fostering freedom online: the role of internet intermediaries*. Paris: UNESCO Publishing, 2015. GRASSER, U.; SCHULZ, W. *Governance of online intermediaries: observations from a series of national case studies*. Cambridge: The Berkman Center for Internet & Society Research Publication, [S.l.], v. 18, p. 283, 2015.

⁷⁷ Art. 19, *Internet intermediaries: Dilemma of Liability Q and A*. 2013. Available at: <https://www.article19.org/resources/internet-intermediaries-dilemma-liability-q/>. Accessed on: 10 November 2021.

KUCZERAWY, A. Intermediary liability & freedom of expression: recent developments in the EU notice & action initiative. *Computer Law and Security Review*, [S.l.], v. 31, n. 1, p. 46-56, 2015.

⁷⁸ BRADLEY, C.; WINGFIELD, R. *A rights-respecting model of online content regulation by platforms*. London: Global Partners Digital, 2018.

⁷⁹ GLOBAL NETWORKING INITIATIVE. *Content regulation and human rights: analysis and recommendations*. 2020. Available at: <https://globalnetworkinitiative.org/wp-content/uploads/2020/10/GNI-Content-Regulation-HR-Policy-Brief.pdf>. Accessed on: 10 Nov. 2021.

⁸⁰ PEN AMERICA. *Forbidden fees: government controls on social media in China*. 2018. Available at: https://pen.org/wp-content/uploads/2018/03/PENAmerica_Forbidden-Feeds-3.13-3.pdf. Accessed on: 18 Nov. 2021.

⁸¹ COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTÉS [CNIL]. *Case n. C-507/17*. 2019.

⁸² UNITED NATIONS. Human Rights Council. *Report n. A/HRC/38/35*. Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, David Kaye. Human Rights Council, 2018. paragraph 19.

⁸³ UNITED NATIONS. Human Rights Council. *Report n. A/HRC/38/35*. Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, David Kaye. Human Rights Council, 2018. See the examples of the European Union Internet Referral Unit or specialized offices in Australia and South-East Asia.

any case, there are good reasons, either under a broad interpretation of UNGPs or within the interpretation of the International Covenant on Economic, Social and Cultural Rights made by its treaty body, to elaborate further on extraterritorial obligations regarding potential infringement of private actors on the right to freedom of expression abroad⁸⁴.

2. *The Corporate Responsibility to Respect Human Rights and Content Moderation*

Pillar two of the UNGPs outlines the independent corporate responsibility to respect human rights. The responsibility to respect human rights requires business enterprises to avoid infringing on human rights and to address any adverse human rights impacts with which they are involved, even in cases where states are unable or unwilling to protect human rights. This principle applies to all businesses of all sizes in all situations. According to Principle 15, in order to “know and show” that they respect human rights, businesses should have in place: (a) a policy commitment to respect human rights, (b) an ongoing process of human rights due diligence, and (c) processes to enable the remediation of any adverse human rights impacts they cause or to which they contribute.

Human rights due diligence needs to cover different types of adverse impacts on all internationally recognised human rights. This includes impacts that the business causes, impacts to which the business contributes, and impacts that are directly linked to business operations, products or services. In the ICT sector, although “Human Rights Due Diligence will look different in different organizations and processes”, they should all have a special consideration for the features of the industry⁸⁵. Actual and potential adverse human rights impacts should be avoided and addressed through human rights due diligence which includes four key steps (a) assessing adverse human rights impacts; (b) integrating and acting upon the findings in the impact assessment, and to do so, the company should integrate these

findings into the company policies and create internal capacity opportunities; (c) tracking and monitoring responses: this should be an ongoing process and it is important to involve affected right-holders and other stakeholders; and (d) communicating and reporting: information about due diligence and impact assessments should be publicly communicated that are accessible, provide sufficient and relevant information, and not pose risks to rights holders or others.

Since the endorsement of the UNGPs by the Human Rights Council in 2011, there has been an increasing commitment to human rights in tech companies’ policy statements. Using the database contained in the Business and Human Rights Resource Centre, we analysed 42 companies in the tech sector⁸⁶. 52% of the companies have available human rights policies and another 10% (including companies such as Facebook and Google), refer to their participation in the Global Network Initiative, and the GNI’s assessment of their policies and practice on human rights, without making public their human rights policies. The influence and experimentalist approach taken by the UNGPs becomes more evident when companies such as Microsoft, Ericsson, Telefonica and Yahoo!, have stated that since the adoption of the UNGPs in 2011, they have either created or adapted their companies’ policies in line with the “Protect, Respect, Remedy” framework.

What is perhaps more important is to evaluate how these statements have been developed in further commitments to respect human rights.⁸⁷ Again, as with states, there is contrasting evidence. Some companies have taken preventative action; others that have only acted reactively. Companies such as AT&T, HP, Telefonica, and Vodafone have argued that in order to act a priori, they have developed comprehensive human rights impact assessments; and CISCO, Deutschland Telekom Intel and Verizon have created formal governance structures such as internal cross-sector human rights

⁸⁴ COMMITTEE ON ECONOMIC, SOCIAL AND CULTURAL RIGHTS. *Report n. E/C.12/GC/24*. General comment No. 24. On State obligations under the International Covenant on Economic, Social and Cultural Rights in the context of business activities. 2017, paragraphs 25-28.

⁸⁵ B-TECH FOUNDATION PAPER. *Key characteristics of business respect for human rights*. 2020. Available at: <https://www.ohchr.org/Documents/Issues/Business/B-Tech/key-characteristics-business-respect.pdf>. Accessed on: 12 Nov. 2021.

⁸⁶ The Resource Centre is an independent non-profit organization based in the United Kingdom and United States. According to their website, the researchers “draw global attention to the human rights impacts (positive and negative) of companies in their region, seek responses from companies when civil society raises concerns, and establish close contacts with grassroots NGOs, local businesspeople, and others.” Available at: <https://www.business-humanrights.org/en/>. Accessed on: 19 November 2021.

⁸⁷ SANTARELLI, Nicolás Carrillo. Declaraciones empresariales “voluntarias” sobre derechos humanos, y la necesidad de una regulación institucional (internacional e interna) externa. *Revista de Direito Internacional*, Brasília, v. 16, n. 3, p. 23-49, 2019.

working groups, human rights executive committees, and human rights contact points. Companies such as CISCO, Deutschland Telekom, Orange, and Qualcomm have said that the UNGPs have allowed them to “deal with the topic of human rights in the greater business context”, to “align internally and provided for a more systematic approach to human rights risks and opportunities” and to “further enhance their approach to incorporate human rights throughout their company”.

Despite these positive developments, there are still important gaps in the ICT sector when it comes to their responsibility to respect human rights, especially freedom of expression and privacy. As the latest RDR corporate accountability index states, “Companies are improving in principle, but failing in practice⁸⁸. Though we have seen important developments and commitments from different ICT companies, RDR highlights that most of these companies do not have the appropriate mechanisms to mitigate the risks posed, for example, by government censorship, malicious activities from other non-state actors and their own business models. But perhaps more important for the focus of this paper, “companies do not adequately inform the public about how content and information flows are policed and shaped through their platforms and services” arguing that usually the decisions undertaken by companies lack accountability and reparation mechanisms⁸⁹. Although several companies have created redress mechanisms, there is no transparency regarding the way in which these remedies are addressed and handled by the companies. The recent creation of Facebook’s Oversight Board may be starting point of an innovative private governance approach that is embedded in broader framework of rule of law values⁹⁰.

When moderating content, companies need to comply with national legislation as well as with international human rights law. As Facebook puts it: “When content is reported as violating local law, but doesn’t go against our Community Standards, we may limit access to that content in the country where the local violation is

alleged”⁹¹. Yet when complying with content regulation requests, companies must be sure that they are meeting international human rights standards, especially when making decisions in national contexts that may have legislation that is vague, subject to varying interpretations or inconsistent with human rights law⁹². As previously explained, the UNGPs provide some preventive tools to minimise those conflicts, such as transparency reports and due diligence duties. Another good example is provided by the GNI Principles which, taking into consideration the UNGPs contain a set of recommendations to respect freedom of expression and privacy, when they deal with governments requests⁹³.

Some companies have gone further and regulated the way they manage content on their platforms, through policies for content moderation or Terms of Service. These can be defined as any rules (regardless of phrasing or format) established by a platform which set out the criteria according to which content will be removed or restricted, or a user’s account deleted or suspended. The repertoire of private rules includes “Community Standards”, “Participation Guidelines”, “Rules”⁹⁴. Yet, the evidence suggests that the relationship between these internal policies and respect for human rights, particularly freedom of expression, is not very encouraging. Taking into consideration international and regional human rights standards, the Center for Technology and Society of Fundação Getulio Vargas Rio de Janeiro Law School analysed the Terms of Service of 50 major online platforms in order to assess how they dealt with human rights, including the right to freedom of expression, privacy and due process. In terms of freedom of expression, the study concluded that 46% of the platforms contained clauses that allowed them to monitor content, without specifying which kind of content, and only 8% explicitly stated either that they will not monitor content, or that, they will do so only to the

⁸⁸ RANKING DIGITAL RIGHTS. *Corporate Accountability Index 2020*, 2020. Available at: <https://rankingdigitalrights.org/index2020>. Accessed on: 10 Nov. 2021.

⁸⁹ RANKING DIGITAL RIGHTS. *Corporate accountability index 2018*. 2018. Available at: <https://rankingdigitalrights.org/index2018/report/executive-summary/>. Accessed on: 12 Nov. 2021.

⁹⁰ KLONICK, K., The Facebook Oversight Board: creating an independent institution to adjudicate online free expression. *The Yale Law Journal*, [s.l.], v. 129, p. 2418-2499, 2020.

⁹¹ PROGRESS on our commitment to transparency – META. 2021. Available at: <https://about.fb.com/news/2021/11/meta-transparency-report-h1-2021/>. Accessed on: 18 Nov. 2021.

⁹² UNITED NATIONS. Human Rights Council. *Report n. A/HRC/38/35*. Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, David Kaye. Human Rights Council, 2018. paragraph 23.

⁹³ Ranking Digital Rights, note 87.

⁹⁴ BRADLEY, C.; WINGFIELD, R. *A rights-respecting model of online content regulation by platforms*. London: Global Partners Digital, 2018.

extent necessary to eliminate materials that violate their policies⁹⁵.

RDR have arrived at similar conclusions. In 2018, the Ranking Digital Rights corporate accountability index reviewed 22 major internet companies and found that internet and mobile ecosystem companies lack transparency about what their rules are and actions they take to enforce them. Their results show that while internet and mobile ecosystem companies disclosed at least some information about what types of content or activities are prohibited by their Terms of Service (Facebook, Kakao and Microsoft leading this indicator), most disclosed nothing about the actions they took to enforce these rules (Twitter and Microsoft being the best ranked, yet still below the 50% on a scale from 1 to 1000). Yet, the same Index recognise that there is some progress in the field, concluding that in 2015 no company disclosed any data about their content moderation. In 2016 three companies did so and four companies disclosed their data about the volume or nature of content or accounts restricted for violating their rules in 2017. In 2018, we can observe, at least five companies as shown in Table 2.

Table 2: Rule Enforcement data Provided

When analysing different content moderation policies, we can identify good and bad practices related to both its content and process. In terms of good procedural practices, companies are increasingly creating internal teams with varied expertise that regularly update these policies. Oath, for example, created a business and human rights team that participates in this process⁹⁶ and Facebook has created an internal team including individuals with experience in a range of sectors, including child safety, hate speech, and terrorism, including human rights lawyers or criminal prosecutors⁹⁷. In terms of content, there are also some platforms that have tailored their Terms of Service specifically for the audiences that use them. That is the case, for example, of the

children platform Scratch, which developed Community Guidelines which are short, clear and child-friendly⁹⁸.

Despite the examples provided before, Terms of Service are usually dense and formulated in language that is hard to be understood by anyone who does not have legal training. That is one of the reasons why people hardly ever read these contracts⁹⁹. When they do, they find them difficult to understand¹⁰⁰. Moreover, regarding clarity, Terms of Service are usually vague. Platforms should ensure that they provide sufficient detail – whether through accompanying documents or in the Terms of Service themselves – to enable users to know, with a reasonable degree of certainty, whether particular content is or is not restricted. This is especially true with terms such as “hate”, “harassment” or “abuse” or even certain counter-terrorism policies. Vague policies in these field carry the risk of excessive limitations to freedom of expression, which has proven to mainly affect “minorities while reinforcing the status of dominant or powerful groups”¹⁰¹.

Apart from the private rules of governance of online speech, and influenced by the increasing costs of content moderation, several platforms have resorted to algorithmic systems to comply with national regulations – such as short timelines for content takedowns- and, allegedly, to international human rights standards¹⁰². In the case of the biggest competitors in the market of online platforms, issues of scale have triggered a shift towards what has been called as “commercial content

⁹⁵ VENTURINI J. et al. *Terms of service and human rights: an analysis of online platform contracts*. Rio de Janeiro: Revan, 2016. p. 152.

⁹⁶ Oath is now branded as Verizon Media. However, during its short lifespan (2017-2019), this special team. OATH. *Our approach: built on seven pillars*. 2019. Available at: <https://www.oath.com/our-story/business-and-human-rights/our-program/>. Accessed on: 15 Jan. 2019.

⁹⁷ ZUCKERBERG, M. *A Blueprint for Content Governance and Enforcement*. 2018. Available at: <https://www.facebook.com/notes/mark-zuckerberg/a-blueprint-for-content-governance-and-enforcement/10156443129621634/>. Accessed on: 15 Jan. 2019.

⁹⁸ LOMBANA-BERMÚDEZ, A. *Moderation and sense of community in a youth-oriented online platform: Scratch's governance strategy for addressing harmful Speech*. 2017. Available at: <https://medium.com/berkman-klein-center/moderation-and-sense-of-community-in-a-youth-oriented-online-platform-scratchs-governance-eeac6941e9c9>. Accessed on: 10 Nov. 2021.

⁹⁹ BYGRAVE, L. *Internet governance by contract*. Oxford: Oxford University Press, 2015.

¹⁰⁰ BAKOS, Y.; MAROTTA-WURGLER, F.; TROSSEN, D. R. *Does anyone read the fine print? consumer attention to standard form contracts*. Law & Economics Research, paper series. 2009. Ideas. Available at: <https://ideas.repec.org/p/net/wpaper/0904.html>. Accessed on: 10 November 2021.

¹⁰¹ UNITED NATIONS. Human Rights Council. *Report n. A/HRC/38/35*. Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, David Kaye. Human Rights Council, 2018. paragraph 27.

¹⁰² According to Expert Market Research, the industry of content moderation will reach a 12 billion dollars for 2027. CONSUMER NEWS AND BUSINESS CHANNEL. *Why content moderation costs billions and is so tricky for Facebook, Twitter, YouTube and others*. 2021. Available at: <https://www.cnb.com/2021/02/27/content-moderation-on-social-media.html>. Accessed on: 17 Nov. 2021.

moderation” or “algorithmic content moderation”, with the expectation that machine learning techniques could handle the delicate balance between freedom of expression and other important rights in the near future. If content moderation is part of the commodity that online platforms compete for, reducing the associated costs is an essential part of the evolving business strategies. However, the turn to AI-based content moderation, as pointed out by several CSOs, “has remained opaque, unaccountable and poorly understood”¹⁰³. For example, as the Facebook Papers revealed recently, most of the integrity systems that use AI systems within Facebook’s content moderation practices are not well trained in languages other than English, triggering several human rights risks¹⁰⁴. Within the normative framework of UNGPs, algorithmic content moderation must be integrated with accountability standards and transparency measures that could make clear when do online platform rely on algorithms for decisions that affect end-users, how are these decisions processed and adopted, and how to challenge those decisions before a human based system. For some critics, these human rights requirements may be under pressure when they entail touching upon “business models that involve the sale of human attention”¹⁰⁵.

Online platforms offer few guarantees in their policies on preserving the right to freedom of expression. There is a lack of clear and specific information in the Terms of Service on what content is allowed or not in the platform. There is also little commitment to offering users justification, notice and the right to be heard when content is removed by the platforms’ own initiative, after notification from third parties or through automated procedures. That is why in the concluding remarks we push for a model of content moderation for online platforms that fully reflects companies’ responsibilities under the UNGPs.

3. *The role of individual and civil society organisations*

¹⁰³ GORWA, R.; BINNS, R.; KATZENBACH, C. Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society*, [S.l.], v. 7, n. 1, p. 15, 2020. Available at: <https://doi.org/10.1177/2053951719897945>. Accessed on: 10 Nov. 2021.

¹⁰⁴ HAUGEN, F. *Formal meeting (oral evidence session)*: draft online safety bill. London: UK Parliament, 2021. Available at: <https://committees.parliament.uk/event/5594>. Accessed at: 28 Oct. 2021.

¹⁰⁵ SANDER, B. Freedom of expression in the age of online platforms: the promise and pitfalls of a human-rights based approach to content moderation. *Fordham International Law Journal*, [S.l.], v. 43, n. 4, p. 939-1006, 2020.

As previously highlighted, the UNGPs have undertaken an experimentalist and polycentric approach. This goes in line with the new governance approaches, which understand regulatory decision-making processes not as unidirectional or authoritative, but rather as one that incorporate “new mechanisms of stakeholder participation and public accountability as a way to retain democratic legitimacy and ensure community responsiveness” of these processes¹⁰⁶. In the ICT sector, which operates across national boundaries and within diverse digital ecosystems, top-down legal rule may prove hard to enforce and provide for its own effectiveness¹⁰⁷. Moreover, addressing human rights challenges in the online world entail multiple obligation bearers that could contribute to decentralized enforcement¹⁰⁸. While we have seen that different voices representing communities, academia, think tanks, companies and governments, participated in the formation of the UNGPs, their contribution during the implementation phase have not been equivalent to each other. In fact, since its adoption, civil society organisations (CSOs) have played only a marginal role in the implementation of the UNGPs¹⁰⁹. For authors such as Tara Melish, UNGPs should incorporate civil society as a key actor in the implementation of business and human rights principles, through the incorporation of a fourth pillar “Participation”. To support this statement, she argues that the “most vocal and consistent critics of the UNGPs are human rights organisations, precisely the groups that have been pushing the longest and hardest for a more effective,

¹⁰⁶ MELISH, T.; MEIDINGER, E. Protect, respect, remedy and participate: “New Governance” lessons for the Ruggie Framework. In: MARES, Radu (ed.). *The UN guiding principles on business and human rights: foundations and implementations*. Leiden: Brill-Nijhoff, 2012.

¹⁰⁷ LAIDLAW, E. B. (ed.). Myth or promise? the corporate social responsibilities of online service providers for human rights. In: TADDEO, M.; FLORIDI, L. *The responsibilities of online service providers*. Springer: Cham, 2017.

¹⁰⁸ SIKKIINK, K. *The hidden face of rights: towards a politics of responsibilities*. London: Yale University Press, 2020. v. 10.

¹⁰⁹ Ultimately, the experimentalist and polycentric approach proposed by the UNGPs reflect the power relations that exist between governments, companies and civil society organizations, usually leaving the last group as the most vulnerable and unattended one. That’s why some authors have proposed to incorporate a fourth pillar in the business and human rights field, the so called “civil society pillar”. MELISH, T.; MEIDINGER, E. Protect, respect, remedy and participate: “New Governance” lessons for the Ruggie Framework. In: MARES, Radu (ed.). *The UN guiding principles on business and human rights: foundations and implementations*. Leiden: Brill-Nijhoff, 2012.

non-business-as-usual approach to corporate human rights abuse”¹¹⁰.

As we have previously noted, the tech sector has been part of the polycentric governance approach developed by the UNGPs, yet such an approach also shows the need for a more open space for CSOs in the field. The polycentric and experimentalist approach on business and human rights field has allowed, for example, the creation of multi-stakeholder initiatives such as the GNI and the development of initiatives coming from government and companies, such as previously highlighted government and companies’ policies, statements, and plans. Yet, as it will be further developed, there is still a gap in the potential impact that the UNGPs may have in the implementation of these initiatives. We have discussed some explanations for this gap such as the late entry of the tech sector in the UNGPs discussions, and the lack of existing evidence when it comes to show the human rights impacts of tech companies. One potential solution to address these issues could be to give a more prominent role to civil society organisations when it comes to implement, discuss and think about UNGPs in the digital environment. In this scenario, it is important to highlight the purpose of the B-Tech Project, which is precisely to engage “diverse stakeholders as part of a global process to produce guidance, tools and practical recommendations to advance implementation of the UN Guiding Principles on business and Human Rights in the technology sector”¹¹¹.

CSOs - mainly human rights groups - have increased their pressure to be part of the implementation process of business and human rights initiatives. Most argue that “genuine social transformation occurs only when affected communities themselves have the power and voice to engage decision-making processes that affect their lives, as active subjects of law, not mere objects”¹¹².

From the perspective of civil society organisation, the lack of influence in the implementation of the UNGPs has to do with power relations, arguing that the interest of states and companies are too strong, and that they have tended to ignore and misrepresent CSOs voices¹¹³. The frustration of organised civil society has resulted in further commitments in the field. They have led, for example, to the creation of the “Treaty Alliance” comprising some 600 NGOs from around the world, who have been deeply engaged in the discussions on a binding treaty on business and human rights¹¹⁴. There are also examples where civil society organisations have been able to conduct impact studies, be involved in the creation of National Action Plans, review due diligence and grievances procedures, usually leading to better results than those processes that did not include their participation¹¹⁵.

However, apart from states, corporations and CSOs, individuals may also have some responsibility in cases where their performance as users of digital platforms may put at risk the rights to privacy, free speech or other rights of the rest of the population. As put by Kahtryn Sikkink, the Cambridge Analytica Scandal and the rise of fake news has highlighted cases where our lack of care may “leave our friends and contacts vulnerable”, or where “disinformation is the result of individuals retweeting, posting, and forwarding the news”¹¹⁶. Considering that regulation is sometimes slower than technological development, there are good reasons to ground a place for individual responsibility where corporate power is not willing to make changes. Of course, there are many challenges regarding digital education and training that depend on state regulation (“digital civics”), but one should not completely avoid individual responsibility in the face of complex decentralized problems such as misinformation or digital privacy, “where the

¹¹⁰ MELISH, T. Putting “human rights” back into the UN guiding principles on business and human rights: shifting frames and embedding participation rights. In: RODRÍGUEZ-GARAVITO, C. *Business and human rights: beyond the end of the beginning*. Cambridge: Cambridge University Press, 2017.

¹¹¹ B-TECH FOUNDATION PAPER. *Designing and implementing effective company-based grievance mechanisms*. 2020. Available at: <https://www.ohchr.org/Documents/Issues/Business/B-Tech/access-to-remedy-company-based-grievance-mechanisms.pdf>. Accessed on: 12 Nov. 2021.

¹¹² MELISH, T. Putting “human rights” back into the UN guiding principles on business and human rights: shifting frames and embedding participation rights. In: RODRÍGUEZ-GARAVITO, C. *Business and human rights: beyond the end of the beginning*. Cam-

bridge: Cambridge University Press, 2017.

¹¹³ MARTENS, J. *Corporate influence on the business and human rights agenda of the United Nations working paper*. Aachen: Bischöfliches Hilfswerk Misereor e.V., 2014.

¹¹⁴ DE SCHUTTER, O. Towards a New Treaty on Business and Human Rights. *Business and Human Rights Journal*, [S.L.], v. 1, n. 1, p. 41-67, 2016.

¹¹⁵ MELISH, T. Putting “human rights” back into the UN guiding principles on business and human rights: shifting frames and embedding participation rights. In: RODRÍGUEZ-GARAVITO, C. *Business and human rights: beyond the end of the beginning*. Cambridge: Cambridge University Press, 2017.

¹¹⁶ SIKKIINK, K. *The hidden face of rights: towards a politics of responsibilities*. London: Yale University Press, 2020. v. 10.

concept of human rights doesn't take us very far"¹¹⁷. Although UNGPs do not address the role of individual users, its polycentric and experimentalist approach should consider what role individuals/users could play in a multi-stakeholder approach to content moderation policy.

5 Conclusions

Through this paper we have seen that there is a gap between the principles promoted by the UNGPs and current regulations in the tech sector, especially among internet platforms. The paper further demonstrates the increasing involvement and commitment of the tech sector in discussions related to business and human rights, as illustrated by initiatives such as GNI or B-Tech Project. These commitments are shown through companies' policies and transparency reports, through novel forms of private governance, through civil society engagement in these issues and governments' increasing interest in developing different forms of regulation. While these examples can be seen as progress in the field, there is still much to do when it comes to make the most of the polycentric approach of the UNGPs.

First, governments have not made too much when it comes to business and human rights in the digital environment. The latter has been demonstrated by a lack of attention to business and human rights issues in the tech sector in NAPs and in a somehow improvised intermediary liability legislation which, in some cases, has proven to be openly against freedom of expression. The current status of content regulation, which attempts to materialize the duty to protect human rights, can be described in an interesting way: paraphrasing the idea that some companies are "too big to fail", coined after the economic crisis of 2007-8, we can label the challenges that arise for online platforms as "too late to regulate". Indeed, current calls for "slowing down the platforms" or design platforms at a "human scale" seem out of touch with the size, scale and character of the business models of the big competitors in this market. On the other hand, the developments that come from companies have been mostly reactive and not preventive. The example of Facebook clarifies this

assertion: under pressure from diverse stakeholders, the recent creation of the Oversight Board "marks the first platform-scaled moment of transnational internet adjudication of online speech"¹¹⁸. However, these novel forms or private regulation seem to be impracticable when confronted with issues of scale, which are embedded in the design features of the biggest online platforms in an increasingly concentrated market. In this scenario, private regulatory attempts, although well-intended, could be deemed as desperate attempts to mitigate the risks of in-built features that were originally designed to harm the conditions of a "democratic culture", allegedly the modern purpose of free speech protection. Moreover, most of the efforts undertaken by platforms have focused on improving their Terms of Service and the quality of their decision-making, to the detriment of implementing ways of challenging wrongful decisions and, if appropriate, remedied.

If UNGPs are to be considered as a normative framework of universal character, applicable to online platforms of all sizes, which operate across national boundaries, there needs to be more research and initiatives on the ways in which these principles could be applied to the complex issues of content regulation/moderation. The non-binding nature of these principles has not been an obstacle for the development of a modern consensus that online platforms should definitely respect human rights, especially after periodical scandals that arise with the big competitors in this billionaire industry. The current wave of transnational and national regulatory attempts of online content needs to start from a human rights-based approach, and the UNGPs serve as a good starting point. However, at some point, as we have argued here, the human rights discourse needs to accommodate issues of personal and collective responsibility for a complex and decentralized problem such as global free expression, which has become even more complex after a global pandemic that has accelerated the digital revolution.

¹¹⁷ SIKKIINK, K. *The hidden face of rights: towards a politics of responsibilities*. London: Yale University Press, 2020. v. 10.

¹¹⁸ KLONICK, K. The Facebook Oversight Board: creating an independent institution to adjudicate online free expression. *The Yale Law Journal*, [S.l.], v. 129, p. 2418-2499, 2020.

Table 1 Intermediary Liability Regimes

Liability regime	Summary	Examples
Strict Liability	Platforms are liable for unlawful or harmful content made available by users, even if they are not aware of the content	Thailand (Section 15 of the Computer Crimes Act 2007) China Cybersecurity Law 2017
Conditional Liability	Platforms are not held liable for unlawful or harmful content made available by users on their platforms provided they do not have any knowledge of the content or, if they do have knowledge, have acted expeditiously to remove it.	European Union's E-Commerce Directive (2000) United States (Section 512 of the Digital Millennium Copyright Act 1998) South Africa's (Chapter XI of the Electronic Communications and Transactions Act 2002)
Broad Liability	Platforms are, as a general rule, not held liable for unlawful or harmful content made available on their platforms, even if they are aware of the content. Some limited exceptions may exist, such as for certain specified crimes or intellectual property.	United States (Section 230 of the Communications Decency Act)

Source: Charles Bradley and Richard Wingfield, "A Rights-Respecting Model of Online Content Regulation by Platforms" (Global Partners Digital, 2018).

Table 2: Rule Enforcement data Provided

Twitter	Facebook	Google	YouTube	Oath
Reporting of data on the enforcement of community guidelines was published for the first time in 2018.	Data provided on enforcement of policies related to: Adult Nudity and Sexual Activity Hate Speech Terrorist Propaganda (ISIS, al-Qaeda and affiliates) Fake Accounts Spam Violence and Graphic Content	Reporting on enforcement of community standards is confined to YouTube (see YouTube section). Reporting of content removal appears to be confined to that required under law. Reporting on application of internal content regulation policy does not seem to be present.	Reporting of the enforcement of YouTube's community standards is structured by content type (Channel, video or comment) and removal reason. The report gives an overview of the source of "flags" about problematic content and provides case studies of the enforcement of the enforcement of YouTube's policies on violent extremism and child safety.	Oath's transparency reporting is very limited and only covers government requests for data and the removal of content as well as data on content removal for copyright and trademark infringement. It does not provide any detail regarding the enforcement of their internal content regulation policies.

Source: own creation.

References

ARNOLD, D. G., Transnational corporations and the duty to respect basic human rights. *Business Ethics Quarterly*, [S.l.], v. 20, n. 3, p. 371-399, 2010.

BAKOS, Y.; MAROTTA-WURGLER, F.; TROSSEN, D. R. *Does anyone read the fine print?* consumer attention to standard form contracts. Law & Economics Research, paper series. 2009. Ideas. Available at: <https://ideas.repec.org/p/net/wpaper/0904.html>. Accessed on: 10 November 2021.

BALKIN, J. Free speech in the algorithmic society: Big Data, Governance, and new school speech regulation. *Davis Law Review*, [S.l.], v. 51, p. 1149-1210, 2020.

BRADLEY, C.; WINGFIELD, R. *A rights-respecting model of online content regulation by platforms*. London: Global Partners Digital, 2018.

B-TECH FOUNDATION PAPER. *Bridging governance gaps in the age of technology*: key characteristics of the State duty to protect. 2020. Available at: <https://www.ohchr.org/Documents/Issues/Business/B-Tech/b-tech-foundational-paper-state-duty-to-protect.pdf>. Accessed on: 10 Nov. 2021.

B-TECH FOUNDATION PAPER. *Designing and implementing effective company-based grievance mechanisms*. 2020. Available at: <https://www.ohchr.org/Documents/Issues/Business/B-Tech/access-to-remedy-company-based-grievance-mechanisms.pdf>. Accessed on: 12 Nov. 2021.

B-TECH FOUNDATION PAPER. *Key characteristics of business respect for human rights*. 2020. Available at: <https://www.ohchr.org/Documents/Issues/Business/B-Tech/key-characteristics-business-respect.pdf>. Accessed on: 12 Nov. 2021.

BYGRAVE, L. *Internet governance by contract*. Oxford: Oxford University Press, 2015.

CARDIA, Ana Cláudia Ruy. Reparação de vítimas à luz de um tratado sobre empresas e direitos humanos. *Revista de Direito Internacional*, Brasília, v. 15, n. 2, p. 2-11, 2018.

CELESTE, E. Social media bans: what platforms can learn from National Courts? 2021. *Blog British Association of Comparative Law*. Available at: <https://british-association-comparative-law.org/2021/05/21/social-media->

- bans-what-platforms-can-learn-from-national-courts-by-edoardo-celeste/. Accessed on: 18 Oct. 2022.
- CENTER FOR STUDIES ON FREEDOM OF EXPRESSION AND ACCESS TO INFORMATION [CELE]. *Submission to the UN SR on the Protection and Promotion of Freedom of Opinion and Expression*. Palermo, 2017.
- CERQUEIRA, D.; MONTGOMERY, A. Extraterritorial obligations: a missing component of the UN Guiding Principles that should be addressed in a binding treaty on business and human rights. 2018. *Due Process Foundation Blog*. Available at: <https://dplfblog.com/2018/02/08/extraterritorial-obligations-a-missing-component-of-the-un-guiding-principles-that-should-be-addressed-in-a-binding-treaty-on-business-and-human-rights/>. Accessed on: 20 Mar. 2022.
- CHEN, A. *Inside Facebook's outsourced anti-porn and gore brigade, where 'camel toes' are more offensive than 'crushed heads'*. Available at: <http://gawker.com/5885714/inside-facebooks-outsourced-anti-porn-and-gore-brigade-where-camel-toes-are-more-offensive-than-crushed-heads>. Accessed on: 20 Nov. 2021.
- COHEN, J. E. *Between truth and power*. [S.l.]: Oxford University Press, 2019.
- COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTÉS [CNIL]. *Case n. C-507/17*. 2019.
- COMMITTEE ON ECONOMIC, SOCIAL AND CULTURAL RIGHTS. *Report n. E/C.12/GC/24*. General comment No. 24. On State obligations under the International Covenant on Economic, Social and Cultural Rights in the context of business activities. 2017, paragraphs 25-28.
- COMMITTEE ON ECONOMIC, SOCIAL AND CULTURAL RIGHTS. *Report n. E/C.12/GC/24*. General comment No. 24. On State obligations under the International Covenant on Economic, Social and Cultural Rights in the context of business activities. 2017, paragraphs 25-28.
- CONSUMER NEWS AND BUSINESS CHANNEL. *Why content moderation costs billions and is so tricky for Facebook, Twitter, YouTube and others*. 2021. Available at: <https://www.cnbc.com/2021/02/27/content-moderation-on-social-media.html>. Accessed on: 17 Nov. 2021.
- DE BÚRCA, G.; KEOHANE, R. O.; SABEL, C. New modes of pluralist global governance. *International Law and Politics*, New York, v. 45, p. 723-739, 2013.
- DE FELICE, D.; GRAF, A. The potential of National Action Plans to implement human rights norms: an early assessment with respect to the UN Guiding Principles on Business and human rights. *Journal of Human Rights Practice*, [S.l.], v. 7, n. 1, p. 40-71, 2015.
- DE FILIPPI, P.; BELLI, L. The Law of the Cloud v the Law of the Land: challenges and opportunities for innovation. *European Journal of Law and Technology*, [S.l.], v. 3, n. 2, p. 23, 2012.
- DE GREGORIO, G. Democratising online content moderation a constitutional framework. *Computer Law & Security Review*, [S.l.], v. 2, p. 105374, 2020.
- DE SCHUTTER, O. Towards a New Treaty on Business and Human Rights. *Business and Human Rights Journal*, [S.l.], v. 1, n. 1, p. 41-67, 2016.
- DEIBERT, R. *et al.* (eds.). *Access Controlled: the shaping of power, rights, and rule in cyberspace*. London: The MIT Press, 2010.
- DEIBERT, R. *et al.* (eds.). *Access denied the practice and policy of global internet filtering*. London: The MIT Press, 2008.
- DIGITAL MARKETS ACT. EU COMMISSION. *Europe fit for the digital age: new online rules for platforms*. 2021. Available at: https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/digital-services-act-ensuring-safe-and-accountable-online-environment/europe-fit-digital-age-new-online-rules-platforms_en. Accessed on: 15 Mar. 2022.
- DUARTE, N.; LLANSO, E.; LOUP, A. *Mixed messages? the limits of automated social media content analysis*. 2018. Available at: <https://apo.org.au/sites/default/files/resource-files/2017-11/apo-nid240471.pdf>. Accessed on: 16 Mar. 2022.
- DUNSTAN, A. Our Human Rights Impact Assessment of Facebook in Myanmar. November 2018. *Business for Social Responsibility*. Available at: <https://www.bsr.org/en/our-insights/blog-view/facebook-in-myanmar-human-rights-impact-assessment>. Accessed on: 10 Oct. 2021.
- EUROPEAN COURT OF HUMAN RIGHTS. *Application No. 5493/7 Handyside v the United Kingdom*. 1976.

- FAZLIOGLU, M. Forget me not: the clash of the right to be forgotten and freedom of expression on the Internet. *International Data Privacy Law*, [S.l.], v. 3, n. 3, p. 149-157, 2013.
- FINLAND. Government of Finland. *National Action Plan for the implementation of the UN Guiding Principles on Business and Human Rights*. 2014.
- FLEW, T.; MARTIN, F.; SUZOR, N. Internet regulation as media policy: rethinking the question of digital communication platform governance. *Journal of Digital Media & Policy*, [S.l.], v. 10, n. 1, p. 33-50, 2019.
- FREEDOM HOUSE. *Freedom on the Net 2018*. 2018. Available at: https://freedomhouse.org/sites/default/files/FOTN_2018_Final.pdf. Accessed on: 15 Oct. 2021.
- FREEDOM HOUSE. *Freedom on the Net 2021: the global drive to control big tech*. 2021. Available at: <https://freedomhouse.org/report/freedom-net/2021/global-drive-control-big-tech>. Accessed on: 15 Nov. 2021.
- GILLESPIE, T. Platforms are not Intermediaries. *Georgetown Law Technology Review*, [S.l.], v. 2, n. 2, p. 198-216, 2018.
- GLOBAL NETWORK INITIATIVES. *GNI Principles on freedom of expression and privacy*. 2017. Available at: <https://globalnetworkinitiative.org/gni-principles/>. Accessed 10 Nov. 2021.
- GLOBAL NETWORKING INITIATIVE. *Content regulation and human rights: analysis and recommendations*. 2020. Available at: <https://globalnetworkinitiative.org/wp-content/uploads/2020/10/GNI-Content-Regulation-HR-Policy-Brief.pdf>. Accessed on: 10 Nov. 2021.
- GORWA, R.; BINNS, R.; KATZENBACH, C. Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society*, [S.l.], v. 7, n. 1, p. 15, 2020. Available at: <https://doi.org/10.1177/2053951719897945>. Accessed on: 10 Nov. 2021.
- GRASSER, U.; SCHULZ, W. *Governance of online intermediaries: observations from a series of national case studies*. Cambridge: The Berkman Center for Internet & Society Research Publication, [S.l.], v. 18, p. 283, 2015.
- HAUGEN, F. *Formal meeting (oral evidence session): draft online safety bill*. London: UK Parliament, 2021. Available at: <https://committees.parliament.uk/event/5594>. Accessed at: 28 Oct. 2021.
- HUMAN RIGHTS COMITEE. *Report n. CCPR/C/GC/34*. General Comment 34. United Nations, 2011, paragraph 12. See also: UN. Report n. A/HRC/38/35. Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression - regulation of user-generated online content. 2018, paragraph 7.
- HUMAN RIGHTS COMITEE. *Report n. CCPR/C/GC/34*. General Comment 34. United Nations, 2011, paragraph 7.
- INTERNATIONAL LABOUR ORGANIZATION. *ILO Declaration on Fundamental Principles and Rights at Work*. International Labour Organization, 1998. Available at: www.ilo.org/declaration/lang-en/index.htm. Accessed on: 5 Nov. 2021.
- IRELAND. Government of Ireland. *National Plan on Business and human rights 2017-2020*. 2017.
- JØRGENSEN, R. F. *A human rights-based approach to social media platforms*. 2021. Available at: <https://berkeleycenter.georgetown.edu/responses/a-human-rights-based-approach-to-social-media-platforms>. Accessed on: 10 Oct. 2021.
- JØRGENSEN, R. F.; ZULETA, L. Private governance of freedom of expression on social media platforms: EU content regulation through the lens of human rights standards. *Nordicom Review*, Copenhagen, v. 41, n. 1, p. 51-67, 2020.
- KHONDKER, H. H. Role of the New Media in the Arab Spring. *Globalizations*, v. 8 n. 5, p. 675-679, 2011.
- KLONICK, K. The Facebook Oversight Board: creating an independent institution to adjudicate online free expression. *The Yale Law Journal*, [S.l.], v. 129, p. 2418-2499, 2020.
- KLONICK, Kate. Law and technology: content moderation modulation. *Communications of the ACM*, [S.l.], v. 64, n. 1, p. 29-31, January 2021. Available at: <https://cacm.acm.org/magazines/2021/1/249450-content-moderation-modulation/fulltext>. Accessed on: 15 Nov. 2021.
- KUCZERAWY, A. Intermediary liability & freedom of expression: recent developments in the EU notice & action initiative. *Computer Law and Security Review*, [S.l.], v. 31, n. 1, p. 46-56, 2015.
- LAIDLAW, E. B. (ed.). *Myth or promise? the corporate social responsibilities of online service providers*

- for human rights. In: TADDEO, M.; FLORIDI, L. *The responsibilities of online service providers*. Springer: Cham, 2017.
- LAIDLAW, E. B. Myth or promise? the corporate social responsibilities of online service providers for human rights (ed.). In: TADDEO, M.; FLORIDI, L. *The responsibilities of online service providers*. Springer: Cham, 2017.
- LOMBANA-BERMÚDEZ, A. *Moderation and sense of community in a youth-oriented online platform*: Scratch's governance strategy for addressing harmful Speech. 2017. Available at: <https://medium.com/berkman-klein-center/moderation-and-sense-of-community-in-a-youth-oriented-online-platform-scratchs-governance-eea-c6941e9c9>. Accessed on: 10 Nov. 2021.
- MACKINNON, R. *et al. Fostering freedom online: the role of internet intermediaries*. Paris: UNESCO Publishing, 2015.
- MARTENS, J. *Corporate influence on the business and human rights agenda of the United Nations working paper*. Aachen: Bischöfliches Hilfswerk Misereor e.V, 2014.
- MELISH, T. Putting "human rights" back into the UN guiding principles on business and human rights: shifting frames and embedding participation rights. In: RODRÍGUEZ-GARAVITO, C. *Business and human rights: beyond the end of the beginning*. Cambridge: Cambridge University Press, 2017.
- MELISH, T.; MEIDINGER, E. Protect, respect, remedy and participate: "New Governance" lessons for the Ruggie Framework. In: MARES, Radu (ed.). *The UN guiding principles on business and human rights: foundations and implementations*. Leiden: Brill-Nijhoff, 2012.
- NETHERLANDS. Government of Netherlands, *National Action Plan on Business and Human Rights*. 2013.
- NOLAN, Justine. Hardening soft law: are the emerging corporate social disclosure laws capable of generating substantive compliance with human rights?. *Revista de Direito Internacional*, Brasília, v. 15, n. 2, p. 64-83, 2018.
- OATH. *Our approach: built on seven pillars*. 2019. Available at: <https://www.oath.com/our-story/business-and-human-rights/our-program/>. Accessed on: 15 Jan. 2019.
- ORGANIZATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT. *OECD Guidelines for Multinational Enterprises*: 2011 Edition. Available at: <http://mneguidelines.oecd.org/text/>. Accessed on: 5 Nov. 2021
- PEN AMERICA. *Forbidden fees: government controls on social media in China*. 2018. Available at: https://pen.org/wp-content/uploads/2018/03/PENAmerica_Forbidden-Feeds-3.13-3.pdf. Accessed on: 18 Nov. 2021.
- POLAND. Government of Poland. *Polish National Action Plan for the Implementation of the United Nations Guiding Principles on Business and Human Rights 2017-2020*. 2017.
- PROGRESS on our commitment to transparency – META. 2021. Available at: <https://about.fb.com/news/2021/11/meta-transparency-report-h1-2021/>. Accessed on: 18 Nov. 2021.
- QIANG, X. How China internet police control speech on the Internet. *Radio Free Asia*, [S.l.], v. 24, 2008. Available at: https://www.rfa.org/english/commentaries/china_internet-11242008134108.html. Accessed on: 29 Jan. 2019.
- RAMÍREZ, I. *Online content regulation and competition policy*. 2020. Available at: <https://orgs.law.harvard.edu/antitrust/2020/12/03/online-content-regulation-and-competition-policy/>. Accessed on: 10 Nov. 2021.
- RANKING DIGITAL RIGHTS. *Corporate accountability index 2018*. 2018. Available at: <https://rankingdigital-rights.org/index2018/report/executive-summary/>. Accessed on: 12 Nov. 2021.
- RANKING DIGITAL RIGHTS. *Corporate Accountability Index 2020*, 2020. Available at: <https://rankingdigital-rights.org/index2020>. Accessed on: 10 Nov. 2021.
- RODRÍGUEZ-GARAVITO, C. *Business and human rights: beyond the end of the beginning*. Cambridge: Cambridge University Press, 2017.
- RUGGIE, J. *Facebook in the rest of the world*. 2018. Available at: https://media.business-humanrights.org/media/documents/files/documents/John_Ruggie_Facebook_15_Nov_2018.pdf. Accessed on: 5 Oct. 2021.
- RUGGIE, J. *Just Business: multinational corporations and human rights*. New York: Norton & Company, 2013.
- SAMWAY, M. The Global Network Initiative: how can companies in the information and communications technology industry respect human rights? In: BAUMANN-PAULY, D.; NOLAN, J. (ed.). *Business*

- and human rights: from principles to practice. London: Routledg, 2016. p. 136-140.
- SANDER, B. Freedom of expression in the age of online platforms: the promise and pitfalls of a human-rights based approach to content moderation. *Fordham International Law Journal*, [S.l.], v. 43, n. 4, p. 939-1006, 2020.
- SANTARELLI, Nicolás Carrillo. Declaraciones empresariales “voluntarias” sobre derechos humanos, y la necesidad de una regulación institucional (internacional e interna) externa. *Revista de Direito Internacional*, Brasília, v. 16, n. 3, p. 23-49, 2019.
- SCHOFIELD, J. Microsoft takes down Chinese blogger. *The Guardian*, [S.l.], 2 January 2006. Available at: <https://www.theguardian.com/technology/blog/2006/jan/04/microsofttakes>. Accessed on: 24 Jan. 2019.
- SIKKIINK, K. *The hidden face of rights: towards a politics of responsibilities*. London: Yale University Press, 2020. v. 10.
- SWEDEN. Government Offices of Sweden. *Action plan for business and human rights*. Here and in the following footnotes, references to NAPs can be found in the OHCHR website. 2015. <https://www.ohchr.org/en/issues/business/pages/nationalactionplans.aspx>. Accessed on: Nov. 2021.
- THOMSOM, C. Google’s China problem (and China’s Google problem). *New York Times*, New York, 23 April 2006. Available at: <https://www.nytimes.com/2006/04/23/magazine/23google.html>. Accessed on: 24 Jan. 2019.
- UK PARLIAMENT. *Joint Committee on the Draft Online Safety Bill*. London, 2021. Available at: <https://publications.parliament.uk/pa/jt5802/jtselect/jtonlinesafety/129/12902.htm>. Accessed on: 10 Mar. 2022.
- UNITED KINGDO. Government of United Kingdo. *Good business: implementing the UN guiding principles on business and human rights*. 2016.
- UNITED NATIONS. Human Rights Council. *Report n. A/HRC/38/35*. Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, David Kaye. Human Rights Council, 2018.
- UNITED NATIONS. Human Rights Council. *Report n. A/HRC/RES/32/13*. The promotion, protection and enjoyment of human rights on the Internet: resolution. Human Rights Council, 2016.
- UNITED NATIONS. *Moderating online content: fighting harm or silencing dissent?* 2021. Available at: <https://www.ohchr.org/EN/NewsEvents/Pages/Online-content-regulation.aspx>. Accessed on: 12 Nov. 2021.
- UTLU, D. Public policy and digital technologies: the role of national human rights institutions in achieving policy coherence. *B-Tech*, [S.l.], May 2021. Available at: <https://www.ohchr.org/Documents/Issues/Business/B-Tech/b-tech-blog-policy-coherence-nhris-tech.pdf>. Accessed on: 10 Nov. 2021.
- VAN DIJK, P. *et al. Theory and practice of the European Convention on Human Rights*. Antwerp: Oxford, 2006.
- VENTURINI J. *et al. Terms of service and human rights: an analysis of online platform contracts*. Rio de Janeiro: Revan, 2016.
- VIGEO EIRIS. *Human rights in a globalised world: why do companies need to pay more attention?* 2018. Available at: https://www.vigeo-eiris.com/wp-content/uploads/2018/11/2018_Human-rights-study_VFok.pdf. Accessed on: 10 Nov. 2021.
- ZITTRAIN, J. A history of online Gatekeeping. *Harvard Journal of Law and Technology*, [S.l.], v. 19, n. 2, p. 253-298, 2006.
- ZUCKERBERG, M. *A Blueprint for Content Governance and Enforcement*. 2018. Available at: <https://www.facebook.com/notes/mark-zuckerberg/a-blueprint-for-content-governance-and-enforcement/10156443129621634/>. Accessed on: 15 Jan. 2019.
- RUGGIE, J. *Just Business: multinational corporations and human rights*. New York: Norton & Company, 2013.
- SAMWAY, M. The Global Network Initiative: how can companies in the information and communications technology industry respect human rights? In: BAUMANN-PAULY, D.; NOLAN, J. (ed.). *Business and human rights: from principles to practice*. London: Routledg, 2016. p. 136-140.
- SANDER, B. Freedom of expression in the age of online platforms: the promise and pitfalls of a human-rights based approach to content moderation. *Fordham International Law Journal*, [S.l.], v. 43, n. 4, p. 939-1006, 2020.

- SANTARELLI, Nicolás Carrillo. Declaraciones empresariales “voluntarias” sobre derechos humanos, y la necesidad de una regulación institucional (internacional e interna) externa. *Revista de Direito Internacional*, Brasília, v. 16, n. 3, p. 23-49, 2019.
- SCHOFIELD, J. Microsoft takes down Chinese blogger. *The Guardian*, [S.l.], 2 January 2006. Available at: <https://www.theguardian.com/technology/blog/2006/jan/04/microsofttakes>. Accessed on: 24 Jan. 2019.
- SIKKIINK, K. *The hidden face of rights: towards a politics of responsibilities*. London: Yale University Press, 2020. v. 10.
- THOMSOM, C. Google’s China problem (and China’s Google problem). *New York Times*, New York, 23 April 2006. Available at: <https://www.nytimes.com/2006/04/23/magazine/23google.html>. Accessed on: 24 Jan. 2019.
- UK PARLIAMENT. *Joint Committee on the Draft Online Safety Bill*. London, 2021. Available at: <https://publications.parliament.uk/pa/jt5802/jtselect/jtonlinesafety/129/12902.htm>. Accessed on: 10 Mar. 2022.
- UNITED NATIONS. Human Rights Council. *Report n. A/HRC/38/35*. Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, David Kaye. Human Rights Council, 2018.
- UNITED NATIONS. Human Rights Council. *Report n. A/HRC/RES/32/13*. The promotion, protection and enjoyment of human rights on the Internet: resolution. Human Rights Council, 2016.
- UTLU, D. Public policy and digital technologies: the role of national human rights institutions in achieving policy coherence. *B-Tech*, [S.l.], May 2021. Available at: <https://www.ohchr.org/Documents/Issues/Business/B-Tech/b-tech-blog-policy-coherence-nhris-tech.pdf>. Accessed on: 10 Nov. 2021.
- VAN DIJK, P. *et al. Theory and practice of the European Convention on Human Rights*. Antwerp: Oxford, 2006.
- VENTURINI J. *et al. Terms of service and human rights: an analysis of online platform contracts*. Rio de Janeiro: Revan, 2016.
- VIGEO EIRIS. *Human rights in a globalised world: why do companies need to pay more attention?* 2018. Available at: https://www.vigeo-eiris.com/wp-content/uplo-ads/2018/11/2018_Human-rights-study_VFok.pdf. Accessed on: 10 Nov. 2021.
- ZITTRAIN, J. A history of online Gatekeeping. *Harvard Journal of Law and Technology*, [S.l.], v. 19, n. 2, p. 253-298, 2006.
- ZUCKERBERG, M. *A Blueprint for Content Governance and Enforcement*. 2018. Available at: <https://www.facebook.com/notes/mark-zuckerberg/a-blueprint-for-content-governance-and-enforcement/10156443129621634/>. Accessed on: 15 Jan. 2019.

Para publicar na Revista de Direito Internacional, acesse o endereço eletrônico
www.rdi.uniceub.br ou www.brazilianjournal.org.
Observe as normas de publicação, para facilitar e agilizar o trabalho de edição.