Don't Shoot the Message: Regulating Disinformation Beyond Content[1]

Não Dispare a Mensagem: Regulamentando a Desinformação Além do Conteúdo

CLARA IGLESIAS KELLER[2]

Leibniz Institute for Media Research | Hans-Bredow-Institute (HBI). Hamburgo, Alemanha.

ABSTRACT: This paper approaches regulatory strategies against disinformation with two main goals: (i) exploring the policies recently implemented in different legal contexts to provide insight into both the risks they pose to free speech and their potential to address the rationales that motivated them, and (ii) to do so by bridging policy debates and recent social and communications studies findings on disinformation. An interdisciplinary theoretical framework informs both the paper's scope (anchored on understandings of *regulatory strategies* and of *disinformation*) and the analysis of the legitimate motivations for states to establish statutory regulation that aims at disinformation. Departing from this analysis, I suggest an organisation of recently implemented and proposed policies into three groups based on their regulatory target: content, data, and structure. Combining the analysis of these three types of policies with the theoretical framework, I will argue that, in the realm of statutory regulation that aims at disinformation. Departing from this analysis, I suggest an organisation of recently implemented and proposed policies into three groups based on their regulatory target: content, data, and structure. Combining the analysis of these three types of policies with the theoretical framework, I will argue that, in the realm of statutory regulation, state action is better off targeted at data or structure, as aiming at content represents disproportional risks to freedom of expression. Furthermore, content targeted regulation shows little potential to address the structural transformations on the public sphere of communications that, among other factors, influence current practices of production and spread of disinformation.

KEYWORDS: Disinformation; regulation; regulatory strategies; fake news; digital platforms.

RESUMO: Esse artigo aborda estratégias regulatórias contra a desinformação com dois objetivos principais. O primeiro é explorar regulações recentemente implementadas em diferentes contextos legais, com foco tanto nos riscos que representam para a liberdade de expressão, quanto no seu

1 I would like to thank my colleagues at the Digital Disinformation Hub of the Leibniz Institute for Media Research, Stephan Dreyer and Amélie Heldt, for the conversations that inspired my ideas and supported the findings in this paper. Furthermore, I would like to thank Stephan Dreyer and Leonard Kamps for their revisions and suggestions, as as well as Lena Hinrichs and Mara Barthelmes for their help with the empirical research.

2 Orcid: https://orcid.org/0000-0001-6454-2616.

potencial para solucionar as questões que as motivaram. O segundo objetivo é fazer tal análise através de uma conexão entre debates regulatórios e os recentes achados das ciências sociais e das ciências da comunicação sobre desinformação. Este marco teórico interdisciplinar informa tanto o escopo do artigo (ancorado em entendimentos de "estratégias regulatórias" e de "desinformação") quanto a análise das motivações legítimas para que os Estados estabeleçam regulação estatutária que vise o combate à desinformação. Partindo desta análise, eu sugiro uma organização de políticas recentemente implementadas e propostas em três grupos, de acordo com o alvo regulatório: conteúdo, dados e estrutura. Combinando a análise desses três tipos de políticas com o marco teórico, eu argumento que, no âmbito da regulamentação estatutária, a ação estatal é melhor direcionada a dados ou estrutura, pois o objetivo de conteúdo não apenas representa riscos desproporcionais à liberdade de expressão, mas também mostra pouco potencial para abordar as transformações estruturais nas comunicações que marcam a esfera pública atual.

PALAVRAS-CHAVE: Desinformação; regulação; estratégias regulatórias; *fake news;* plataformas digitais.

INTRODUCTION

Following the 2016 Brexit referendum, a series of electoral processes drew the world's attention to the possibility of the intentional and massive spread of false information through digital means[3]. Shortly after, this sort of practice soon revealed itself to be a threat in and outside of electoral processes, as digital disinformation became a familiar contingency to various debates. Among public health, climate change, historic revisionism and others, disinformation revealed itself to be a dynamic phenomenon that has become more and more entrenched in contemporary communications by presenting itself in different forms.

As a result, we have witnessed different academic research strains aimed at understanding the multiple dimensions of disinformation and assessing suitable responses in recent years. Multidisciplinary works have

3 CADWALLADR, Carole, The great British Brexit robbery: how our democracy was hijacked, *The Guardian*, 2017; EVANGELISTA, Rafael; BRUNO, Fernanda, WhatsApp and political instability in Brazil: targeted messages and political radicalisation, *Internet Policy Review*, v. 8, n. 4, 2019; FARIS, Robert M. *et al*, *Partisanship, Propaganda, and Disinformation: Online Media and the 2016 U.S. Presidential Election*, Cambridge, U.S.: Berkman Klein Center for Internet & Society at Harvard University, 2017; DAS, Anupam; SCHROEDER, Ralph, Online disinformation in the run-up to the Indian 2019 election, *Information, Communication & Society*, p. 1-17, 2020.

covered disinformation's conceptual implications[4], different forms[5], and potential effects[6], as well as possible countermeasures[7]. While some of these strings are being untangled, other aspects remain under dispute or are simply not proven, like the debated existence of political bots (whose functions would include spreading disinformation)[8] or the assertion of a correlation between such strategies and poll results[9]. Even as these and other questions are publicly raised, disinformation remains a relevant phenomenon that bears fundamental societal risks.

In all the varied approaches to the topic, some common ground has been found in the idea that there is no single silver bullet. As a multi-layered phenomenon, disinformation is an expression of tensions and transformations that, despite being linked to digitalisation, result from a broader constellation of technological, social, and political factors[10]. For the policy debate, this means that fighting disinformation requires different responses from different actors. So, for instance, fact-checking and media literacy can be performed by journalistic or civil society organisations, and digital platforms can implement their own moderation and certification procedures. As I argue in this article, there is also an important role for state regulation.

This is where we start to enter even more complex territory, as governmental action towards digital disinformation mingles with the ever-delicate exercise of regulating freedom of expression – i.e., establishing rules regarding what can and cannot be said, published, and distributed. Even

4 MARWICK, Alice; LEWIS, Rebecca, *Media manipulation and disinformation online*, [s.l.]: Data & Society Research Institute, 2020.

5 ROSSINI, Patrícia *et al*, Dysfunctional information sharing on WhatsApp and Facebook: The role of political talk, cross-cutting exposure and social corrections, *New Media & Society*, v. 23, n. 8, p. 2430-2451, 2021; DAN, Viorela *et al*, Visual Mis- and Disinformation, Social Media, and Democracy, *Journalism & Mass Communication Quarterly*, v. 98, n. 3, p. 641-664, 2021.

6 FARIS *et al*, *Partisanship, Propaganda, and Disinformation: Online Media and the 2016 U.S. Presidential Election*; KARPF, David, On Digital Disinformation and Democratric Myths, *MediaWell, Social Science Research Council*, 2019.

7 NEO, Ric, The International Discourses and Governance of Fake News, *Global Policy*, v. 12, n. 2, p. 214-228, 2021; SCHULZ, WOLFGANG, *Roles and Responsibilities of Information Intermediaries: Fighting Misinformation as a Test Case for Human-Rights Respecting Governance of Social Media Platforms*, [s.l.]: Hoover Institution, Stanford University, 2019.

8 RAUCHFLEISCH, Adrian; KAISER, Jonas, The False positive problem of automatic bot detection in social science research, *PLOS ONE*, v. 15, n. 10, p. e0241045, 2020.

9 FARIS *et al*, *Partisanship, Propaganda, and Disinformation: Online Media and the 2016 U.S. Presidential Election*.

10 This is a research perspective on technological transformation called "mediated democracy", which will be further explored in Section 3 of this paper. In general, see HOFMANN, Jeanette, Mediated democracy – Linking digital technology to political agency, *Internet Policy Review*, v. 8, n. 2, 2019.

though every right can be subject to restrictions, freedom of expression is an indisputable pillar of modern democracies, and the line that separates these restrictions from state censorship can be thin.

In light of this, this paper's main goals are (i) to explore regulatory strategies against disinformation currently on the table and provide insight into both the risks they pose to free speech and their potential to address the rationales that motivated them and (ii) to do so by bridging policy debates and recent social and communications studies findings on disinformation. To do that, I suggest to organise recently implemented and proposed policies into three groups based on their regulatory target: content, data, and structure. After this analysis, I will argue that, in terms of statutory regulation, state action is better off targeted at data or structure, as aiming at content represents disproportional risks to freedom of expression. Furthermore, content targeted regulation shows little potential to address the structural transformations on the public sphere of communications that, among other factors, condition current practices of production and spread of disinformation.

The paper is organised as follows. Section 1 explains the scope decisions and technical definitions that ground the key objects of study: regulatory strategies and disinformation. Section 2 approaches the regulation of disinformation, starting with an outline of the adequate regulatory rationales under consideration (item 2.1). Item 2.2 sets out a set of regulatory strategies against disinformation, classified according to their regulatory target. Item 3 analyses these strategies to extract key takeaways for the policy debate, and the paper is finished with the conclusions.

1 DEFINITIONS AND SCOPE

This paper rests on two main scope decisions. The first, regarding the concept of *regulatory strategies*, refers to the promulgation of rules by governments, accompanied by mechanisms for monitoring and enforcement[11]. While recognising the need for actions from different stakeholders, I will focus on "the main instruments that the state can use to regulate directly"[12], i.e., statutory regulation proposed to or enacted by

11 BLACK, J., Decentring Regulation: Understanding the Role of Regulation and Self-Regulation in a "Post-Regulatory" World, *Current Legal Problems*, v. 54, n. 1, p. 103-146, 2001.

12 BALDWIN, Robert; CAVE, Martin; LODGE, Martin, *Understanding regulation: theory, strategy, and practice*, 2nd ed. New York: Oxford University Press, 2012, p. 105.

parliament, the declared goal, rationales, or formal or informal motivations of which relate to the spread of online disinformation.

Disinformation is understood as "false or misleading information that is intentionally spread for profit, to create harm, or to advance political or ideological goals"[13]. This definition is in tune with conceptual literature on the topic, which attempts to make sense of the differing yet blurred communication practices that are part of what has been referred to as an "information disorder"[14]. Since the expansion of practices related to the spread of false information through digital means, the terminology has evolved to allow greater accuracy and differentiation among varying phenomena.

Even though the term "fake news" is still used, its mixed applications restrict its theoretical and technical relevance[15]. The conceptual debate in the English language builds, in large part, on the distinction between disinformation and misinformation – the latter being commonly referred to as information that is "false by definition"[10] but has not been disseminated with a specific purpose to cause harm. Intent is recognised as the element that differentiates mis- and disinformation[16], as "[d]isinformation is meant to deceive, while misinformation may be inadvertent or unintentional"[17]. While these two concepts hold structural relevance, different classifications vary in the additional typologies to this list[18]. These taxonomies allow for the distinction among practices with one element – falsehood – in common but that still vary considerably in terms of the risks they entail for individual and collective rights.

13 MARWICK, Alice *et al*, *Critical Disinformation Studies – A Syllabus*, [s.l.]: Center for Information, Technology and Public Life – University of North Carolina at Chapel Hill, 2021.

14 WARDLE, Claire; DERAKHSHAN, Hossein, *Information Disorder: Toward an interdisciplinary framework for research and policy making*, [s.l.]: Council of Europe, 2017.

15 For exceptions, DAN, Viorela *et al*, Visual Mis- and Disinformation, Social Media, and Democracy, *Journalism & Mass Communication Quarterly*, v. 98, n. 3, p. 641-664, 2021.

16 For all, see WARDLE; DERAKHSHAN, *Information Disorder: Toward an interdisciplinary framework for research and policy making*.

17 GUESS, Andrew M.; LYONS, Benjamin A., Misinformation, Disinformation and Online Propaganda, *in*: PERSILY, Nathaniel; TUCKER, Joshua A. (Eds.), *Social Media and Democracy. The State of the Field, Prospects for Reform.*, [s.l.]: Cambridge University Press, 2020.

18 WARDLE; DERAKHSHAN, *Information Disorder: Toward an interdisciplinary framework for research and policy making*; GUESS; LYONS, Misinformation, Disinformation and Online Propaganda.; FARIS *et al*, *Partisanship, Propaganda, and Disinformation: Online Media and the 2016 U.S. Presidential Election*; EGELHOFER, Jana Laura; LECHELER, Sophie, Fake news as a two-dimensional phenomenon: a framework and research agenda, *Annals of the International Communication Association*, v. 43, n. 2, p. 97-116, 2019.

The distinction between mis- and disinformation holds fundamental relevance for the debate about regulatory policies in liberal democracies, where counteractions against these practices mingle with the ever-delicate exercise of regulating freedom of expression. Because classifying and countering disinformation intrinsically depends on a judgement on the substance of expression – i.e., it is false? – countermeasures hold an increased risk of promoting chilling effects or even censorship. Ultimately, conceptions of truth and fact are a matter of perspective, the meaning of which, in a democracy, should be fairly and equally disputed by society[19]. If statutory policy targets cases where the circulation of inaccurate or false information is not intentional, regulatory initiatives would be pushed even further away from legitimacy and compromise the "benefits of a noisy and unruly public arena"[20]. Throughout this paper, I will show that these risks may still not be neutralised even if we narrow policy debates down to disinformation.

Nonetheless, disinformation is still a phenomenon that bears fundamental societal risks, especially when it stands in the way of equal and fair participation in public debate. There are legitimate reasons for states to be concerned about and regulate disinformation, but they must be equally concerned that these interventions do not pose risks to freedom of speech.

2 REGULATING DISINFORMATION

2.1 Disinformation and regulatory rationales

Choosing regulatory strategies that impose minimal restrictions on free speech is in line with recent findings in social and communications sciences on the emergence and effects of disinformation trends. Beyond the usual rationales for regulating freedom of expression (e.g., by setting remedies for abuses) or media and communications structures (e.g., to assure access to plural information), I argue that the legitimate reasons to regulate in this context relate to (i) recent transformations in social communications

19 HABERMAS, Jürgen. O Estado Democrático de Democrático de Direito: uma amarração paradoxal de princípios contraditórios? In: *Era das Transições*. Rio de Janeiro: Editora Tempo Brasileiro, 2003.

20 JUNGHERR, Andreas; SCHROEDER, Ralph, Disinformation and the Structural Transformations of the Public Arena: Addressing the Actual Challenges to Democracy, *Social Media + Society*, v. 7, n. 1, 2021, p. 2.

and (ii) the "second-order effects"[21] that disinformation potentially has on democracy.

The policy debate often rests on assumptions that digital disinformation, employed in and outside of electoral processes, is harming democratic institutions, the public arena of debate, and the integrity of electoral processes in particular. In this sense, disinformation is often linked to political polarisation[22], elections disruption[23], and far-right communication strategies, to mention a few. Although it is relevant to investigate current socio-political phenomena, empirical evidence on these theories is still disputed[24], which weakens their potential to support policy proposals.

Sharing Andreas Jungherr and Ralph Schroeder's outlook, this paper approaches disinformation as "a symptom, and not a cause"[25] of structural tensions and transformations that have "impacted information flows and attention allocation"[26] in the public arena. This perspective carries far-reaching implications because if these transformations are not accounted for and consequently addressed by public policy, the dispute over truth or falsity by itself has little to no potential to remedy the structural challenges and harmful effects that accrue from the current digital communications landscape. These transformations are in great part connected to digitalisation and the emergence of digital platforms as information intermediaries. In an effort to name these transformations, the authors highlight how digital platforms are now "an integral part of the public arena as they provide complementary opportunities for distributing information and political messages in addition to those provided by news media and political organizations"[27]. These opportunities are shaped by the practices through which platforms exert influence over information fluxes (like content moderation and algorithm curation), which ultimately means that "how messages are disseminated on these platforms and their internal governance processes matter now beyond

21 KARPF, On Digital Disinformation and Democratric Myths.

22 BARBERÁ, Pablo, Social Media, Echo Chambers, and Political Polarization, *in*: PERSILY, Nathaniel; TUCKER, Joshua A. (Orgs.), *Social media and democracy: the state of the field, prospects for reform*, Cambridge: Cambridge University Press, 2020, p. 345.

23 FARIS *et al*, *Partisanship, Propaganda, and Disinformation: Online Media and the 2016 U.S. Presidential Election*.

24 For an overview of these disputes, see BARBERÁ, Social Media, Echo Chambers, and Political Polarization.

25 JUNGHERR; SCHROEDER, Disinformation and the Structural Transformations of the Public Arena, p. 2.

26 *Ibid*.

27 JUNGHERR; SCHROEDER, Disinformation and the Structural Transformations of the Public Arena, p. 4.

the narrow confines of their businesses"[28]. Connected to the emergence of digital intermediaries are other transformations relating to forms of message amplification, impacts on legacy media business models, and lack of transparency on how these intermediaries operate[29].

Although digitalised communications play an important role, the transformations in communication and socio-political practices to which disinformation refers are not connected exclusively to digitalisation. As proposed by Jeanette Hofmann, the relationship between technology and democracies is better approached by a "co-constitution" lense, according to which democracy and technology are connected "through a co-evolutionary process of mutual enabling"[30] rather than a causal link. Ultimately, this means not only that communication technologies and democracy shape each other but also that they are inserted in a "macro-level constellation of social change", which is affected by different socio-political factors. Building on Manuel Castells, Hofmann expressly approaches the current crisis of western democracies as an example of this multisided relationship, arguing that "the decay of conventional channels of political expression" cannot be single handily pinned on digitalisation since "core representative institutions began losing support and stability long before the internet advanced as a medium for 'mass self-communication'"[31]. It is worth highlighting that, for the policy debate, the adoption of this co-constitution perspective of democracy and technology does not imply that democracy is unaffected by digital technologies or even by disinformation. Rather, it means that addressing disinformation as a single negative externality overlooks other relevant factors and stifles the search for solutions. Yochai Benkler adopted a similar approach when writing about foreign interference in the 2016 American elections. Grounded in the realisation that "evidence of action is not evidence of influence", the author states that the eventual success of disinformation or propaganda strategies in the country must be interpreted "in the context of long-term patterns of loss of trust in institutions, including mainstream media, and the deep alienation of the past decade since the Great Recession"[32].

28 JUNGHERR; SCHROEDER, Disinformation and the Structural Transformations of the Public Arena, p. 4.

29 JUNGHERR; SCHROEDER, Disinformation and the Structural Transformations of the Public Arena, p. 5-8.

30 HOFMANN, Mediated democracy – Linking digital technology to political agency.

31 *Ibid*.

32 BENKLER, Yochai, Cautionary Notes on Disinformation and the Origins of Distrust, *MediaWell, Social Science Research Council*, 2019. Available at: https://mediawell.ssrc.org/expert-reflections/cautionary-notes-

These approaches provide two key takeaways for the policy debate. First, in terms of strategy, it means that, as highlighted before, countermeasures will come from multilevel actors, and state regulation is far from being a single effective solution. This coincides with much of the literature acknowledging that fighting disinformation requires different responses from different actors[33]. Whilst fact-checking and media literacy can be performed by journalistic or civil society organisations, and digital platforms can implement their own moderation and certification procedures, state regulation also plays an important role in the form of public policies implemented by legislation, which will be discussed in the next section. Second, in terms of regulatory legitimacy, motivations for state action against disinformation should not be founded on necessary causal links between disinformation and concrete results such as electoral outcomes or polarisation. Even though this link might exist, indicate that disinformation and such results are related is a non-structural way, rather then as causal pointers.

This is a crucial point because, as noted by David Karpf, "online disinformation and propaganda do not have to be particularly effective at duping voters or directly altering electoral outcomes in order to be fundamentally toxic to a well-functioning democracy"[34]. Assumptions that disinformation leads to polarization or that it affects electoral outcomes are not only disputed but also unnecessary because there are "second-order effects" that already undermine democratic institutions and the "governing norms that stand as a bulwark against elite corruption and abuse of power"[35]. The author's perception aligns with other theories that identify risks that disinformation poses to trust in democratic institutions[36] and even to fundamental rights. For instance, disinformation can "contribute to increased doubts in political and media institutions and [...] contribute to the destabilization of political systems"[37]. It "often targets institutions

on-disinformation-benkler/. Accessed on: 10 dec. 2019.

33 VALENTE, Jonas C. L., Regulando desinformação e fake news: um panorama internacional das respostas ao problema, *Comunicação pública*, n. Vol.14 nº 27, 2019.

34 KARPF, On Digital Disinformation and Democratic Myths. Available at: https://mediawell.ssrc.org/expert-reflections/on-digital-disinformation-and-democratic-myths/. Accessed on: 10 jan. 2020.

35 *Ibid*. Available at: https://mediawell.ssrc.org/expert-reflections/on-digital-disinformation-and-democratic-myths/. Accessed on: 10 jan. 2020.

36 OGNYANOVA, Katherine *et al*, Misinformation in action: Fake news exposureis linked to lower trust in media, higher trust in government when your side is in power, *Harvard Kennedy School Misinformation Review*, 2020.

37 JUNGHERR; SCHROEDER, Disinformation and the Structural Transformations of the Public Arena, p. 3.

and individuals in vulnerable situations and affects a wide range of human rights, including economic, social, cultural, civil and political rights, in which cases its effects surpass communications and rhetoric to translate into discrimination and hatred against minorities, immigrants and other marginalized communities"[38]. Disinformation also undermines "public confidence in mainstream media"[39] in different ways (such as by discrediting, impersonating, or accusing) and "the very existence of online misinformation resembling a journalistic product can diminish the credibility of legitimate news"[40].

This shows that appropriate regulatory rationales for disinformation cannot be narrowed down to a dispute over facts. Legitimate motivation for statutory regulation are those that address digital disinformation beyond an information quality perspective. This does not exclude speech-related counteractions from the broader debate – it is still important to dispute facts and meanings in the public sphere, this should be done outside of statutory and abstractly applicable regulation. Let aside the role of courts – who, restrained by their procedural and constitutional limitations, legitimately seek for the truth in an *ex post* case-by-case basis – there is not a legitimate locus for arbitration inside state coercion.

2.2 Regulatory strategies against disinformation

Statutory regulation initiatives against disinformation have continuously grown in different national contexts since 2017, notably after Brexit and the 2016 American election campaigns[41]. Since then, several

38 KHAN, Irene, *Disinformation and freedom of opinion and expression.*, [s.l.]: United Nations, General Assembly, 2021.

39 OGNYANOVA *et al*, Misinformation in action.

40 OGNYANOVA *et al*, Misinformation in action.

41 The role of states in disinformation counteraction is not restricted to formal legislation. It also includes other sorts of public policy beyond this paper's scope, like police task forces, institutional support, encouraging fact-checking and media literacy initiatives (MARSDEN, Chris; MEYER, Trisha; BROWN, Ian, Platform values and democratic elections: How can the law regulate digital disinformation?, *Computer Law & Security Review*, v. 36, p. 105373, 2020, p. 3.) and even enhancing cybersecurity. Further, as this paper looks exclusively at statutory legislation, it will not approach institutional solutions decentred from the state, such as the negotiation of voluntary measures, for example, the European Code of Conduct (See DURACH, Flavia; BÂRGĂOANU, Alina; NASTASIU, Cătălina, Tackling Disinformation: EU Regulation of the Digital Space, *Romanian Journal of European Affairs*, v. 20, n. 1, 2020) and the Australian Code of Practice on Disinformation and Misinformation (The Code was elaborated by digital platform providers represented by the Digital Industry Group Inc. (DIGI) upon a recommendation of the Australian Media and Communications Authority (ACMA). The Code is available at: https://digi.org.au/disinformation-code/. Accessed on: 10 jun. 2021. For more on the Australian framework, see CARSON, Andrea; FALLON, Liam, *Fighting Fake News: A study of online misinformation regulation in the Asia Pacific.*, [s.l.]: La Trobe University, 2021).

countries have enacted regulations aiming to combat disinformation in and outside of electoral processes, a trend enhanced in volume and justification by the Covid-19 pandemic (which brought the spread of public health-related disinformation to light)[42]. Previous works have accounted for these regulatory strategies against disinformation and classified them based on different criteria, such as the institutional arrangement[43].

In this paper, I propose that we separate statutory regulation (approved or under discussion in parliament) into three groups, according to their regulatory target: those aiming at individual expression, i.e., the content of the message itself; those aiming at the collection, handling, and use of personal data for disinformation ends; and those that implement structural regulation of digital intermediaries. These three categories are the result of research on global regulatory strategies against disinformation collected from pre-existing policy repositories that describe these initiatives in English[44].

It should be highlighted that these three categories of regulatory targets are not static. A regulatory target is understood here as the object to which regulation intends to conform[45], not necessarily the subject affected by the regulation (even though this element is still important). So, for instance, while policies that regulate data or structure have the potential to bind corporate bodies almost exclusively, content regulation refers to individual or corporate behaviour (as will be clarified in the following section).

42 WISEMAN, Jamie, Rush to pass 'fake news' laws during Covid-19 intensifying global media freedom challenges, *International Press Institute*, 2020. According to Irene Khan, at least "17 states adopted legislation to address pandemic-related problematic disinformation". KHAN, *Disinformation and freedom of opinion and expression.*, p. 11.

43 VALENTE, Regulando desinformação e fake news.

44 The repositories that were initially consulted are the Poynter Institute's "A guide to anti-misinformation actions around the world". Available at: https://www.poynter.org/ifcn/anti-misinformation-actions/. Accessed on: 10 jan. 2020); the Law Library of Congress Reports "Initiatives to Counter Fake News in Selected Countries". Available at: https://digitalcommons.unl.edu/scholcom/179/. Accessed on: 06 jun. 2021; and "Government Responses to Disinformation on Social Media Platforms". Available at: https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1180&context=scholcom. Accessed on: 06 jun. 2021); and CARSON; FALLON, *Fighting Fake News: A study of online misinformation regulation in the Asia Pacific*. A Study of Online Misinformation Regulation in the Asia Pacific, La Trobe University. Available at: https://www.latrobe.edu.au/__data/assets/pdf_file/0019/1203553/carson-fake-news.pdf. Accessed on: 06 jun. 2021). Besides these repositories, further academic literature and media reports were used to provide insights and context for regulatory experiences. All these other sources are cited throughout the paper.

45 This is similar to the explanation used by Lyria Bennet-Moses in MOSES, Lyria Bennett, How to Think about Law, Regulation and Technology: Problems with 'Technology' as a Regulatory Target, *Law, Innovation and Technology*, v. 5, n. 1, p. 1-20, 2013. The literature accounts for different ways of referencing regulatory targets, which can also be understood as the "the individual or organization to which a regulatory instrument applies" (COGLIANESE, Cary, Engaging Business in the Regulation of Nanotechnology, *in*: BOSSO, Christopher J. (Org.), *Governing uncertainty: environmental regulation in the age of nanotechnology*, Washington, DC: RFF Press, 2010.)

Nevertheless, the element that the policies aim to influence – the element on which the rationale rests – is content.

Also, as I will show, a single strategy can affect more than one target. Additionally, these three groups do not intend to consider all possible targets or regulatory strategies related to disinformation. Since disinformation is such an intricate phenomenon, the idea behind this taxonomy is to better understand policy options currently on the table, their rationales, and implications for fundamental rights. In the next subsections, I will describe each one of these groups of policies and their recent implementation or consideration in different experiences.

2.2.1 Content

Policies aimed at content deem disinformation an illegal type of speech and thus something that must be banned from circulation. This includes a variety of mechanisms with the potential to restrict freedom of expression to different degrees.

Some of these policies are directed at deterring individuals from producing, publishing, distributing, or spreading disinformation in any way by imposing criminal or civil liability. This can be done through a diverse set of commands, like, for instance, the creation of new criminal provisions. In this sense, Brazilian Law 13.834/2019 criminalises accusing someone of a crime or infraction of which they are innocent with electoral purposes, thus prompting administrative and criminal investigations[46]. Other examples have gained international attention, such as the case of Ethiopia, where Proclamation 1185/2020 provides for imprisonment sanctions against individuals who disseminate disinformation or hate speech by means of broadcasting, print, or social media[47]. Similar provisions were also approved, for instance, in Malaysia[48], Cambodia[49] and Kenya[50]. Besides the creation of

46 Law 13.834/2019, art. 2o. Available at: http://www.planalto.gov.br/ccivil_03/_ato2019-2022/2019/lei/L13834.htm. Accessed on: 30 sep. 2021.

47 Proclamation 1185/2020, as per the translation available at: https://chilot.me/2020/04/05/proclamation-no-1185-2020-hate-speech-and-disinformation-prevention-and-suppression/. Accessed on: 13 sep. 2021.

48 SCHULDT, Lasse, The rebirth of Malaysia's fake news law – and what the NetzDG has to do with it, *Verfassungsblog.* Available at: https://verfassungsblog.de/malaysia-fake-news/. Accessed on: 13 sep. 2021.

49 Cambodian Center for Human Rights, *Submission to the Special Rapporteur on the promotion and protection of the rights to freedom of opinion and expression.* Available at: Cambodia-Centre-for-human-rights.pdf (ohchr.org). Accessed on: 13 sep. 2021.

50 SUGOW, Abdulmalik, MUNGAI, Beatrice, WANYAMA, Jentrix, "The regulation of fake news in Kenya under the coronavirus threat", available at: https://cipit.strathmore.edu/the-regulation-of-fake-news-in-kenya-under-the-coronavirus-threat/. Accessed on: 06 jun. 2021.

new types of crimes, the enhancement of pre-existing penalties for speech abuses is also possible. E.g. in Denmark[51] a 2019 amendment to the Criminal Act added activities that affect public opinion among the unlawful speech abuses by foreign governments as a clear way to avoid foreign disruption in elections[52].

Policies aimed at content can be directed not only at individuals but also at information intermediaries. These policies include legislations that impose duties of removal of disinformation content on digital platforms, which have the potential to disproportionally curtail protected speech, as they force intermediaries "to make highly context-sensitive decisions within tight time frames and based on insufficient available information"[53]. This is the case for China, where article 47 of the Cybersecurity Law requires platforms to take appropriate measures in the face of disinformation[54]. Regulation can also target media services and digital platforms, as does French Law 2018--2012, which "enables the transmission of foreign state-controlled radio and television services that broadcast disinformation to be curtailed, or temporarily suspended, prior to elections"[55]. In the face of disinformation, an interested party can apply to a judge for an expedited order requiring that providers of online communication services take necessary steps to prevent continuing diffusion of false information[56]. The interruption of communication services as a remedy for disinformation can also take more authoritarian forms. In Belarus, media legislation was amended in 2018 to allow the Ministry of Information to block social media platforms and hold website owners liable for hosting content deemed false, defamatory, or harmful to the national

51 Law Library of Congress Reports, "Initiatives to Counter Fake News in Selected Countries". Available at: https://digitalcommons.unl.edu/scholcom/179/. Accessed on: 06 jun. 2021.

52 https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1180&context=scholcom

53 SCHULZ, Wolfgang, *Roles and Responsibilities of Information Intermediaries: Fighting Misinformation as a Test Case for Human-Rights Respecting Governance of Social Media Platforms*, [s.l.]: Hoover Institution, Stanford University, 2019, p. 17.

54 QI, Aimin; SHAO, Guosong; ZHENG, Wentong, Assessing China's Cybersecurity Law, *Computer Law & Security Review*, v. 34, n. 6, p. 1342-1354, 2018, p. 12.

55 CRAUFURD SMITH, Rachael, Fake news, French Law and democratic legitimacy: lessons for the United Kingdom?, *Journal of Media Law*, v. 11, n. 1, p. 52-81, 2019, p. 52. As the author highlights, the French Law adopts "a more holistic approach" (p. 58) based on three strands designed to curb foreign state disinformation: to prevent further online transmission of false information prior to elections (i.e., the case of the policy described here); to ensure greater transparency in the operation of online communication platforms; and to stimulate new educational initiatives. Some of these strategies do not encompass content regulation and will be discussed in other sections of this paper.

56 *Ibid*., p. 60.

interest (without warning or judicial oversight)[57]. In this case, the provision was accompanied by other internet and media targeted policies that have so far resulted in arrests of journalists and persecution of political dissent[58]. Similarly, in 2019, the Cambodian government announced it would revoke licenses "of print and online media outlets distributing 'fake news' deemed to be a danger to national security"[59], and even though the measure did not come into force, local NGOs still report the use of cybercrime and other legislations under the guise of disinformation combat to crack down on political dissent[60].

For the sake of nuance, it is important to highlight that these examples encompass policies with very different impacts on speech. For instance, subjecting proponents of a certain type of individual expression to incarceration jeopardises individual liberties more severely than generating incentives for intermediaries to over-restrict access to content would. Even though the latter still represents a threat of troubling effects that should not be taken lightly[61], in the first case, risks to freedom of expression is worsened by the threat of incarceration, which also jeopardises one's right to liberty and possibly physical and mental integrity. Similarly, employing judicial oversight over a possible disinformation countermeasure (like in France) is within the design of courts in liberal democracies[62] and does not threaten freedom of expression to the same extent as arbitrary executive branch decisions. National context is also a relevant factor. The examples of Belarus and Cambodia show us that when regulation against disinformation is part of a broader authoritarian legal framework, it is mixed with subterfuge for

57 Belarus, *Freedom House*. Available at: https://freedomhouse.org/country/belarus/freedom-net/2021. Accessed on: 13 sep. 2021.

58 Belarus, *Freedom House*. Available at: https://freedomhouse.org/country/belarus/freedom-net/2021. Accessed on: 13 sep. 2021.

59 Cambodian Center for Human Rights, *Submission to the Special Rapporteur on the promotion and protection of the rights to freedom of opinion and expression*. Available at: https://www.ohchr.org/Documents/Issues/Expression/disinformation/2-Civil-society-organisations/Cambodia-Centre-for-human-rights.pdf. Accessed on: 13 sep. 2021.

60 Cambodian Center for Human Rights, *Submission to the Special Rapporteur on the promotion and protection of the rights to freedom of opinion and expression*. Available at: https://www.ohchr.org/Documents/Issues/Expression/disinformation/2-Civil-society-organisations/Cambodia-Centre-for-human-rights.pdf. Accessed on: 13 sep. 2021.

61 KHAN, Irene, *A/HRC/47/25 – Disinformation and freedom of opinion and expression*, Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, p. 3. Available at: https://www.ohchr.org/EN/Issues/FreedomOpinion/Pages/Report-on-disinformation.aspx. Accessed on: 01 jun. 2021.

62 IGLESIAS KELLER, Clara, Policy by judicialisation: the institutional framework for intermediary liability in Brazil, *International Review of Law, Computers & Technology*, p. 1-19, 2020.

persecution political opponents – as was reported to have happened in Egypt, China, and the United Arab Emirates as well[63].

Nevertheless, provisions that establish countermeasures for disinformation based on content essentially rest on an understanding of truth or falsity that can either be pre-established in legislation or left to the discretion of the body responsible for enforcing them. This will ultimately lead to imposing one version of facts – whether it be judges', executive authorities', or platforms' – over others. Even though it is to varying degrees, strategies based on a concept of disinformation will inevitably steer the dispute over truth and fact away from where it belongs – in society and public debate.

Provisions for abuses of freedom of expression are entailed in liberal democracies' constitutional systems, which presume the possibility of restricting fundamental rights when it conflicts with other guarantees. Terrorist content and hate speech, for instance, are prohibited because they notably impose severe risks to third-party rights. But disinformation rests in a greyer area. Other than the cases where it is paired with other sorts of online harms, like hate speech, defamation, or harassment (and can ultimately lead to concrete violence against vulnerable groups[64]), disinformation should not be necessarily illegal, as a great part of its forms can be considered legitimate expression. Depending on the conceptual framework, many different conducts can qualify as such, from information that is displaced from its original context to parodies and completely made-up facts – all of which could still be placed in the realm of legitimate speech, depending on the circumstances of each case. In fact, one could argue that even the regular adjudication of plain illegal speech, which is usually left up to courts, will inevitably rely on the judge's interpretation and perspective of the facts under consideration[65]. This inherent risk in arbitrating speech is severely higher if this task is handed to administrative authorities without judicial oversight, given their unfettered discretion, opening up the possibility for abuse and arbitrary decision-making[66].Finally, content-targeted policies do

63 Law Library of Congress Reports, "Initiatives to Counter Fake News in Selected Countries". Available at: https://digitalcommons.unl.edu/scholcom/179/. Accessed on: 06 jun. 2021.

64 SCHULZ, Wolfgang, *Roles and Responsibilities of Information Intermediaries: Fighting Misinformation as a Test Case for Human-Rights Respecting Governance of Social Media Platforms*, [s.l.]: Hoover Institution, Stanford University, 2019, p. 17

65 MACEDO JUNIOR, Ronaldo Porto, Freedom of Expression: what lessons should we learn from US experience?, *Revista Direito GV*, v. 13, n. 1, p. 274-302, 2017, p. 2.

66 KHAN, *Disinformation and freedom of opinion and expression.,* p. 11.

not promise efficiency, especially when they are applied to individuals. They overlook how orchestrated and structured disinformation campaigns operate, particularly when what differentiates their reach potential is not individual conduct but the characteristics of digital communications that they rest on – for instance, the possibility of using personal data for political microtargeting. In fact, digital platforms' influence over information and attention fluxes rests in large part on the use of personal data, which also makes it a relevant target for regulatory policies.

2.2.2 Data

Policies that target *data* are represented by legal frameworks that regulate the collection, treatment, and storage of personal data for different purposes. These regulations approach data protection across sectors to protect rights holders in increasingly digitalised economies. This strain is represented by the enactment or updating of laws dedicated to data protection, most often based on "the guarantee of a fundamental right and the realization of this right by means of a legal regime of data protection, in the form of a general law on the subject"[67].

In digital disinformation debates, the use of data is specifically related to the use of political microtargeting techniques that are meant to distribute content to a segmented audience[68]. Microtargeting is "a form of online targeted advertising that analyses personal data to identify the interests of a specific audience or individual in order to influence their actions"[69]. This logic is embedded, for instance, in the core of social media platforms' business models, which target users with advertisements tailored to their preferences. Similarly, political microtargeting "involves collecting and analysing people's personal data to send them tailored political messages"[70] and thus introduces information "while targeting promising individuals or groups specifically and out of sight of the public arena"[71]. The main fuel for

67 MENDES, Laura Schertel, *Privacidade, proteção de dados e direito do consumidor: linhas gerais de um novo direito fundamental.,* São Paulo: Saraiva, 2014, p. 47.

68 WALKER, Shawn; MERCEA, Dan; BASTOS, Marco, The disinformation landscape and the lockdown of social platforms, *Information, Communication & Society*, v. 22, n. 11, p. 1531-1543, 2019.

69 *UK Information Commissioner's Office*, Microtargeting, ICO website. Available at: https://ico.org.uk/your-data-matters/be-data-aware/social-media-privacy-settings/microtargeting/. Accessed on: 10 sep 2021.

70 ZAROUALI, Brahim *et al*, Using a Personality-Profiling Algorithm to Investigate Political Microtargeting: Assessing the Persuasion Effects of Personality-Tailored Ads on Social Media, *Communication Research*, p. 009365022096196, 2020.

71 JUNGHERR; SCHROEDER, Disinformation and the Structural Transformations of the Public Arena, p. 3.

political microtargeting is "data gathered from citizens' online presentation and behaviour, including from their social media use"[72]. This means that the kind of data that informs political microtargeting is likely to include ethnicity, ideologies, and political and religious beliefs, among other types of information concerning which users can be further discriminated. Differentiating political microtargeting from mere microtargeting for policy purposes entails a complicated exercise of interpretation – what is political, after all? – with potential to impose stricter rules based on content.

Even though microtargeted disinformation can be used in different contexts, it is in the realm of electoral legislation that data appears as a target for regulation against disinformation, as "online disinformation and unlawful political microtargeting represent a threat to elections around the globe"[73]. Data-based political advertisement has been at the centre of disinformation strategies in different national contexts[74], raising concerns beyond the protection of citizens' privacy and data protection[75]. Electoral frameworks can vary across countries, but they are commonly built on the assumption that the electoral period requires qualified protection for speech, access to information, and opinion formation guarantees[76]. This presumes that all information will circulate to everyone. In this sense, political microtargeting has a doubled potential to interfere with elections, both when it harmfully manipulates information being distributed and when it restricts plurality by excluding people who are not targeted from the debate[77]. Distribution of disinformation in electoral contexts has been associated with "harming the political debate, excluding populations from it, and even making individual autonomy vulnerable in invisible and unexpected ways" [78].

72 NENADIĆ, Iva, Unpacking the "European approach" to tackling challenges of disinformation and political manipulation, *Internet Policy Review*, v. 8, n. 4, 2019, p. 6.

73 *Ibid*., p. 2.

74 CADWALLADR, The great British Brexit robbery: how our democracy was hijacked; EVANGELISTA; BRUNO, WhatsApp and political instability in Brazil; DOBBER, Tom; Ó FATHAIGH, Ronan; ZUIDERVEEN BORGESIUS, Frederik J., The regulation of online political micro-targeting in Europe, *Internet Policy Review*, v. 8, n. 4, 2019.

75 However, guaranteeing the integrity of data-driven elections encompasses concerns that go beyond disinformation. See BENNET, Colin J.; ODURO-MARFO, Smith, *Privacy, Voter Surveillance and Democratic Engagement: Challenges for Data Protection Authorities.,* [s.l.]: University of Victoria, 2019. BENNETT, Colin J.; LYON, David, Data-driven elections: implications and challenges for democratic societies, *Internet Policy Review*, v. 8, n. 4, 2019.

76 CRUZ, Francisco Brito, *Novo jogo, velhas regras: democracia e direito na era da nova propaganda politica e das fake news*, Belo Horizonte, MG: Grupo Editorial Letramento, Casa do Direito, 2020, p. 297.

77 CRUZ, *Novo jogo, velhas regras.*

78 *Ibid*., p. 377.

Thus, there is a case for "strengthening enforcement of data protection legislation in electoral contexts"[79]. In general, data protection legislation already serves different aspects that regulate the use of political microtargeting in electoral processes. As noted by Francisco Brito Cruz, they can limit data collection (e.g., by prohibiting unlawful surveillance or commercialisation of voters databases), data sharing (by prohibiting international data transfer in specific purposes, possibly electoral), and data management, as in cases where there is a deviation from the purposes authorised by the agent. In several countries, these constraints have been carried out by personal data protection rules and by authorities that enforce these rules[80]. Specific mechanisms may include restrictions on data gathering and accumulation for political microtargeting purposes – like in Japan, where "the capture of personal data on the electorate, and the communication of personalised political messaging" is understood to be "largely prohibited" by the current legislation[81].

Depending on how they are carved in legislation, limitation or prohibition of use of data for "political purposes" are likely to lead to another intricate exercise of interpretation regarding what is a political purpose or not. The electoral time frame can provide a more stable criterion – however, in a highly digitalised public sphere, electoral campaigning is submitted to transformations of its own, which means that the relevance of such time frames as the key moment for political communication is diminished. Political content standards based on characteristics of a certain kind of message will probably be fully subjected to the perception of their enforcer, and therefore provide less legal certainty.

Either via general or disinformation specific regulation, data is a relevant target for public policies because of its role in feeding the dynamics of information distribution in digital platforms. This means that aiming at data would in theory promise better results. Along with how and what for these companies use data, there a series of other aspects of their business models that are a relevant for disinformation regulation, as I will show in the next section.

79 NENADIĆ, Unpacking the "European approach" to tackling challenges of disinformation and political manipulation, p. 13; CRUZ, *Novo jogo, velhas regras,* p. 376-378.

80 CRUZ, *Novo jogo, velhas regras,* p. 377.

81 BENNET; ODURO-MARFO, *Privacy, Voter Surveillance and Democratic Engagement: Challenges for Data Protection Authorities.*, p. 6.

2.2.3 Structure

As per the third and last group of measures, policies focusing on structure regulate the digital platforms serving as one of the means to distribute disinformation[82]. These initiatives target the business models of such actors, with a particular focus on technological tools, practices, and the criteria on which they operate.

By "policies aimed at structure", I mean the mechanisms meant to imbue digital platform's business models with "a new ethics of responsible platforms, which can provide certainty, fairness and accountability of enforcement of speech rules, but ensure that speech control is strictly autonomous from the state"[83]. Overall, they implement "incentives for the platforms to modify their operations" through different means, including "the introduction of government mandated responsibilities; data and privacy protection measures; the use of codes of practice; and measures to strengthen skills and training policies"[84]. Regulations focused on structure do not hang on disinformation specifically but rather address aspects of platforms' business models that exert influence on information and attention fluxes of different natures, thus impacting the distribution of a diversity of online harms. Even so, mis- and disinformation are usually perceived[85], or even expressly listed[86], as regulatory motivations of structure regulation proposals.

Some part of European literature distinguishes a regulatory trend towards structural regulation through expressions such as "a shift from liability to responsibility"[87] or "from liability to duty"[88]. These terms attempt to distinguish regulatory initiatives in recent years that entail "the need for

82 In the context of digital platform regulation, regulating structure can also refer to antitrust legislation (see *Tackling the Information Crisis: A Policy Framework for Media System* Resilience, The Report of the LSE Commission on Truth, Trust and Technology. Available at: https://www.lse.ac.uk/media-and-communications/truth-trust-and-technology-commission. Accessed on: 10 aug. 2021). Despite this application of the term, this article's scope does not go so far as to encompass the analysis of antitrust legislation.

83 TAMBINI, Damian, Rights and Responsibilities of Internet Intermediaries in Europe: The Need for Policy Coordination.

84 MANSELL, Robin; STEINMUELLER, W. Edward, *Advanced introduction to platform economics,* Cheltenham, UK; Northampton, MA: Edward Elgar Publishing, 2020, p. 101.

85 SCHULZ, WOLFGANG, *Roles and Responsibilities of Information Intermediaries: Fighting Misinformation as a Test Case for Human-Rights Respecting Governance of Social Media Platforms.*

86 Brazilian Federal House of Representatives Bill of Law 2.630/2020. Available at: https://www.camara.leg.br/proposicoesWeb/fichadetramitacao?idProposicao=2256735. Accessed on: 10 jun. 2021.

87 FROSIO, Giancarlo, Why keep a dog and bark yourself? From intermediary liability to responsibility, *International Journal of Law and Information Technology*, p. 1-33, 2017.

88 MAC SÍTHIGH, Daithí, The road to responsibilities: new attitudes towards Internet intermediaries, *Information & Communications Technology Law*, v. 29, n. 1, p. 1-21, 2020.

proactive measures"[89] from dominant liability-centred models, which still characterise online content regulation frameworks by carving out conditions for the civil liability of digital platforms over user-generated content[90]. While different authors will include different types of regulatory strategies under what is understood here as the responsibility of digitals platforms[91], there are a few common suspects – like duties of notification and due process in content moderation, obligations of setting user-centred flagging tools[92], and the very popular obligations of transparency.

In fact, the latter is a key mechanism of structure regulation. It seeks to bring more clarity to digital platforms' operations, like the criteria for and effects of content moderation decisions[93] and "the black box of algorithm decision-making"[94]. Notwithstanding critiques of its ambiguity and flexibility[95], or even of the use of transparency as a "policy panacea"[96], recent regulatory approaches provide, for instance, the delivery of transparency reports on content moderation decisions – a trend that came to light with the German law *Netzwerkdurchsetzungsgesetz* – NetzDG[97] and is now part of policy proposals under discussion in Brazil[98] and in Europe (where transparency is one of the pillars of the Digital Services Act proposals[99]). Requirements

89 KUCZERAWY, Aleksandra. *General Monitoring Obligations: A New Cornerstone of Internet Regulation in the EU?*. Available at: https://ssrn.com/abstract=3449170. Accessed on: 06 mar. 2020.

90 GASSER, Urs; SCHULZ, Wolfgang, Governance of Online Intermediaries: Observations from a Series of National Case Studies, *SSRN Electronic Journal,* 2015.

91 See, for instance, Tambini, 2017 and Frosio, 2017.

92 CRAUFURD SMITH, Fake news, French Law and democratic legitimacy, p. 62.

93 SUZOR, Nicolas P. *et al*, What Do We Mean When We Talk About Transparency? Toward Meaningful Transparency in Commercial Content Moderation, *International Journal of Communication,* v. 13, p. 1526-1543, 2019.

94 RIEDER, Bernhard; HOFMANN, Jeanette, Towards platform observability, *Internet Policy Review,* v. 9, n. 4, 2020.

95 GORWA, Robert; ASH, Timonthy Garton, Democratic Transparency in the Platform Society, *in*: PERSILY, Nathaniel; TUCKER, Joshua A. (Orgs.), *Social media and democracy: the state of the field, prospects for reform,* Cambridge: Cambridge University Press, 2020, p. 287.

96 RIEDER; HOFMANN, Towards platform observability.

97 The 2017 *Netzwerkdurchsetzungsgesetz* – NetzDG Act is not a disinformation-targeted law. It requires social media platforms to implement procedures that allow users to report illegal content, notably, the 22 criminal conducts already provided in Germany's Criminal Code. According to its terms, "'manifestly unlawful"' content needs to be removed within 24 hours of notification (or possibly after seven days or more, with terms to be agreed upon with law enforcement authority). Beside removals, Section 2 requires platforms to periodically publish transparency reports on the number of complaints received and how they were handled by the platform. HELDT, Amélie, Reading between the lines and the numbers: an analysis of the first NetzDG reports, *Internet Policy Review*, v. 8, n. 2, 2019.

98 Brazilian Federal House of Representatives Bill of Law 2.630/2020. Available at: https://www.camara.leg.br/proposicoesWeb/fichadetramitacao?idProposicao=2256735. Accessed on: 10 jun. 2021.

99 European Commission, Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC. Available at: https://

of transparency for advertisers and especially of political campaigns is also a trend reflected, for example, in the 2021 European Commission public consultation on improving transparency in political advertisement online and offline[100], in the Brazilian proposals mentioned before, and French Law 2018-1202 (according to which platforms must publish details of the measures taken and report annually on progress in these areas[101]).

Also, in the realm of structure there are a diversity of measures that can generate different types of incentives and impact platforms and users' rights to different degrees. Obligations of content removal, for instance, require attention for the incentives they generate[102]. There is a mix of regulatory targets in these cases, as monitoring obligations are aimed at platforms, but obligations of removal are very much centred on content. Here, the lines between content and structure regulations are blurry, and warnings of risks to freedom of expression are due. Obligations of content removal[103] are known to delegate the job of arbitrating the scope of freedom of expression to private platforms, raising immediate concerns about incentives for over-blocking[104].

As a relatively recent trend, structural regulation does not yet rest on empirical results. In turn, its legitimacy can be justified in different ways, among which are social media platforms' "unique public role" in communications[105]; their "systemic opinion power [...] to create dependences and influence other players in a democracy"[106]; or the fact that such companies' decide on collective behaviour behind closed doors and therefore need to be imbued with "procedural values", such as "the rule of law, due process and transparency", as well as "participation in

digital-strategy.ec.europa.eu/en/library/proposal-regulation-european-parliament-and-council-single-market-digital-services-digital-services. Accessed on: 10 jun. 2021.

100 https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12826-Transparency-of-political-advertising/public-consultation_en.

101 CRAUFURD SMITH, Fake news, French Law and democratic legitimacy, p. 62.

102 NENADIĆ, Unpacking the "European approach" to tackling challenges of disinformation and political manipulation.

103 An example would be the German NetzDG mentioned above.

104 SCHULZ, WOLFGANG, *Roles and Responsibilities of Information Intermediaries: Fighting Misinformation as a Test Case for Human-Rights Respecting Governance of Social Media Platforms*.

105 SYLVAIN, Oliver, Internet governance and democratic legitimacy, *Federal Communications Law Journal*, v. 62, n. 2, p. 205-274, 2010, p. 209.

106 HELBERGER, Natali, The Political Power of Platforms: How Current Attempts to Regulate Misinformation Amplify Opinion Power, *Digital Journalism*, v. 8, n. 6, p. 842-854, 2020, p. 846.

decision making"[107]. For aiming at defining aspects of platforms' business practices – including the ones that allow them to exert influence over online communications – structural regulation bears relevant potential to countermeasure disinformation and other sorts of online harms. However, their potential to actually shift power imbalances does not meet the same enthusiasm.

3 ANALYSIS

The literature analysed in section 2.1 suggests that disinformation as a socio-political phenomenon cannot be approached as the single cause of democratic crisis or presumed shifts in opinion formation processes – either due to the lack of consensus on empirical evidence, or simply because this approach overlooks its complexity. As proposed by the concept of mediated democracy, a series of different co-existing conditions enable "possibilities of political action without determining them"[108], from which I inferred that understanding recent forms of disinformation and its effects goes beyond looking at how technologies impact democratic institutions. Among the different conditions that deserve attention, this paper looks at what is needed from statutory regulation, notably in the realm of digital communications policies.

Out of the three groups of regulatory strategies herein described, the ones aimed at content pose the most severe risks to freedom of expression and even to civil liberties. This is because strategies aimed at banning content (i) will inevitably rest on a certain understanding of truth, and when enforced, (ii) will privilege the understanding of the enforcer over others, removing the dispute over facts and perspectives from where it originally belongs – in societal debate. Further, strategies aimed at content show the least promise of addressing the transformations in communication practices introduced by digitalisation, which sets the conditions for the spread of disinformation and other sorts of harms in a wide extent. Therefore, statutory regulation aiming at providing countermeasures for disinformation should not be aimed at content. This does not necessarily mean that content is not a target for statutory regulation under any circumstances. Liberal democracies

107 SUZOR, Nicolas; VAN GEELEN, Tess; MYERS WEST, Sarah, Evaluating the legitimacy of platform governance: A review of research and a shared research agenda, *International Communication Gazette*, v. 80, n. 4, p. 385-400, 2018, p. 391-392.

108 HOFMANN, Mediated democracy – Linking digital technology to political agency.

that are supported, among other pillars, by a right to freedom of expression must also provide and prepare for hypotheses of abuses to this right and its collision with other fundamental guarantees. Freedom of expression is usually legitimately restricted, for instance, to avoid the circulation of illegal harmful speech, but this is not necessarily the case of disinformation.

Considering the policy options approached in this paper, free speech will be restricted the least if regulation aims at data and structure, which are policies aimed at curbing disinformation or minimising its impacts, instead of plainly removing or criminalising disputed facts. Still, this does not mean these policies represent no risks at all. For instance, regulating data usage affects information circulation, and specifically in electoral contexts, it can end up limiting the circulation of legitimate communication from political parties. Furthermore, and as mentioned previously, restrictions of data usage pinned on "political content" or "political purposes" will also ultimately entail subjective interpretations. Nonetheless, data remains a relevant target for disinformation regulation. Both general data protection regulations, provisions aiming at elections or even at political purposes, will present proportionally less risks to speech guarantees and fundamental liberties. Setting aside the importance of general data protection regulation in digitalised societies, they do implement minimum safeguards to different stages of data collection and treatment, binding digital platforms and other actors that engage in illegal uses of data for disinformation purposes. In fact, even provisions that restrict the use for political purposes hold at least rhetorical relevance, as they are part of a bigger societal conversation about what sort of economic, political or ideological interests microtargeting technologies should support.

Similarly, depending on the kind of incentive generated by structure regulation, risks for speech will still be relevant. Incentives for removal of content by platforms deserve double attention, as they might overlap with targeting content. The shift towards structural responsibility for digital platforms is welcome, at least as a new attempt to steer digital platforms' operations towards the public interest. Nevertheless, these policies still function very much inside the power structures in place instead of actually challenging them, and their results are yet to be seen. In fact, one could even argue that they grant even more power to digital platforms, since they

recognize the legitimacy of their current operations and enable internal decision making on compliance standards[109].

Aiming at structure and data is promising because it aligns with the legitimate rationales for disinformation regulation described in section 2.1. The use of data by and general practices of digital platforms are relevant factors that contribute to the transformations of the public sphere[110], allowing for the spread of the phenomena included in the current information disorder, as well as for a series of online harms. Influence in information and attention fluxes, for instance, are ultimately promoted by the technological tools that generate engagement by determining content distribution, allowing for digital platforms' advertising targeting business models to thrive. Moreover, criteria upon which these fluxes operate remains opaque, and thus above public scrutiny. The impacts of these transformations go beyond the experience of online communications and affect legacy media and governments, which increasingly rely on platforms to communicate with their constituencies[111].

A grain of salt is due, as a reminder that regulation of data and structure are not complete responses to the scenarios of disinformation and distrust in democratic institutions. Still, they allow us to address some of these transformations without creating ineffective and illegitimate policies with the potential to become weapons of mass repression. On the other hand, their early age means that empirical data on the success of these regulatory strategies is still missing.

Lastly, while content should stay out of statutory regulation, it remains very much at the centre of countermeasures against disinformation performed by non-state actors. Independent fact-checking, media literacy, and professional journalism should take the lead in assuring dissent and democratic dispute over facts and truth.

CONCLUSIONS

As regulatory efforts towards disinformation continuously increase, this paper offered an analysis of a set of policies under discussion and implementation in different legal backgrounds. Its main finding is that current

109 Natali Helberger has argued that recent attempts to "infuse some public value standards into corporations" formalises "the role of platforms as governors of online speech" and reinforcing their political power. HELBERGER, The Political Power of Platforms, p. 848.

110 JUNGHERR; SCHROEDER, Disinformation and the Structural Transformations of the Public Arena.

111 HELBERGER, The Political Power of Platforms, p. 847.

disinformation practices are fuelling the proposal of policies over restrictive of speech in several experiences, without any promise of efficiency. In fact, there is relevant indication that these policies can be used as proxies to regulate undesired speech and crack down on political dissent. Privileging regulatory strategies that aim at data and structure rather than content is not only important to preserve freedom of expression and societal debate as pillars of democracy. These regulatory strategies also bear greater potential to address the transformations promoted by digital communication practices. It should be highlighted, however, that this is a small piece of the puzzle, as the spread and effects of disinformation in current scenarios is shaped by various conditions of different nature, that go way beyond the ways through which states regulate digital communications.

BIBLIOGRAPHY

BALDWIN, Robert; CAVE, Martin; LODGE, Martin. *Understanding regulation: theory, strategy, and practice*. 2nd ed. New York: Oxford University Press, 2012.

BARBERÁ, Pablo. Social Media, Echo Chambers, and Political Polarization. *In*: PERSILY, Nathaniel; TUCKER, Joshua A. (Orgs.). *Social media and democracy: the state of the field, prospects for reform*. Cambridge New York Port Melbourne New Delhi Singapore: Cambridge University Press, 2020, p. 345.

BENKLER, Yochai. Cautionary Notes on Disinformation and the Origins of Distrust. *MediaWell, Social Science Research Council,* 2019. Disponível em: https://mediawell.ssrc.org/expert-reflections/cautionary-notes-on-disinformation-benkler/. Acesso em: 10 dez. 2019.

BENNET, Colin J.; ODURO-MARFO, Smith. *Privacy, Voter Surveillance and Democratic Engagement: Challenges for Data Protection Authorities.* [s.l.]: University of Victoria, 2019.

BENNETT, Colin J.; LYON, David. Data-driven elections: implications and challenges for democratic societies. *Internet Policy Review*, v. 8, n. 4, 2019. Disponível em: https://policyreview.info/node/1433. Acesso em: 10 out. 2021.

BLACK, J. Decentring Regulation: Understanding the Role of Regulation and Self-Regulation in a "Post-Regulatory" World. *Current Legal Problems*, v. 54, n. 1, p. 103-146, 2001.

CADWALLADR, Carole. The great British Brexit robbery: how our democracy was hijacked. *The Guardian*, 2017. Disponível em: https://www.theguardian.com/technology/2017/may/07/the-great-british-brexit-robbery-hijacked-democracy. Acesso em: 10 maio 2020.

CARSON, Andrea; FALLON, Liam. *Fighting Fake News: A study of online misinformation regulation in the Asia Pacific.* [s.l.]: La Trobe University, 2021.

Disponível em: https://www.latrobe.edu.au/__data/assets/pdf_file/0019/1203553/carson-fake-news.pdf. Acesso em: 10 jun. 221DC.

COGLIANESE, Cary. Engaging Business in the Regulation of Nanotechnology. *In*: BOSSO, Christopher J. (Org.). *Governing uncertainty: environmental regulation in the age of nanotechnology*. Washington, DC: RFF Press, 2010.

CRAUFURD SMITH, Rachael. Fake news, French Law and democratic legitimacy: lessons for the United Kingdom? *Journal of Media Law*, v. 11, n. 1, p. 52-81, 2019.

CRUZ, Francisco Brito. *Novo jogo, velhas regras: democracia e direito na era da nova propaganda politica e das fake news.* Belo Horizonte, MG: Grupo Editorial Letramento, Casa do Direito, 2020.

DAN, Viorela; PARIS, Britt; DONOVAN, Joan; *et al*. Visual Mis- and Disinformation, Social Media, and Democracy. *Journalism & Mass Communication Quarterly,* v. 98, n. 3, p. 641-664, 2021.

DAS, Anupam; SCHROEDER, Ralph. Online disinformation in the run-up to the Indian 2019 election. *Information, Communication & Society*, p. 1-17, 2020.

DOBBER, Tom; Ó FATHAIGH, Ronan; ZUIDERVEEN BORGESIUS, Frederik J. The regulation of online political micro-targeting in Europe. *Internet Policy Review*, v. 8, n. 4, 2019. Disponível em: https://policyreview.info/node/1440. Acesso em: 16 set. 2021.

DURACH, Flavia; BÂRGĂOANU, Alina; NASTASIU, Cătălina. Tackling Disinformation: EU Regulation of the Digital Space. *Romanian Journal of European Affairs*, v. 20, n. 1, 2020. Disponível em: http://rjea.ier.gov.ro/wp-content/uploads/2020/05/RJEA_vol.-20_no.1_June-2020_Full-issue.pdf#page=6.

EGELHOFER, Jana Laura; LECHELER, Sophie. Fake news as a two-dimensional phenomenon: a framework and research agenda. *Annals of the International Communication Association*, v. 43, n. 2, p. 97-116, 2019.

EVANGELISTA, Rafael; BRUNO, Fernanda. WhatsApp and political instability in Brazil: targeted messages and political radicalisation. *Internet Policy Review*, v. 8, n. 4, 2019. Disponível em: https://policyreview.info/node/1434. Acesso em: 27 jun. 2021.

FARIS, Robert M.; ROBERTS, Hal; ETLING, Bruce; *et al*. *Partisanship, Propaganda, and Disinformation: Online Media and the 2016 U.S. Presidential Election*. Cambridge, U.S.: Berkman Klein Center for Internet & Society at Harvard University, 2017. Disponível em: http://nrs.harvard.edu/urn-3:HUL.InstRepos:33759251.

FROSIO, Giancarlo. Why keep a dog and bark yourself? From intermediary liability to responsibilit. *International Journal of Law and Information Technology*, p. 1-33, 2017.

GASSER, Urs; SCHULZ, Wolfgang. Governance of Online Intermediaries: Observations from a Series of National Case Studies. *SSRN Electronic Journal,* 2015. Disponível em: http://www.ssrn.com/abstract=2566364. Acesso em: 6 out. 2021.

GORWA, Robert; ASH, Timonthy Garton. Democratic Transparency in the Platform Society. *In*: PERSILY, Nathaniel; TUCKER, Joshua A. (Orgs.). *Social media and democracy: the state of the field, prospects for reform.* Cambridge New York Port Melbourne New Delhi Singapore: Cambridge University Press, 2020.

GUESS, Andrew M.; LYONS, Benjamin A. Misinformation, Disinformation and Online Propaganda. *In*: PERSILY, Nathaniel; TUCKER, Joshua A. (Eds.). *Social Media and Democracy. The State of the Field, Prospects for Reform.* [s.l.]: Cambridge University Press, 2020.

HELBERGER, Natali. The Political Power of Platforms: How Current Attempts to Regulate Misinformation Amplify Opinion Power. *Digital Journalism*, v. 8, n. 6, p. 842-854, 2020.

HELDT, Amélie. Reading between the lines and the numbers: an analysis of the first NetzDG reports. *Internet Policy Review*, v. 8, n. 2, 2019. Disponível em: https://policyreview.info/node/1398. Acesso em: 10 out. 2021.

HOFMANN, Jeanette. Mediated democracy – Linking digital technology to political agency. *Internet Policy Review,* v. 8, n. 2, 2019. Disponível em: https://policyreview.info/node/1416. Acesso em: 23 ago. 2021.

IGLESIAS KELLER, Clara. Policy by judicialisation: the institutional framework for intermediary liability in Brazil. *International Review of Law, Computers & Technology*, p. 1-19, 2020.

JUNGHERR, Andreas; SCHROEDER, Ralph. Disinformation and the Structural Transformations of the Public Arena: Addressing the Actual Challenges to Democracy. *Social Media + Society,* v. 7, n. 1, p. 205630512198892, 2021.

KARPF, David. On Digital Disinformation and Democratric Myths. *MediaWell, Social Science Research Council,* 2019. Disponível em: https://mediawell.ssrc.org/expert-reflections/on-digital-disinformation-and-democratic-myths/. Acesso em: 10 jan. 2020.

KHAN, Irene. *Disinformation and freedom of opinion and expression.* [s.l.]: United Nations, General Assembly, 2021.

MAC SÍTHIGH, Daithí. The road to responsibilities: new attitudes towards Internet intermediaries. *Information & Communications Technology Law*, v. 29, n. 1, p. 1-21, 2020.

MACEDO JUNIOR, Ronaldo Porto. Freedom of Expression: what lessons should we learn from US experience? *Revista Direito GV*, v. 13, n. 1, p. 274-302, 2017.

MANSELL, Robin; STEINMUELLER, W. Edward. *Advanced introduction to platform economics.* Cheltenham, UK ; Northampton, MA: Edward Elgar Publishing, 2020. (Elgar advanced introductions).

MARSDEN, Chris; MEYER, Trisha; BROWN, Ian. Platform values and democratic elections: How can the law regulate digital disinformation? *Computer Law & Security Review,* v. 36, p. 105373, 2020.

MARWICK, Alice; KUO, Rachel; CAMERON, Shanice Jones; *et al. Critical Disinformation Studies – A Syllabus.* [s.l.]: Center for Information, Technology and Public Life – University of Noth Carolina at Chapel Hill, 2021. Disponível em: https://citap.unc.edu/research/critical-disinfo/. Acesso em: 5 ago. 2021.

MARWICK, Alice; LEWIS, Rebecca. *Media manipulation and disinformation online.* [s.l.]: Data & Society Research Institute, 2020. Disponível em: https://datasociety.net/wp-content/uploads/2017/05/DataAndSociety_MediaManipulationAndDisinformationOnline-1.pdf. Acesso em: 10 nov. 2020.

MENDES, Laura Schertel. *Privacidade, proteção de dados e direito do consumidor: linhas gerais de um novo direito fundamental.* São Paulo: Saraiva, 2014.

MOSES, Lyria Bennett. How to Think about Law, Regulation and Technology: Problems with 'Technology' as a Regulatory Target. *Law, Innovation and Technology*, v. 5, n. 1, p. 1-20, 2013.

NENADIĆ, Iva. Unpacking the "European approach" to tackling challenges of disinformation and political manipulation. *Internet Policy Review*, v. 8, n. 4, 2019. Disponível em: https://policyreview.info/node/1436. Acesso em: 27 jun. 2021.

NEO, Ric. The International Discourses and Governance of Fake News. *Global Policy*, v. 12, n. 2, p. 214-228, 2021.

OGNYANOVA, Katherine; LAZER, David; ROBERTSON, Ronald E.; *et al.* Misinformation in action: Fake news exposureis linked to lower trust in media, higher trust in government when your side is in power. *Harvard Kennedy School Misinformation Review*, 2020. Disponível em: https://misinforeview.hks.harvard.edu/?p=1689. Acesso em: 30 ago. 2021.

QI, Aimin; SHAO, Guosong; ZHENG, Wentong. Assessing China's Cybersecurity Law. *Computer Law & Security Review*, v. 34, n. 6, p. 1342-1354, 2018.

RAUCHFLEISCH, Adrian; KAISER, Jonas. The False positive problem of automatic bot detection in social science research. *PLOS ONE*, v. 15, n. 10, p. e0241045, 2020.

RIEDER, Bernhard; HOFMANN, Jeanette. Towards platform observability. *Internet Policy Review*, v. 9, n. 4, 2020. Disponível em: https://policyreview.info/articles/analysis/towards-platform-observability. Acesso em: 6 out. 2021.

ROSSINI, Patrícia; STROMER-GALLEY, Jennifer; BAPTISTA, Erica Anita; *et al.* Dysfunctional information sharing on WhatsApp and Facebook: The role of

political talk, cross-cutting exposure and social corrections. *New Media & Society*, v. 23, n. 8, p. 2430-2451, 2021.

SCHULDT, Lasse. The rebirth of Malaysia's fake news law – and what the NetzDG has to do with it. Disponível em: https://verfassungsblog.de/malaysia-fake-news/. Acesso em: 13 set. 2021.

SCHULZ, Wolfgang. *Roles and Responsibilities of Information Intermediaries: Fighting Misinformation as a Test Case for Human-Rights Respecting Governance of Social Media Platforms*. [s.l.]: Hoover Institution, Stanford University, 2019. (Aegis Series). Disponível em: https://www.hoover.org/sites/default/files/research/docs/schulz_webreadypdf.pdf. Acesso em: 12 dez. 2019.

SUZOR, Nicolas P.; WEST, Sarah Meyers; QUODLING, Andrew; *et al*. What Do We Mean When We Talk About Transparency? Toward Meaningful Transparency in Commercial Content Moderation. *International Journal of Communication*, v. 13, p. 1526-1543, 2019.

SUZOR, Nicolas; VAN GEELEN, Tess; MYERS WEST, Sarah. Evaluating the legitimacy of platform governance: A review of research and a shared research agenda. *International Communication Gazette*, v. 80, n. 4, p. 385-400, 2018.

SYLVAIN, Oliver. Internet governance and democratic legitimacy. *Federal Communications Law Journal*, v. 62, n. 2, p. 205-274, 2010.

TAMBINI, Damian. Rights and Responsibilities of Internet Intermediaries in Europe: The Need for Policy Coordination. Disponível em: https://www.cigionline.org/articles/rights-and-responsibilities-internet-intermediaries-europe-need-policy-coordination. Acesso em: 14 abr. 2021.

VALENTE, Jonas C. L. Regulando desinformação e fake news: um panorama internacional das respostas ao problema. *Comunicação pública*, v.14 n. 27, 2019. Disponível em: http://journals.openedition.org/cp/5262. Acesso em: 27 jun. 2021.

WALKER, Shawn; MERCEA, Dan; BASTOS, Marco. The disinformation landscape and the lockdown of social platforms. *Information, Communication & Society*, v. 22, n. 11, p. 1531-1543, 2019.

WARDLE, Claire; DERAKHSHAN, Hossein. *Information Disorder: Toward an interdisciplinary framework for research and policy making*. [s.l.]: Council of Europe, 2017. (Council of Europe Report).

WISEMAN, Jamie. Rush to pass 'fake news' laws during Covid-19 intensifying global media freedom challenges. *International Press Institute*, 2020. Disponível em: https://ipi.media/rush-to-pass-fake-news-laws-during-covid-19-intensifying-global-media-freedom-challenges/.

ZAROUALI, Brahim; DOBBER, Tom; DE PAUW, Guy; *et al*. Using a Personality-Profiling Algorithm to Investigate Political Microtargeting: Assessing the Persuasion Effects of Personality-Tailored Ads on Social Media. *Communication Research*, p. 009365022096196, 2020.

Sobre a autora:

Clara Iglesias Keller | *E-mail:* c.keller@leibniz-hbi.de

Coordenadora do Digital Disinformation Hub no Leibniz Institute for Media Research | Hans-Bredow-Institute e Pesquisadora Senior em Políticas da Digitalização no WZB Berlin Social Science Center. Doutora e Mestre em Direito Público pela Universidade do Estado do Rio de Janeiro – UERJ e L.LM em Direito da Tecnologia da Informação e da Mídia pela London School of Economics and Political Science. Autora dos livros "Regulação Nacional de Serviços na Internet: exceção, legitimidade e o papel do Estado" (LumenJuris, 2019) e "Media Law in Brazil" (International Encyclopeadia of Laws, Forthcoming).

Artigo convidado.