



Universidade do Estado do Rio de Janeiro

Centro de Ciências Sociais

Faculdade de Direito

Vinício Guimarães Salvarezza

#Falecommoderação:

**Fundamentos, Possibilidades e Limites para a regulação pública da
liberdade de expressão nas redes sociais**

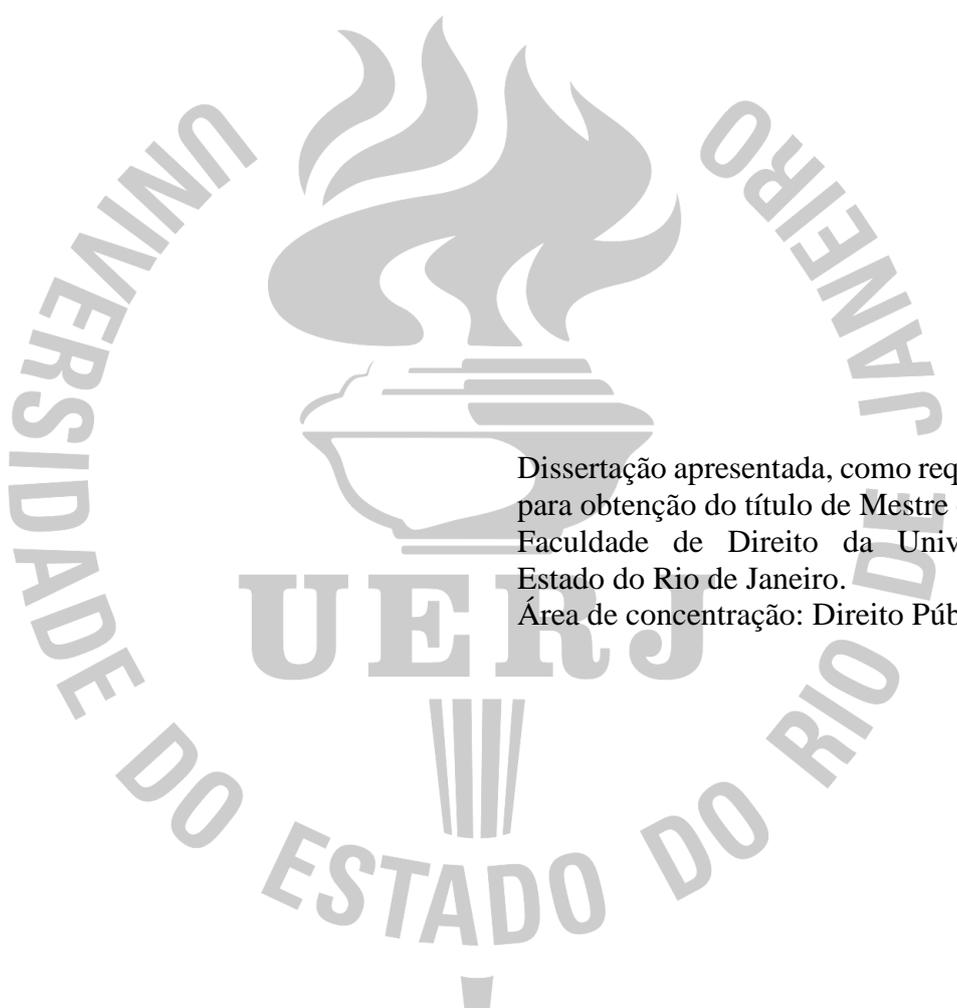
Rio de Janeiro

2023

Vinício Guimarães Salvarezza

#Falecommoderação:

Fundamentos, Possibilidades e Limites para a regulação pública da liberdade de expressão nas redes sociais



Dissertação apresentada, como requisito parcial para obtenção do título de Mestre em Direito, à Faculdade de Direito da Universidade do Estado do Rio de Janeiro.
Área de concentração: Direito Público.

Orientador: Prof. Dr. José Vicente Santos de Mendonça

Rio de Janeiro

2023

CATALOGAÇÃO NA FONTE
UERJ/REDE SIRIUS/BIBLIOTECA CCS/C

S182

Salvarezza, Vinicio Guimarães

#Falecommoderação: fundamentos, possibilidades e limites para a regulação pública da liberdade de expressão nas redes sociais/ Vinicio Guimarães Salvarezza. - 2023.
199 f.

Orientador: Prof. Dr. José Vicente Santos de Mendonça
Dissertação (Mestrado). Universidade do Estado do Rio de Janeiro,
Faculdade de Direito.

1. Liberdade de expressão - Teses. 2.Redes sociais – Teses. 3.Regulação – Teses. I.Mendonça, José Vicente Santos de. II. Universidade do Estado do Rio de Janeiro. Faculdade de Direito. III. Título.

CDU 316.422

Bibliotecária: Marcela Rodrigues de Souza CRB7/5906

Autorizo, apenas para fins acadêmicos e científicos, a reprodução total ou parcial desta tese, desde que citada a fonte.

Assinatura

Data

Vinício Guimarães Salvarezza

#Falecommoderação:

Fundamentos, Possibilidades e Limites para a regulação pública da liberdade de expressão nas redes sociais

Dissertação apresentada, como requisito parcial para obtenção do título de Mestre em Direito, à Faculdade de Direito da Universidade do Estado do Rio de Janeiro.
Área de concentração: Direito Público.

Aprovada em 01 de março de 2023.

Banca Examinadora:

Prof. Dr. José Vicente Santos de Mendonça (Orientador)
Faculdade de Direito – UERJ

Prof. Dr. Gustavo Binenbojm
Faculdade de Direito – UERJ

Prof^a. Dra. Clara Iglesias Keller
The Wissenschaftszentrum Berlin für Sozialforschung

Rio de Janeiro

2023

AGRADECIMENTOS

Após período tão intenso quanto gratificante como o mestrado faz-se necessário um agradecimento especial a todos aqueles que contribuíram e se fizeram especial nessa jornada.

Aos meus pais, meus maiores incentivadores, agradeço pelo apoio que sempre dispensaram a todos os meus projetos de vida e pela educação que me conferiu a disciplina necessária para passar por um período tão extenuante como o mestrado com a resiliência que o momento exigia.

À minha esposa e melhor amiga Ana Gabriela, agradeço não só pela paciência para lidar com os momentos de minha ausência, mas pela segurança que sempre me passou ao acreditar em mim e no meu trabalho, mesmo nos meus piores dias.

Ao José Vicente, meu orientador, sou grato não só pelas argutas observações e sugestões apresentadas à dissertação, mas pela disposição sempre imediata de me auxiliar nessa laboriosa jornada, tanto na orientação conferida durante estágio-docente, quanto na viabilização da montagem da banca tal como idealizado por mim.

Aos professores Gustavo Binenbojm e Clara Keller, meu especial agradecimento pela pronta aceitação ao convite para participação da banca examinadora. Me causa enorme alegria poder falar de um tema tão importante quanto atual na presença de duas pessoas cujos trabalhos foram inspirações para as ideias aqui expostas.

Aos demais professores da UERJ, agradeço pelas trocas, experiências e aprendizados que tornaram minha passagem pela universidade especialmente rica e justificaram os motivos pelo qual a linha de Direito Público da Instituição é considerada a melhor do país.

Aos colegas, tanto do mestrado quanto da Procuradoria do Município de Niterói, agradeço pelos momentos de leveza e dicas acerca dos temas aqui desenvolvidos. Agradeço, em especial, Arthur Pinel e Vitor Gomes, amigos que dedicaram parte substancial do seu tempo para leitura do trabalho e apresentação de sugestões pontuais à dissertação.

Agradeço, por fim, ao meu filho Bernardo, grata surpresa que tive no último ano. Desde o seu nascimento, se tornou o centro de minha vida e a principal razão para a alegria dos meus dias. A ele, também dedico este trabalho.

RESUMO

SALVAREZZA, Vinicio Guimarães. **#Falecommoderação: fundamentos, possibilidades e limites para a regulação pública da liberdade de expressão nas redes sociais**. 2023. 199 f. Dissertação (Mestrado em Direito) – Faculdade de Direito, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2023.

A presente dissertação tem por objetivo propor um modelo de regulação pública das redes sociais calcada na construção de um arranjo corregulatório, que permita a responsabilização administrativa de tais plataformas, sem esquecer os especiais impactos que uma proposta deste tipo pode ter sobre o direito fundamental à liberdade de expressão. O trabalho se divide em quatro capítulos. No primeiro, são firmadas premissas a serem consideradas quando se estabelece limites ao exercício da liberdade de expressão. No segundo capítulo, são apresentados os riscos que o poder de governança das principais plataformas de redes sociais sobre o discurso representa à sociedade, bem como as novas formas de censura à disposição do Estado. No terceiro capítulo, expõe-se as principais experiências regulatórias sobre o tema e, na última parte, se defende a criação de deveres a serem observados pelas redes sociais tanto na moderação, quanto na definição dos critérios de visibilidade/invisibilidade dos conteúdos publicados necessários à efetiva constitucionalização da esfera digital

Palavras-chave: Liberdade de Expressão. Redes Sociais. Corregulação. Desordem informacional. Constitucionalismo Digital.

ABSTRACT

SALVAREZZA, Vinicio Guimarães. **#Speakwithmoderation. Fundamentals, possibilities and limits for the public regulation of freedom of speech in social networks.** 2023. 199 f. Dissertação (Mestrado em Direito) – Faculdade de Direito, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2023.

This paper aims to propose a model of public regulation of social networks based on the construction of a co-regulatory arrangement that allows the administrative accountability of such platforms, without forgetting the special impacts that a proposal may have on the fundamental right to freedom of speech. The paper is divided into four chapters. The first chapter establishes the premises to be considered when establishing limits to the exercise of freedom of speech. The second chapter presents the risks that the power of governance of the main social networks over speech represents to society, as well as the new forms of censorship available to the State. In the third chapter, the main regulatory experiences are exposed and, in the last part, the creation of duties to be observed by social networks is defended, both in moderation and in the definition of the criteria of visibility/invisibility of the published content, in order to ensure an effective constitutionalization of the digital sphere.

Keywords: Freedom of Speech. Social Networks. Co-regulation. Information Disorder. Digital Constitutionalism.

SUMÁRIO

INTRODUÇÃO	8
1 REGULAÇÃO PÚBLICA E LIBERDADE DE EXPRESSÃO	13
1.1 A concepção libertária e democrática da liberdade de expressão	13
1.2 As razões para a proteção da liberdade de expressão	20
1.2.1 <u>A metáfora do “livre mercado de ideias”</u>	20
1.2.2 <u>A proteção da democracia</u>	26
1.2.3 <u>A proteção da autonomia</u>	30
1.3 Os riscos da regulação pública	34
1.3.1 <u>A Censura Estatal e o Autoritarismo</u>	34
1.3.2 <u>O chilling effect</u>	39
1.4 Liberdade de expressão e regulação de arquiteturas	47
1.5 Parâmetros para a regulação pública da liberdade de expressão	50
2 A LIBERDADE DE EXPRESSÃO NO SÉCULO XXI: NOVOS DESAFIOS E RISCOS	53
2.1 A liberdade de expressão nas camadas da internet: da neutralidade à imunidade dos intermediários	53
2.2 A Seção 230 do CDA	58
2.3 A imunidade dos intermediários no Brasil: a construção do art. 19 do Marco Civil da Internet	60
2.4 A teoria triangular da liberdade de expressão	65
2.5 Estados e redes sociais: compreendendo os novos tipos de censura	69
2.5.1 <u>A censura colateral</u>	69
2.5.2 <u>A censura reversa</u>	74
2.6 Redes sociais e esfera pública: os novos desafios	79
2.6.1 <u>As externalidades algorítmicas</u>	79
2.6.2 <u>Moderação e censura privada</u>	88
2.7 A insuficiência do art. 19 do Marco Civil da Internet: a necessária busca por um modelo de regulação das plataformas de redes sociais	94
3 CONTEÚDO, ARQUITETURAS E (COR)REGULAÇÃO: UMA ANÁLISE DAS ESTRATÉGIAS E ARRANJOS REGULATÓRIOS	100

3.1	Regulação e combate à desinformação: conteúdo x arquitetura	100
3.2	Aclaramentos conceituais: autorregulação, correção e autorregulação regulada	112
3.3	Um novo modelo autorregulatório: uma avaliação do <i>Oversight Board</i>	118
3.4	O exemplo correção alemão: a <i>NetzDg</i>	126
3.5	Fundamentos constitucionais para a priorização da correção das redes sociais	132
4	A CONSTITUCIONALIZAÇÃO DO ESPAÇO DIGITAL E OS DEVERES DAS REDES SOCIAIS	137
4.1	Considerações gerais	137
4.2	Os deveres das redes sociais	139
4.2.1	<u>O dever de transparência</u>	140
4.2.2	<u>O devido processo legal</u>	145
4.2.3	<u>A proteção à boa-fé dos usuários</u>	153
4.3	Da responsabilidade civil à responsabilidade administrativa: reflexões sobre o papel do Poder Judiciário em um contexto correção	160
4.4	Os limites à regulação pública das novas redes de comunicação	167
	CONCLUSÕES	175
	REFERÊNCIAS	181

INTRODUÇÃO

“Proteger os valores da liberdade de expressão na era digital será cada vez menos um problema de direito constitucional, embora essas proteções se mantenham muito importantes, e cada vez mais um problema de tecnologia e regulação administrativa”.¹

Segundo pesquisas recentes, as redes sociais somam hoje mais de 4,62 bilhões de usuários ao redor do mundo.² O Brasil se destaca neste cenário, representando hoje o terceiro país do mundo em que as pessoas passam mais tempo na *internet*³, sendo as redes sociais, com ligeira vantagem sobre a televisão, a principal fonte de informação e comunicação no país⁴.

Tais dados revelam o impacto das redes sociais para o debate público, transformando-as na “praça pública” do século XXI, e tiveram o mérito de descentralizar o controle sobre a cultura, a produção de informação e sobre o acesso aos meios de comunicação de massa, mas também transformaram a cultura e a informação em valiosas *commodities* que permitem a mercantilização do conhecimento, pautando-o por uma lógica de captura da atenção.

Este é o plano de fundo contextual que embasa a afirmativa do professor Balkin, que inicia este trabalho.

Durante séculos, o debate estatal sobre a liberdade de expressão se concentrou em uma visão dicotômica que opõe, de um lado, o Estado e, de outro, a sociedade civil formada por uma miríade de indivíduos com múltiplas ideias, crenças e opiniões. Neste campo, o enfoque das discussões sobre a liberdade de expressão fora sempre refletir sobre os limites da intervenção estatal sobre a esfera discursiva, ora percebido como ameaça, ora como um possível promotor de valores fundantes.

A ascensão da *internet* e, especialmente, das redes sociais embaralha um pouco este debate. Permanece, sem dúvida, o receio do Leviatã Estatal, mas surgem novos agentes, que tal como o Estado gozam de efetivo poder de governança para influenciar o exercício da liberdade de expressão na esfera pública. A liberdade de expressão passa a estar não só sob ameaça de pretensos ditadores, mas também de conglomerados que administram redes sociais, plataformas

¹ BALKIN, Jack M. O Futuro da Liberdade de Expressão na Era Digital. In: SAMPAIO, José Adércio Leite (Coord.). *Liberdade de Expressão no Século XXI*. Belo Horizonte: Del Rey, 2016, p. 370.

² A pesquisa pode ser conferida em artigo de Simon Kemp (Digital 2022: global overview report. *Datareportal*, [s.l.], 26 jan. 2022. Acesso em: <https://datareportal.com/reports/digital-2022-global-overview-report>. Acesso em: 07 fev. 2022.).

³ Com uma média de 10 horas e 19 minutos de uso por dia. O resultado pode ser conferido na pesquisa supracitada.

⁴ O estudo da Reuters Digital pode ser conferido em: Pela 1ª vez, rede social é mais citada que TV como fonte de notícia no Brasil. *Poder 360*, [s.l.], 16 jun. 2020. Disponível em: <https://www.poder360.com.br/midia/pela-1a-vez-rede-social-e-mais-citada-que-tv-como-fonte-de-noticia-no-brasil/>. Acesso em: 07 fev. 2022.

privadas que atuam como canais de intermediação para comunicação e troca de toda espécie de dados no âmbito da *internet*. Tais plataformas, como se verá, apresentam uma substancial capacidade de influência sobre o comportamento dos cidadãos, tendo, ainda, poderes para limitar a manifestação e divulgação de opiniões, informações e comportamentos ao redor do globo terrestre.

E, aqui, situa-se o principal ponto de perplexidade que move o propósito desta pesquisa. Dada a relevância e o poder das plataformas de redes sociais, na condição de canais de difusão de conhecimento, informações e manifestações em geral, é possível ao Estado impor limites ao poder de governança que os atores privados exercem sobre estes canais discursivos? E, em caso positivo, que limites seriam estes e quais os riscos da eventual ação estatal?

Algumas explicações conceituais introdutórias fazem-se necessárias para a adequada compreensão do escopo do trabalho que se apresentará.

A primeira é a ideia de “regulação pública”, que consta do título. Como se aprofundará ao longo da dissertação, o conceito de regulação adotado é bastante amplo, compreendendo todo processo de influência sobre o comportamento de terceiros, com o propósito de alcançar um resultado específico. Nesta toada, não somente a lei seria um instrumento regulatório, mas também normas sociais, normas privadas de mercado e o código (as “arquiteturas” das redes, segundo Lessig).⁵

O uso da expressão “pública” no título tem, portanto, o propósito de evidenciar que a ideia principal deste trabalho é discutir o papel que caberia ao Estado, na regulação das plataformas privadas e eventuais impactos desta ação sobre a liberdade de expressão, *por meio de lei*, uma vez que salvo uma lacônica previsão de responsabilização civil no art. 19 do Marco Civil da *Internet* (MCI), não há qualquer tipo de norma jurídica que regule a ação dos agentes privados gerenciadores das principais redes sociais do país.

O segundo conceito que se antecipa é o de *desordens informacionais*. Diante das variadas expressões que surgiram nos últimos anos (*fake news*, *ciberbullying*, *trolling*, *desinformação* e afins) para identificar os danos discursivos produzidos na *internet*, há um esforço doutrinário a nível internacional para aglutinar estes rótulos sob o gênero de “desordens informacionais” e que, conforme se sustenta neste trabalho, é a principal externalidade negativa oriunda do poder de governança das redes sociais. O termo engloba três espécies de “desordens”: a *disinformation* (desinformação), que seria toda informação enganosa, com aparência jornalística, divulgada por meio de redes sociais, com o propósito de obter proveito

⁵ Cf. LESSIG, Lawrence. *Code and Others Laws of Cyberspace, version 2.0*. Nova Iorque: Basic Books, 2006.

político ou econômico; a *misinformation*, que é a informação enganosa que circula nas redes sociais, mas reproduzida por pessoa de boa-fé; e a *mal-information*, informação maliciosamente divulgada e que engloba um conjunto de discursos danosos, mas que não são propriamente falsos (vazamentos de fatos privados, fatos descontextualizados, discursos de ódio e assédios). Embora não seja possível ao longo da dissertação abordar em minúcias a peculiaridade de cada uma das espécies, a menção ao gênero e a identificação de como as redes sociais contribuem para este ciclo se faz importante por ser este um dos elementos justificadores da regulação pública das plataformas.

O plano de trabalho da pesquisa se dividirá em quatro capítulos.

O primeiro capítulo é focado no Estado e suas formas de intervenção sobre a liberdade de expressão. Essa intervenção pode ocorrer diretamente sobre o conteúdo (por meio da criação de novas ilicitudes discursivas) ou, ainda, com fundamento no art. 174 da Constituição da República, quando a regulação incide sobre as ações de um agente econômico (tais como as plataformas privadas de redes sociais ou canais de mídia de massa), tendo impacto reflexo sobre a liberdade de se expressar.

O escopo deste primeiro capítulo é analisar seriamente os fundamentos que justificam a proteção da liberdade de expressão, bem como os riscos atribuídos à intervenção estatal sobre a esfera pública (o deslize ao autoritarismo e o *chilling effect*). Argumentar-se-á ao longo do capítulo que o peso dado aos riscos e virtudes de eventual regulação pública sobre a liberdade de expressão e canais discursivos é fortemente influenciado pela tradição que inspira a concepção da própria ideia de liberdade de expressão (libertária ou democrática). Enquanto no primeiro modelo há uma ênfase quase que absoluta na liberdade de quem discursa, a concepção democrática busca equilibrar a legítima pretensão do falante com a liberdade dos ouvintes, que pode vir a ser cerceada ou prejudicada por eventuais abusos discursivos. O estudo dessas duas tradições, como premissa inicial para definição de parâmetros para a regulação estatal em matéria de liberdade de expressão, é necessário tanto em razão do crescimento na difusão de ideias libertárias no Brasil⁶, quanto pelo fato histórico de que a própria construção da *internet* foi fortemente influenciada por inspirações libertárias.

O capítulo segundo direciona o olhar para as plataformas de redes sociais e sua evolução dentro da *internet*. Será feita uma análise histórica do contexto que permeou a construção da

⁶ A respeito, confira-se MENDES, Conrado Hubner. Freedom Without Responsibility: the Promise of Bolsonaro's COVID-19 Denial. *Jus Cogens*, [s.l.], v. 3, p. 181-207, 2021).

norma da “imunidade dos intermediários”⁷, oriunda dos Estados Unidos, e como esta regra foi importada para o Brasil, levando à atual redação do art. 19 do MCI. A insuficiência da atual regra jurídica vigente no país restará evidenciada pela incapacidade de o Poder Judiciário lidar de forma sistêmica com os desafios impostos pela liberdade de expressão na era digital.

Para corroborar esta análise, será feita referência à estrutura triangular da liberdade de expressão, à qual o professor Jack Balkin tem aludido em trabalhos recentes, buscando enriquecer suas considerações gerais com estudos empíricos e teóricos de outros especialistas na matéria. Em síntese, serão enfatizados os riscos de captura pelo Estado da infraestrutura privada destas plataformas para vigilância e cerceamento da liberdade de expressão (por meio dos fenômenos como a censura colateral e reversa) e os riscos que estas plataformas privadas representam aos cidadãos em geral, em razão da sua opacidade tecnológica e ausência de *accountability*.

No capítulo terceiro, será apresentada uma avaliação das estratégias regulatórias que têm sido utilizadas ao redor do mundo para resolução do problema dos abusos da liberdade discursiva nas redes sociais. O capítulo terceiro dialoga com o primeiro, conferindo-lhe maior concretude, já que as estratégias têm se dividido entre aquelas que apostam na limitação direta do discurso, por meio de criminalização de condutas como “divulgação de *fake news*”, e estratégias que apostam na reorientação dos algoritmos destas redes a fins constitucionalmente legítimos, por meio da exigência de maior transparência sobre seu funcionamento.

Há ainda distinção no arranjo regulatório idealizado para o alcance deste objetivo. Enquanto a maior parte dos Estados democráticos tensionam para a instituição de regimes corregulatórios, as principais plataformas privadas tentam sofisticar a autorregulação que já exercem. Ao longo do capítulo serão analisados os casos específicos do *Oversight Board*, instituído pela empresa Meta, e a legislação alemã que regula aplicação da lei no âmbito das redes sociais (*NetzDg*), com o propósito de conferir maior riqueza à exploração das virtudes e defeitos de cada uma das propostas. Por fim, será defendida a constitucionalidade de estratégias que priorizem o maior controle sobre a arquiteturas das redes, viabilizadas por meio de arranjos corregulatórios.

O quarto e último capítulo terá por objetivo detalhar um parâmetro mínimo para o conjunto de obrigações ao qual as plataformas de redes sociais estarão vinculadas no caso de instituição de um regime regulatório público, partindo-se da ideia de uma necessária constitucionalização do espaço digital. Fala-se em parâmetro mínimo porque as redes sociais

⁷ Regra que alude à impossibilidade de as plataformas serem responsabilizadas por conteúdos *postados* por terceiros.

são diversas e a evolução tecnológica muito acelerada, o que demanda sempre a necessidade de uma margem plausível de discricionariedade para as plataformas de redes sociais no cumprimento destes objetivos. Ainda será avaliado como este novo regime de responsabilização das redes sociais poderá dialogar com o art. 19 do MCI.

A dissertação tem duas finalidades. De um lado busca descrever os danos que o abuso do exercício da liberdade de expressão tem produzido na atualidade, impulsionado muitas vezes pela negligência e imprudência das plataformas das redes sociais, bem como situar as principais propostas estatais regulatórias sobre a matéria. Há, também, uma finalidade prescritiva: a defesa de que o regime jurídico atual, com uma previsão de limitada responsabilidade civil, com base no art. 19 do MCI, é incapaz de lidar com os desafios que estas novas tecnologias impõem. Propõe-se assim um modelo complementar de responsabilização.

A pesquisa utilizou como referência uma bibliografia especializada sobre a matéria, englobando tanto materiais teóricos quanto pesquisas empíricas, análise de legislações e documentos oficiais, bem como a jurisprudência de tribunais pátrios (com ênfase no Supremo Tribunal Federal) e tribunais estrangeiros.

O trabalho, por óbvio, não exaure o enorme desafio que estas novas tecnologias suscitam em matéria de proteção de direitos fundamentais, nem tem pretensão de esgotar o tema. Há uma série de outras abordagens possíveis, que poderiam dar ênfase a outros aspectos, como: proteção de dados pessoais, regulação da concorrência entre plataformas, o anonimato na *internet* e a necessidade de sua compatibilização com o poder de polícia (especialmente, na área penal) do Estado. Embora parte destes temas sejam abordados de forma incidental, o foco deste trabalho está na construção de arranjos regulatórios, que estimulem arquiteturas tecnológicas mais responsivas aos anseios de seus usuários, o que requer transparência na definição tanto dos critérios que orientam a curadoria algorítmica (definição daquilo que terá maior ou menor visibilidade nas plataformas), quanto no exercício da moderação (controle sobre aquilo que será ou não postado, de acordo com os termos de uso e normas comunitárias das redes sociais). Entende-se que estes elementos são os mais importantes quando se debate os riscos impostos pelas redes sociais à liberdade de expressão.

1 REGULAÇÃO PÚBLICA E LIBERDADE DE EXPRESSÃO

1.1 A concepção libertária e democrática da liberdade de expressão

A liberdade de expressão consiste em um direito considerado fundamental pelas sociedades democráticas e que abarca tanto a liberdade de expressar, ideias, opiniões e pensamentos (liberdade de expressão em sentido estrito), quanto a liberdade de informar (que se refere a fatos noticiáveis). A liberdade de expressão é um direito de todos e, quando exercido por entidades de comunicação em geral, é comumente referenciado como “liberdade de imprensa”, que pode abarcar tanto a liberdade de informação (ao noticiar fatos) quanto a liberdade de expressão em sentido estrito (ao reportar ideias e pensamentos).⁸

Apesar da sua natureza fundamental, a liberdade de expressão não é um direito absoluto, podendo ser objeto de limitações por força de leis ou decisões judiciais.

A restrição oriunda da lei pode ocorrer em duas hipóteses. Na primeira, quando a lei cria objetivamente uma restrição jurídica a uma determinada espécie de discurso, fundada em uma ameaça de sanção penal, civil ou administrativa. Considere, por exemplo, o caso brasileiro do crime de calúnia (art. 138 CP), que penaliza o discurso falso que implique imputação de crime.

Na segunda, as restrições podem ainda ser direcionadas a uma atividade econômica, com o pretense propósito de orientá-la a fins constitucionalmente legítimos. Trata-se assim de uma medida pública regulatória indireta, cujo fundamento decorre do art. 174 da Constituição da República. Valendo-se das lições de Alexandre Santos de Aragão, a regulação, quando pública, pode ser compreendida como “medidas legislativas, administrativas, convencionais, materiais ou econômicas, abstratas ou concretas, pelas quais o Estado, de maneira restritiva da autonomia empresarial ou meramente indutiva, determina, controla, ou influencia o comportamento dos agentes econômicos, evitando que lesem os interesses sociais definidos no marco da Constituição e os orientando em direções socialmente desejáveis”.⁹

Quando essa ação regulatória tem o propósito de reorientar atividades de empresas privadas das áreas de comunicação (o que abarca tanto mídias de massa quanto redes sociais)

⁸ A respeito, cf. CHEQUER, Cláudio. *A Liberdade de Expressão como Direito Preferencial Prima Facie: análise crítica de revisão ao padrão jurisprudencial brasileiro*. 2. ed. Rio de Janeiro: Lumen Juris, 2017, p. 17).

⁹ ARAGÃO, Alexandre Santos de. *Curso de Direito Administrativo*. 2. ed. Rio de Janeiro: Forense, 2013, p. 420.

há um evidente impacto sobre o exercício da liberdade de expressão, pois estes agentes privados são essenciais para a construção de uma esfera pública comunicacional.

É possível, evidentemente, que o direito à liberdade de expressão seja também restringido por decisões judiciais, observado os limites do ordenamento jurídico. O próprio Supremo Tribunal Federal, na interpretação de expressões excessivamente lacunosas, abrangentes e, não raro, contraditórias da Constituição da República, constrói novos condicionamentos ao seu exercício no confronto com outros direitos fundamentais¹⁰. A atribuição de limites mais ou menos restritos, vale dizer, é reflexo muitas vezes do contexto jurídico, político e cultural nacional.¹¹

O grande questionamento que se apresenta quando falamos nestes limites, por lei ou decisão judicial, às liberdades discursivas é a busca por um critério que assente a razoabilidade da sua limitação.

As soluções aventadas são fortemente influenciadas pela tradição incorporada à compreensão do alcance da liberdade de expressão. Embora este direito fundamental tenha surgido como um elemento de contraposição ao poder das antigas monarquias absolutistas¹², duas foram as tradições que se desenvolveram desde a sua positivação na Declaração de Direitos dos Homens e dos Cidadãos, em 1789¹³ e na Primeira Emenda à Constituição Norte-Americana, em 1791¹⁴.

¹⁰ Enfatizando a distinção entre texto e norma e reconhecendo a dimensão construtiva da atividade interpretativa, cf. ÁVILA, Humberto. *Teoria dos princípios: da definição à aplicação dos princípios jurídicos*. 16. ed. São Paulo: Malheiros, 2015.

¹¹ Em interessante artigo, Frederick Schauer destaca a importância do contexto cultural e jurídico para a evolução da compreensão sobre os limites à liberdade de expressão. Cf. Schauer (Liberdade de Expressão em Decisões na Europa e nos Estados Unidos: Um Estudo de Caso na Arquitetura Constitucional Comparada In: SAMPAIO, José Adércio Leite (Coord.). *Liberdade de Expressão no Século XXI*. Belo Horizonte: Del Rey, 2016).

¹² A certidão de nascimento da “liberdade de expressão” remonta aos episódios da Revolução Gloriosa, em 1689, na Inglaterra. Com a destituição do Rei Jaime II da Inglaterra e a assunção do trono pelo Rei Guilherme de Orange (Chefe de Estado da Holanda à época) e sua esposa, fora elaborado o texto da *Bill of Rights*, que estabeleceu uma série de limitações ao exercício do poder pelo monarca, cabendo aqui destacar a prescrição de que a liberdade de expressão, debates ou procedimentos no parlamento não devem ser impedidos ou questionados por qualquer tribunal ou local fora do parlamento. Para maior aprofundamento do histórico de evolução da liberdade de expressão, recomenda-se a leitura do texto de Pamplona (O conteúdo do direito da liberdade de expressão, o discurso de ódio e a resposta democrática. *Revista Brasileira de Direito*, Passo Fundo, v. 14, n. 1, p. 297-316, jan./abr., 2018).

¹³ Aqui, vale destacar que a liberdade de expressão só é prescrita como uma liberdade indefinida pela Constituição Jacobina em 1793. Antes desta data, o texto continha uma acentuada tendência à censura prévia. Cf. CHEQUER, Cláudio. *A Liberdade de Expressão como Direito Preferencial Prima Facie: análise crítica de revisão ao padrão jurisprudencial brasileiro*. 2. ed. Rio de Janeiro: Lumen Juris, 2017, p. 11.

¹⁴ Em tradução livre, assim dispõe a primeira emenda: “O Congresso não deverá fazer qualquer lei a respeito de um estabelecimento de religião, ou proibir o seu livre exercício; ou restringindo a liberdade de expressão, ou da imprensa; ou o direito das pessoas de se reunirem pacificamente, e de fazerem pedidos ao governo para que sejam feitas reparações de queixas” (no original: *Congress shall make no law respecting an establishment of religion, or prohibiting the free exercise thereof; or abridging the freedom of speech, or of the press; or the right of the people peaceably to assemble, and to petition the Government for a redress of grievances*). Embora o texto da Primeira

Uma — mais antiga — é a concepção libertária (ou *ultraliberal*)¹⁵ da liberdade de expressão, que vê no direito fundamental uma garantia principalmente do falante (*speaker*) de exercer sua autonomia, tese de forte tração nos Estados Unidos.¹⁶ Assim, regulações públicas ou privadas sobre o discurso ou canais discursivos somente se justificariam em situações extremamente excepcionais e não poderiam proibi-los por causa do ponto de vista que ele expressa, cabendo ao Estado atuar com absoluta neutralidade na imposição de qualquer limite ao exercício deste direito.¹⁷ Como se verá com maior profundidade no próximo capítulo, esta é a lógica que embasou o modelo autorregulatório de governança privada como o padrão para o desenvolvimento de novas plataformas tecnológicas no âmbito da *internet*.¹⁸

É preciso um adendo. Mesmo entre os defensores de uma concepção libertária da liberdade de expressão, cresce o entendimento de que as principais plataformas de redes sociais são agentes quase tão poderosos quanto o Estado na definição de limites ao exercício da liberdade de expressão e, portanto, precisam de maior controle. O apego, contudo, à tese da prevalência da autonomia do falante, sem maiores preocupações com os impactos concretos do discurso, tem levado ao reconhecimento da necessidade de uma regulação cujo principal objetivo seja reduzir substancialmente a autonomia destas empresas para moderarem conteúdos, reforçando um dever de respeito a todos os pontos de vista. Para o libertário do século XXI, as plataformas privadas que operam as redes sociais precisam de controles, pois moderam demais. Assim, moderar melhor é necessariamente moderar menos.¹⁹

Emenda da Constituição Norte-Americana aluda apenas ao Congresso, é aceito que esta cláusula também delimita os poderes do Estado e do Governo Federal em sentido amplo, em virtude de sua incorporação pela Décima Quarta Emenda (Stromberg v. Califórnia, US 283 359 (1931); Gitlow v. New York, 268 US 652 (1925)).

¹⁵ Há alguma resistência dentre os defensores da “teoria libertária”, em reconhecê-la por esta alcunha, pois seria uma referência aparentemente pejorativa. Argumenta-se que mesmo seus defensores tradicionais (dentre os quais se destaca o *Justice* Holmes) admitem algumas restrições, ainda que em hipóteses bastante estritas. Seria assim uma “teoria bem liberal”, mas não propriamente libertária. Neste sentido, Bruno Lemos Silva (*O paradoxo do direito de manifestação: a regulação que liberta*. 2016. 252 f. Dissertação (Mestrado em Direito do Estado) – Faculdade de Direito, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2016, p. 103).

¹⁶ Orientação que guarda sustentação na clássica obra de John Stuart Mill (*Sobre a Liberdade*. Trad. Denise Bottmann. Porto Alegre: L&PM, 2019).

¹⁷ Trata-se do princípio da neutralidade adotado pela jurisprudência da Suprema Corte Norte-Americana (cf. SUNSTEIN, Cass. *Republic.com 2.0*. Princeton, Oxford: Princeton University Press, 2009, p. 183-186).

¹⁸ O objetivo principal foi justamente assegurar a autonomia ampla dos usuários e das plataformas privadas para inovarem — livre de qualquer amarra estatal — no âmbito da *internet*.

¹⁹ Este foi o raciocínio que motivou a “ordem executiva para a prevenção da censura *online*” assinada pelo Presidente Donald Trump em maio de 2020, sob o argumento de que “Twitter, Facebook, Instagram e Youtube exercem imenso poder para modelar a interpretação de eventos públicos; e para censurar, apagar ou fazer sumir informações; e controlar os que as pessoas podem ou não ver”. Sugere, assim, um dever de neutralidade na regulação do ponto de vista dos agentes discursivos, *speakers*, aproximando as obrigações destes fóruns às obrigações impostas ao Estado norte-americano à luz da interpretação jurisprudencial que se faz naquele país da primeira emenda. A íntegra da ordem executiva pode ser conferida em **EXECUTIVE order on preventing online censorship. White House**, [s.l.], 28 mai. 2020. Disponível em: <https://www.whitehouse.gov/presidential-actions/executive-order-preventing-online-censorship/>. Acesso em: 25 12. 2022. Esta lógica também pautou lei do

Trata-se de teoria que enfatiza, especialmente, a dimensão defensiva e subjetiva dos direitos fundamentais. A liberdade de expressão é, antes de tudo, um direito subjetivo que confere ao indivíduo a prerrogativa de exigir que o Estado — e, numa visão mais recente, plataformas privadas na *internet* — não interfira nas suas relações comunicativas, assegurando uma livre esfera pública.

Por outro lado, a teoria democrática da liberdade de expressão vê no Estado um papel de promotor do debate público, a quem cabe não só observar a autonomia dos falantes (*speakers*), mas fomentar a construção de uma esfera pública plural, aberta e reflexiva. A intervenção estatal poderia se justificar, portanto, não só em nome de outros interesses relevantes (e.g., a igualdade), mas também como ferramenta de promoção da própria liberdade de expressão em sua dimensão objetiva.²⁰ A restrição à “liberdade do falante” não implica necessariamente uma restrição ao direito à liberdade de expressão enquanto valor jurídico fundamental. Por vezes, regular é necessário para ampliar o número de vozes a serem ouvidas, promover o debate ou para findar o silêncio daqueles que não têm voz. Para o adepto desta teoria, a ascensão de estruturas privadas, como plataformas de redes sociais, deve ser regulada pelo Estado para, principalmente, conferir maior transparência aos critérios com que exercem seu poder de governança discursiva de modo a orientá-lo a propósitos minimamente democráticos. Neste contexto, a qualidade da atividade de controle de conteúdo não tem, necessariamente, correlação direta com a quantidade de discurso que é restringido.

A linha ora defendida se aproxima da última concepção, não só por parecer uma teoria melhor, mas também por ser aquela que melhor dialoga com nosso ordenamento constitucional.

É preciso, entretanto, reconhecer que tal posicionamento implica alguns desafios próprios. De um lado, a atribuição de um papel ao Estado de “promotor” da liberdade de

Estado do Texas que impunha às redes sociais a obrigação de não realizar qualquer discriminação contra usuário com base em seu ponto de vista. A norma fora declarada inconstitucional pela Suprema Corte Norte-Americana por um apertado placar de 5 a 4. Informação disponível em Fung e Vogue (Texas asks Supreme Court to keep state social media censorship law in effect, *CNN*, [s.l.], 19 mai. 2022. Disponível em: <https://www.cnn.com/2022/05/18/tech/texas-supreme-court-social-media/index.html>. Acesso em: 01 jun. 2022, e JUSTIÇA dos EUA bloqueia lei do Texas que restringe empresas de redes sociais. *CNN Brasil*, [s.l.], 31 mai. 2022. Disponível em: <https://www.cnnbrasil.com.br/tecnologia/justica-dos-eua-bloqueia-lei-do-texas-que-restringe-empresas-de-redes-sociais/>. Acesso em: 01 jun. 2022. No Brasil, raciocínio similar embasou a elaboração da Medida Provisória nº 1068/2021. Este caso será aprofundado no próximo capítulo.

²⁰ O aprofundamento desta distinção pode ser feito a partir da leitura de obra célebre do Professor Owen Fiss (*A Ironia da Liberdade de Expressão: Estado, Regulação e Diversidade na Esfera Pública*. Trad. Gustavo Binenbojm e Caio Mário da Silva Pereira Neto. Rio de Janeiro: Renovar, 2005, p. 33-67). No Brasil, tal distinção é trazida em relevante artigo do professor Daniel Sarmento, no qual sustenta a possibilidade de regulação das mídias de massa, de modo a conferir maior pluralidade à esfera pública (SARMENTO, Daniel. Liberdade de expressão, pluralismo e o papel promocional do Estado. *Revista Diálogo Jurídico*, [s.l.], n. 16 mai./ago. 2007, p. 2-39). No mesmo sentido, conferindo ênfase à questão da regulação das mídias de massa, cf. Taveira (*Democracia e Pluralismo na esfera comunicativa: uma proposta de reformulação do papel do Estado na garantia da Liberdade de Expressão*. 2010. Tese (Doutorado em Direito) – Faculdade de Direito, Universidade do Rio de Janeiro, Rio de Janeiro, 2010).

expressão demanda um ônus argumentativo diferenciado, pois defender uma atuação proativa estatal na esfera discursiva é comumente associado com governos autoritários.

De outro, é preciso compreender como o poder estatal se relaciona com as redes sociais e como tal ponto se entrecruza com a temática da liberdade de expressão, o que requer melhor conhecimento do seu funcionamento. Neste capítulo inicial aprofunda-se o primeiro aspecto: definir as virtudes e riscos da intervenção legislativa estatal que opere sobre canais de comunicação privados essenciais para a construção da opinião pública.

Do ponto de vista histórico, é imperioso destacar que o reconhecimento do Estado como agente legitimado à imposição de limites à liberdade discursiva consolida-se ao longo do século XX, especialmente após um movimento para universalização deste direito fundamental, que o eleva ao *status* de um direito humano. Inúmeros tratados internacionais contêm cláusulas que garantem a ampla proteção da liberdade de expressão, considerando-a um direito universal, mas permitem seu condicionamento em nome de outros interesses relevantes (tais como saúde, segurança e ordem pública e, ainda, para possibilitar a proteção de grupos alvos de discriminação). Destacam-se como principais instrumentos neste sentido: o Pacto Internacional de Direitos Civis e Políticos — PIDCP (art. 19)²¹; a Declaração Universal de Direitos Humanos — DUDH (art. 19)²²; a Convenção Americana Sobre Direitos Humanos — CADH (art. 13)²³; e, em especial, a Convenção Internacional sobre a Eliminação de todas as formas de Discriminação Racial — CIETFDR (art. 4º)²⁴.

No âmbito nacional, mesmo com a herança maldita legada pelo histórico de censuras²⁵ e perseguições a jornalistas (especialmente durante o período da ditadura militar), foi atribuído ao poder estatal algumas possibilidades restritivas em matéria de intervenção sobre a liberdade de expressão.

No art. 5º, são múltiplas as passagens sobre o tema. Há enunciação dos direitos fundamentais à “liberdade de pensamento”, no inciso IV; à “liberdade, intelectual, artística e científica, independentemente de censura e licença”, no inciso IX; e à “liberdade de informação,

²¹ Incorporado ao ordenamento jurídico nacional pelo Decreto nº 592, de 06 de julho 1992.

²² Ratificada pelo Brasil em 10/12/1948, na Assembleia Geral das Nações Unidas. Por se tratar de um tratado de *soft law* (de observância não obrigatória), não passou por processo de ratificação interna.

²³ Incorporado pelo Decreto nº 678, de 06 de novembro de 1992.

²⁴ O tratado, incorporado pelo Decreto nº 65.810/1969, prevê em seu art. 4º limitações ao exercício da liberdade de expressão ao condenar “toda forma de propaganda e todas as organizações que se inspirem em ideias ou teorias baseadas na superioridade de uma raça ou de grupos de pessoas de uma certa cor ou de uma certa origem étnica (...)”.

²⁵ Luís Roberto Barroso relembra que a história da censura no Brasil remonta ao período do seu descobrimento, com censura de trechos da carta de Pero Vaz de Caminha, considerada a certidão de nascimento do que viria a se chamar Brasil, pelo Padre Manuel Aires do Casal (BARROSO, Luís Roberto. *Sem data venia: um olhar sobre o Brasil e o Mundo*. [S.l.]: História Real, 2020, p. 111).

resguardado o sigilo da fonte”, no inciso XIV, todos imperativos do afastamento estatal da esfera de pensamento, manifestação e informação. Nada obstante, no mesmo art. 5º, há prescrições que condenam o racismo e qualquer ato discriminatório atentatório aos direitos fundamentais (art. 5º, LI e LII).

Em um segundo momento, no capítulo da “Da comunicação social”, a Constituição prescreve de forma mais clara parte dos direitos já mencionados no art. 5º, enunciando a manifestação de pensamento, criação, expressão e, em especial, o direito à informação livre de qualquer restrição (art. 220, *caput*), mas condicionando a liberdade de informação jornalística aos limites estabelecidos nos incisos IV, V, X (proteção aos direitos da personalidade), XIII e XIV (qualificação e sigilo profissional) do art. 5º da Constituição da República. Sem prejuízo da vedação em absoluto à censura de natureza política e ideológica (art. 220, § 2º), há ainda autorização para regulação de diversões, espetáculos públicos e programas de rádio e televisão (informações não jornalísticas), em prol de outros interesses públicos como a proteção à família, saúde pública e promoção de valores nacionais e regionais (arts. 220, § 3º e 221).

Veja-se, portanto, que a despeito da especial ênfase que a Carta Magna dá ao direito da liberdade de expressão em suas variadas acepções (liberdade de se expressar, criar e informar), com múltiplas menções ao longo de seu texto, atribui-lhe alguns limites expressos e, ainda, prevê a possibilidade de o Estado legislar — em alguma medida — sobre a programação em mídias de massa, não abordando uma concepção substancial necessariamente neutra como preconizada pelos libertários.

Tais limites não são os únicos. Embora a evolução jurisprudencial no trato da liberdade de expressão nem sempre tenha seguido um caminho retilíneo e nitidamente coerente, há múltiplos precedentes do Supremo Tribunal Federal que impõem outras limitações ao exercício desta liberdade em caso de conflito com outros interesses fundamentais. É possível encontrar precedentes mais preocupados com o “ouvinte”, ao criminalizar o discurso daquele que ataca ou deprecia grupos historicamente perseguidos;²⁶ há orientações que embora reconheçam primazia *prima facie* da liberdade de expressão quando relacionada à divulgação de informações de interesse público²⁷, certificam o critério da “veracidade subjetiva” como limite

²⁶ Dentre outros precedentes, pode se referenciar o multicitado caso Ellwanger. Julgamento paradigmático sobre o tema que assentou a constitucionalidade do art. 20 da Lei nº 7716/89 e confirmou a prática de crime de racismo por sujeito que editou, distribuiu e vendeu ao público obras de conteúdo antissemita (HC 82.424/RS, Ministro Relator Maurício Corrêa). Recentemente, o Supremo Tribunal Federal categorizou a homofobia e transfobia como crime abrangido pela Lei de Racismo (Lei nº 7716/89), ante a omissão do Congresso Nacional em legislar sobre o tema (*vide* ADO 26, Min. Rel. Celso de Mello e MI 4733/DF, Min. Rel. Edson Fachin).

²⁷ O aspecto preferencial da liberdade de expressão fora assentado, originariamente, na ADPF nº 130/DF, que aboliu a lei da censura.

interno ao direito de informar²⁸; passando, ainda, por precedentes significativamente restritivos que vislumbram a possibilidade de se flexibilizar imunidades discursivas de parlamentares quando utilizadas para incitação ao ataque a outras autoridades públicas e instituições.²⁹

A imposição de condicionantes, legislativas ou judiciais, comungam de uma mesma premissa que é a tentativa de articular suas razões mais ou menos restritas sobre os limites à liberdade de expressão, sempre a partir de uma visão sobre os fundamentos que justificam a proteção a este direito.

Assim, da mesma forma que os Poderes Legislativo e Judiciário devem procurar nos motivos justificadores da natureza fundamental da liberdade de expressão os argumentos para suas decisões, uma pesquisa sobre os limites da ação legislativa estatal sobre redes sociais também depende uma investigação preliminar acerca dos fins que fundam a proteção à liberdade de expressão.

Boa parte destes fundamentos são tradicionalmente apresentados em obras sobre o tema à luz de uma visão estritamente subjetiva, com enfoque quase que exclusivo no direito do indivíduo de ter sua autonomia protegida de qualquer intervenção estatal. O próximo tópico busca rerepresentar tais temas não somente sob esta ótica, mas também dando ênfase a como o agir do Estado pode ser importante para o cumprimento dos objetivos idealizados de quando se preconizou a natureza de direito fundamental à liberdade de expressão.

²⁸ Nesta esteira, afirma Barroso “A informação não pode prescindir da verdade – ainda que uma verdade subjetiva e apenas possível (..) pela circunstância de que é isso que as pessoas legitimamente supõem estar conhecendo ao busca-la” (BARROSO, Luís Roberto. Colisão entre Liberdade de expressão e Direitos da Personalidade: critérios de ponderação. Interpretação constitucionalmente adequada do Código Civil e da Lei de Imprensa. *Revista de Direito Administrativo*, Rio de Janeiro, v. 235, p. 1-36, jan./mar. 2014, p. 18).

²⁹ Neste sentido caminhou o Supremo Tribunal Federal na ADPF 572/DF, que julgou a constitucionalidade de inquérito inaugurado pelo Supremo Tribunal Federal para apurar ameaças à vida e integridade de Ministros do Supremo Tribunal Federal. A tese subjacente às decisões é a da “democracia militante”, isto é, que o limite para o exercício das liberdades discursivas, especialmente de agentes políticos, são ataques e ameaças às instituições (e seus membros) responsáveis pela salvaguarda dos direitos fundamentais de minorias. Ainda que sem menção direta à noção de “democracia militante”, entendimento similar foi reafirmado em denúncia oferecida pela Procuradoria Geral da República em face de parlamentar que capitaneou uma série de ameaças e incitações de ataque aos ministros do Supremo Tribunal Federal. Na hipótese, o Supremo Tribunal Federal considerou a norma da imunidade parlamentar inaplicável ao caso, condenando o réu à prisão (AP 1044/DF, Min. Rel. Alexandre de Moraes, Plenário, julgado em 20/04/2022).

1.2 As razões para a proteção da liberdade de expressão

As razões justificadoras da proteção à liberdade de expressão é ponto abordado em diversos títulos e obras sobre o tema, podendo-se invocar quase uma dezena de motivos para tanto³⁰.

Para os fins desta obra, cumpre indicar apenas os três principais, quais sejam: a) a obtenção da “verdade”, como meio viabilizador do progresso (o “livre mercado de ideias”); b) a preservação do regime democrático; e c) a proteção à autonomia individual. Faz-se este recorte por duas razões.

Primeira, pois os demais fundamentos usualmente elencados (promoção de direitos fundamentais, proteção do patrimônio cultural e científico, maior estabilidade, redução de violência etc.) parecem mera decorrência lógica das três primeiras causas, o que torna um tanto quanto repetitivo descer em maiores minúcias sobre seus aspectos.

Segunda, pois tais elementos já são suficientes para antecipar uma correlação inicial com os novos desafios suscitados pelo poder de influência das redes sociais sobre a esfera pública.

1.2.1 A metáfora do “livre mercado de ideias”

A liberdade de expressão é um direito fundamental, pois é a melhor forma de se alcançar a verdade, possibilitando a constante evolução e progresso social. Mesmo o discurso equivocado, banal, ultrajante e falso deve persistir na esfera pública, livre de qualquer censura estatal, sendo o próprio discurso (*counterspeech*) o melhor remédio para combatê-lo³¹.

³⁰ A título de exemplo, Aline Osório descreve cinco razões possíveis para a proteção à liberdade de expressão, conferindo ênfase, no entanto, às três que apresentamos neste capítulo (OSÓRIO, Aline. *Liberdade de expressão e Direito Eleitoral*. Belo Horizonte: Fórum, 2017, p. 53-97). Cláudio Chequer indica quatro elementos justificadores da proteção à liberdade de expressão: proteção à autonomia, garantia da democracia, avanço do conhecimento e manutenção da balança entre estabilidade e a mudança da sociedade (CHEQUER, Cláudio. *A Liberdade de Expressão como Direito Preferencial Prima Facie: análise crítica de revisão ao padrão jurisprudencial brasileiro*. 2. ed. Rio de Janeiro: Lumen Juris, 2017, p. 17-34). André Gustavo Corrêa de Andrade, por sua vez, chega a elencar oito razões como justificativa para a relevância axiológica deste direito (ANDRADE, André. *Liberdade de expressão em tempos de cólera*. Rio de Janeiro: GZ, 2020, p. 60-80).

³¹ A doutrina do *counterspeech* surge nos Estados Unidos em 1926 como consequência natural da tese do livre mercado de ideias. Nas palavras do *Justice* Louis Brandeis em *Whitney v. California*: “[i]f there be time to expose

O desenvolvimento dessa tese ocorre ao longo de três momentos-chave que culminam com a síntese da tese do “livre mercado de ideias” (*marketplace of ideas*), usualmente evocado para sustentar um dever de neutralidade do governo, quando da regulamentação da liberdade de expressão.

O fundamento da relevância do discurso livre como mecanismo promotor do progresso é originariamente apresentado no discurso “Aeropagítica”, de *Sir John Milton* para o Parlamento Inglês em 1644, em um contexto de alta tensão teológica marcada por tentativas de censura pela Igreja sobre a produção literária, cultural e científica na Europa.³² Para Milton, o mundo seria constituído de uma batalha espiritual eterna entre a verdade e o erro. A batalha não demandaria o uso da força, mas sim um confronto livre e aberto de ideias. Somente a partir do confronto entre o erro e a virtude é que se alcançaria a verdade³³.

Os argumentos trazidos por Milton são refinados pelo utilitarista John Stuart Mill em sua obra célebre “*On Liberty*”, escrita em 1859, tendo como principal alvo também a Igreja. Na visão do autor, a liberdade de expressão é fundamental para a ampla circulação de ideias e opiniões, fornecendo não só o instrumento mais adequado para se alcançar a verdade, como também por permitir o desenvolvimento da sociedade a partir da compatibilização e acomodação de ideias distintas. Com efeito, a própria verdade perderia com o silenciamento da opinião dissidente.³⁴

Ao contrário de Milton, contudo, Mill não vê a verdade como resultado necessário do livre debate, mas advoga pela excepcionalidade da restrição à livre circulação de ideias, já que a censura suprime a mera possibilidade de uma ideia verdadeira vir a ser apresentada ao debate público. Há um risco concreto — e quase sempre ignorado — de falibilidade dos mecanismos censuradores.³⁵ Assim, apenas em circunstâncias em que um dano ao indivíduo ou à sociedade seja claro e manifesto (o famoso *princípio do dano*) é que se justifica a imposição de alguma

through discussion the falsehood and fallacies, to avert the evil by the processes of education, the remedy to be applied is more speech, not enforced silence (em tradução livre: “se há tempo para expor por meio do debate falsidades e falácias, para reverter o mal por meio de processos de educação, o remédio a ser aplicado é mais discurso, não o silêncio forçado”). 274 U.S. 357, 377 (1927).

³² A crítica de Milton era direcionada a um mecanismo de licença prévia estabelecido pelo Parlamento Inglês.

³³ MILTON, John. *Aeropagítica. Discurso sobre a Liberdade de Expressão*. Trad. Benedita Bettencourt. Coimbra: Almedina, 2009.

³⁴ MILL, John Stuart. *Sobre a Liberdade*. Trad. Denise Bottmann. Porto Alegre: L&PM, 2019, p. 81.

³⁵ Nos dizeres do autor: “todo silenciamento de um debate é uma pretensão de infalibilidade (...) embora todos saibam muito bem que são falíveis, poucos julgam necessário se precaver contra a sua própria falibilidade ou admitem supor que alguma opinião sua, da qual se sentem muito certos, pode constituir um exemplo do erro a que se reconhecem sujeitos” (Ibid., p. 125).

restrição à sua liberdade.³⁶ Mesmo o dano meramente provável não deve justificar a restrição ao exercício da autonomia individual pelo Estado, pois se presume que não há ninguém mais apto a conduzir qualquer assunto de forma adequada e racional do que o próprio cidadão.³⁷

No início do Século XX, a obra de Mill é revisitada por Oliver Wendell Holmes, que a complementa com a metáfora do “livre mercado de ideias” para fundamentar a proteção diferenciada a este direito. Ao associar a ideia de liberdade de expressão com a de livre mercado, criada por Adam Smith, o juiz argumenta que, assim como na economia de mercado os consumidores são beneficiados pela livre concorrência entre produtos e serviços, a ampla e descentralizada troca de ideias faria triunfar a verdade no debate público. O melhor teste para a verdade seria submetê-la, portanto, ao teste de mercado, possibilitando sua livre circulação.³⁸

Parece lógico que a livre circulação de ideias é essencial para um contínuo escrutínio sobre a ideia de “verdade” — compreendida aqui como uma proposição validada pelos testes da experiência³⁹ —, possibilitando assim uma contínua evolução da sociedade guiada pela razão e, por certo, isso é um argumento que milita a favor da natureza fundamental deste direito, justificando uma avaliação mais cética sobre a imposição de limitações ao seu exercício pelo Poder Público.

O ponto que suscita dúvidas, no entanto, é que se extraia da premissa de seu caráter fundamental que o Estado seja necessariamente um obstáculo à liberdade de expressão. À luz da concepção libertária, o poder regulamentar estatal seria algo inerentemente ruim, uma vez

³⁶ Essa conclusão pode ser extraída de diversas passagens da obra: “Ninguém pretende que as ações sejam livres como opiniões. Pelo contrário, até as opiniões perdem sua imunidade quando são expressas em circunstâncias tais que convertem essa sua expressão em franca instigação a alguma ação perniciososa”. Em outra passagem acrescenta: “Ninguém deve ser punido por estar simplesmente embriagado; mas cabe punição ao soldado ou policial que estiver embriagado no cumprimento de um dever. Em suma, quando há claro prejuízo ou risco de claro prejuízo, seja ao indivíduo ou ao público, o caso sai do campo da liberdade e ingressa no campo da moralidade ou da lei” (MILL, John Stuart. *Sobre a Liberdade*. Trad. Denise Bottmann. Porto Alegre: L&PM, 2019, p. 87 e 125, respectivamente).

³⁷ *Ibid.*, p. 164.

³⁸ O fundamento consta do voto dissidente do *Justice* Oliver Wendell Holmes em julgamento ocorrido em 1919 (SCOTUS, *Abraham vs United States*, 250 U.S 616 (1919)). Em 1918, os Estados Unidos participaram de uma operação militar em solo russo contra a Alemanha depois que a Revolução Russa derrubou o regime czarista. Imigrantes russos nos EUA circularam cartazes pedindo uma greve geral nas fábricas de munição para minar o esforço de guerra dos EUA. Os réus foram condenados por dois panfletos jogados de uma janela da cidade de Nova York. Um denunciou o envio de tropas americanas para a Rússia, e o segundo denunciou a guerra e defendeu a cessação da produção de armas para serem usadas contra os “soviéticos operários da Rússia”. Eles foram condenados a 20 anos de prisão.

³⁹ Mesmo que a definição de “verdade” seja um problema filosófico ainda sem solução, a influência do raciocínio de Holmes é claramente pragmatista. Neste sentido, afirma Dewey — um notório pragmatista — que toda assertiva que a proposição faz de sua própria verdade é apenas condicional e deve ser submetida a contínuos processos de revisão e investigação empírica (*The Problem of Truth*. In: *The Essential Dewey*, p. 113-114, v. 2 apud POGREBINSCHI, Thamy. *Pragmatismo: Teoria Social e Política*. Rio de Janeiro: Relume Dumará, 2005, p. 46).

que a “mão invisível” do mercado levaria naturalmente à circulação e à vitória das melhores ideias, como pressuposto pelo *Justice Holmes*.

Três críticas podem ser feitas a esta conclusão.

A primeira se direciona à premissa de racionalidade subjacente à tese tanto de Holmes quanto de Mill ou Milton de que, em regra, todo indivíduo terá condições de discernir em concreto aquilo que é mais benéfico aos seus interesses⁴⁰ à medida que o debate na esfera pública avança e amadurece à luz das razões apresentadas ao longo do tempo.

Esta premissa teórica é refutada por estudos mais recentes das ciências comportamentais que indicam o sistema cognitivo humano como mais complexo do que se imagina, não produzindo necessariamente resultados mais racionais.

Em resumo, o sistema cognitivo humano pode ser explicado por meio de dois sistemas. O primeiro — denominado de sistema 1 — mais intuitivo, rápido, associativo e suscetível às emoções e ao ambiente externo; e um sistema 2, mais lento e reflexivo, controlado e autoconsciente.

A esse sistema 2 cabe monitorar e controlar pensamentos e ações “sugeridos” pelo sistema 1, responsável pela construção das “primeiras impressões”, mas este autocontrole, por demandar energia, é uma tarefa muitas vezes exaustiva e desagradável. Embora a interação entre estes sistemas seja, em geral, uma adaptação evolutiva⁴¹, o risco de falhas cognitivas decorre, justamente, do conforto propiciado pela criação de atalhos mentais intuitivos para a tomada de decisões e realização de julgamentos.⁴² Estas falhas, conhecidas como “vieses cognitivos” se manifestam, muitas vezes imperceptivelmente, como resultado do mau uso dos atalhos mentais produzidos pelo sistema 1, que levam a julgamentos que se afastam, de forma sistemática, daqueles que seriam o melhor ao caso.

Tal observação é relevante, porque a lógica do livre mercado de ideias só prospera quando os participantes do debate estão dispostos a informar e aprender (valendo-se do sistema 2). O que os defensores do “livre mercado de ideias” ignoram, no entanto, é que a conversação

⁴⁰ O conceito de interesse tem como parâmetro a ideia de homem racional adotada pela teoria econômica clássica, que aponta como propósito da ação humana a maximização da satisfação individual a partir de uma análise de custo-benefício. (v. POSNER, Richard. *The Economic Approach to Law. Chicago Unbound*, Chicago: University of Chicago Law School, v. 53, 1975, p. 757-782).

⁴¹ Tal afirmação foi a conclusão de experiências levadas a cabo pelo neurocientista Antônio Damásio com pacientes com danos específicos em uma parte específica do cérebro — o ventromedial cortex pré-frontal (uma região logo atrás e acima do nariz) —, que levava a redução da emotividade e intuição a quase zero. Pacientes que sofreram estes danos pontuavam muito bem em testes de coeficiente de inteligências, mas tinham extrema dificuldade para a tomada de decisões individuais quando diante de múltiplas opções. Sem as emoções, uma opção era tão boa quanto qualquer outra (v. DAMASIO, Antônio R. *O erro de Descartes: Emoção, Razão e o Cérebro Humano*. São Paulo: Companhia das Letras, 2020).

⁴² KAHNEMAN, Daniel. *Rápido e Devagar: Duas Formas de Pensar*. São Paulo: Objetiva, 2018, p. 52-66.

pura e simples não é usada somente para este fim, mas também para bloquear perspectivas, intensificar vieses e aumentar preconceitos. Nos dizeres de Jason Stanley, “o argumento do modelo do ‘mercado aberto de ideias’ para a liberdade de expressão funciona somente se a disposição subjacente à sociedade é aceitar a força da razão sobre o poder de ressentimentos irracionais e do preconceito”⁴³. Do contrário, o agir comunicativo pode ser direcionado à exploração de vieses, cujos efeitos (como se verá adiante) podem ser drásticos.

A tese de Mill, reforçada por Holmes, precisa ser também devidamente contextualizada. Tratava-se de um momento em que vários pensadores foram perseguidos por suas opiniões conflitantes com a opinião dominante, espelho daquilo que a religião cristã afirmava. O pluralismo de ideias ainda estava em processo de formação, portanto, sua teoria buscava convencer uma sociedade, até então homogênea, a se abrir a novos espaços. Em síntese, buscava-se combater a falta de informação.⁴⁴

Com o advento dos meios de comunicação em massa (televisão e rádio, em especial) e, ainda mais, com a ascensão da *internet*, transitamos de um modelo no qual a informação, antes escassa, é agora produzida em excesso, disseminando-se em segundos por todos os cantos do mundo. A noção estrita de claro e manifesto risco de dano a que alude Mill e, posteriormente, Holmes precisa ser, portanto, revisitada à luz dos novos tempos em que mais do que a escassez de informação, o que preocupa são excessos e distorções no exercício da liberdade de expressão.

Os exemplos de restrição mencionado pelos autores limitam-se, em geral, à imediatividade de um dano físico. Na visão de Holmes, mesmo a mais rigorosa proteção à liberdade de expressão “não protegeria um homem que falsamente grita “fogo” em um teatro causando pânico”⁴⁵. Do mesmo modo, Mill admite a possibilidade de se restringir o discurso, quando este for exposto “sem justa causa, perante uma turba exaltada”.⁴⁶ No caso dos exemplos, a iminência dos danos provocados pelo discurso — um falso e outro de incitação — é decorrência da proximidade física entre aquele que discursa e a plateia ouvinte. Por certo, que hoje tal proximidade não é mais necessária. Uma mentira ou um ataque por meio da *internet* pode viralizar em segundos, provocando pânico e agitando turbas exaltadas em diversos cantos do mundo. A *internet* confere, em tese, a todos esta capacidade de, a um custo muito baixo e em poucos segundos, produzir danos a um número gigantesco de pessoas por meio de um

⁴³ STANLEY, Jason. *Como Funciona o Fascismo: a política do “Nós” e “Eles”*. 2. ed. Porto Alegre: L&PM, 2019, p. 75.

⁴⁴ V. FADEL, Anna Laura Maneschy. *O discurso de ódio é um limite legítimo ao exercício da liberdade de expressão?* Uma análise das teorias de Ronald Dworkin e Jeremy Waldron a partir da herança do liberalismo de John Stuart Mill. Rio de Janeiro: Lumen Juris, 2018, p. 21.

⁴⁵ Schenk v. United States, 249 US 47,52 (1919).

⁴⁶ MILL, John Stuart. *Sobre a Liberdade*. Trad. Denise Bottmann. Porto Alegre: L&PM, 2019, p. 88.

discurso falso ou de incitação⁴⁷. Ante a maior possibilidade de o discurso produzir danos na esfera virtual, não é irrazoável reconhecer que talvez seja necessário repensar os paradigmas que orientam uma noção muito estrita de dano para fins de justificação da regulação discursiva.

A terceira e última crítica se refere ao uso da metáfora de “livre mercado”. Trata-se de alegação cuja existência empírica não é verificável. A ideia de um mercado perfeito e infalível, cuja “mão invisível” produziria os melhores resultados, não se verifica na realidade historicamente observável, funcionando, portanto, mais como “um ideal de eficiência a ser perseguido por mecanismos de correção de falha de mercado”⁴⁸ do que um fato empírico.

Não é diferente com a liberdade de expressão. O que se observa é que a esfera discursiva autorregulada não necessariamente leva à prevalência das melhores ideias. Ao revés, conduz apenas ao trunfo da ideia mais procurada pelo público, o que muitas vezes inviabiliza o ciclo de progresso e desenvolvimento idealizado por Mill e Holmes. Basta observar que uma mentira tem, em média, 70% mais chances de ser compartilhada e leva seis vezes menos tempo para alcançar seus leitores.⁴⁹ Por sua vez, mensagens negativas são cinco vezes mais eficazes em sua influência do que as positivas.⁵⁰

Não se trata, portanto, de uma competição livre de ideias nos quais as melhores prevalecerão. Vieses cognitivos e discriminatórios podem prejudicar a compreensão e aceitação das melhores ideias, trazendo distorções significativas para o debate público.

Com efeito, há apenas uma interpretação possível a ser conferida à tese do “livre mercado de ideias”. Tal como a lógica do livre mercado é útil como um ideal de eficiência para o exercício do poder ordenador estatal, a metáfora de um livre mercado de ideias pode ser utilizada como referência para regulações que busquem estabelecer uma condição social ideal, na qual todos tenham igual oportunidade de fala e as ideias fluam e compitam de forma genuína em uma esfera pública robusta⁵¹, possibilitando a prevalência não só da ideia mais popular,

⁴⁷ É neste sentido que Cass Sunstein afirma que hoje temos inúmeras pessoas gritando fogo em um teatro lotado. Por essa razão seria possível considerar declarações falsas como não abarcadas pela garantia da liberdade de expressão se o governo conseguir comprovar que elas representam um risco de sério prejuízo, que não pode ser evitado por meio de uma rota mais protetiva ao discurso (SUNSTEIN, Cass R. *Liars: Falsehoods and Free Speech in Age of Deception*. Nova Iorque: Oxford University Press, 2021, p. 2-4).

⁴⁸ BINENBOJM, Gustavo. *Poder de Polícia, Ordenação e Regulação: Transformações Político-Jurídicas, Econômicas e Institucionais do Direito Administrativo Ordenador*. 3. ed. Belo Horizonte: Fórum, 2020, p. 166.

⁴⁹ Segundo estudo publicado no jornal científico Science, em pesquisa realizada no Twitter (ARAL, Sinan; ROY, Deb; VOSOUGH, Soroush. The spread of true and false news online. *Science*, [s.l.], v. 359, n. 6380, p. 1146-1151, mar. 2028. Disponível em: <https://www.science.org/doi/full/10.1126/science.aap9559>. Acesso em 04 fev. 2022).

⁵⁰ CASTELLS, Manuel. *A ruptura: a crise da democracia liberal*. Rio de Janeiro: Zahar, 2018, p. 27.

⁵¹ MITCHELL-BRADDON, David; WEST, Caroline. What is free speech? *The Journal of Political Philosophy*, [s.l.], v. 12, n. 4, p. 437-460, 2004.

apelativa ou com menor custo, mas também aquela mais bem fundamentada e aderente aos fatos observáveis.

Quando antigos e novos limites ao exercício da liberdade de expressão estiverem em debate, tais como o discurso de ódio ou a desinformação⁵², é sempre interessante questionar à luz destes parâmetros em que medidas estes discursos contribuem ou implicam algum prejuízo verificável ao atendimento dessa condição ideal.

1.2.2 A proteção da democracia

A vinculação da liberdade de expressão com a democracia decorre da percepção de que sem liberdade de expressão não é possível conceber um projeto de autogoverno coletivo. Todo processo democrático de decisões pressupõe, necessariamente, a formação de uma opinião pública acerca de debates públicos.⁵³

Nesta linha, Alexander Meiklejohn sustenta que a liberdade de expressão é relevante, pois possibilita ao cidadão comunicar e transmitir ideias sobre o bem comum de forma informada e inteligente. Para o autor, o mais significativo é a proteção da informação que seja relevante para a tomada de decisões no projeto de autogoverno democrático, importando mais aquilo que chega ao conhecimento dos ouvintes do que as palavras dos emissores, a ponto de afirmar que o “essencial não é que todos devam falar, mas que tudo que valha a pena seja dito”.⁵⁴ A tese é passível de crítica, visto que parece atribuir ao Estado o papel de arbitrar o que é digno ou não de ser objeto de discurso, ampliando o risco de censura.

Robert Post também conecta a relevância da liberdade de expressão à ideia de construção da democracia, trazendo, no entanto, um argumento distinto. Para Post, o objetivo da liberdade de expressão não é assegurar a circulação das ideias mais relevantes, mas sim garantir o “valor da autoria”⁵⁵, que é da essência da democracia. Na visão do professor, a democracia só é alcançada quando “aqueles que estão sujeitos à lei, enquanto destinatários,

⁵² Há artigos, por exemplo, que indicam o “discurso de ódio” como uma categoria discursiva particularmente resistente ao *counterspeech*. Cf. Richard Delgado e David Yun (The Speech We Hate: First Amendment Totalism, the ACLU, and the Principle of Dialogic Politics. *Alabama Law*, Tuscaloosa: University of Alabama, v. 27, p 1281-1300, 1995).

⁵³ CHEQUER, Cláudio. *A Liberdade de Expressão como Direito Preferencial Prima Facie: análise crítica de revisão ao padrão jurisprudencial brasileiro*. 2. ed. Rio de Janeiro: Lumen Juris, 2017, p. 29.

⁵⁴ MEIKLEJONH, Alexander. *Free speech and its relation to self-government*. Nova Iorque: Harper & Brothers Publishers, 1948, p. 25.

⁵⁵ POST, Robert. Participatory Democracy and Free Speech. *Virginia Law Review*, [s.l.], v. 97, n. 3, 2011, p. 482.

acreditam que também possam ser autores da lei”⁵⁶, e isso só é possível por meio de um processo de comunicação na esfera pública que permita a todos se tornarem potenciais construtores da opinião pública. A liberdade de expressão, portanto, é um direito fundamental que visa resguardar os atos discursivos e os meios de comunicação de mídia, que são os meios apropriados e necessários para a participação na formação da opinião pública.⁵⁷

A defesa da liberdade de expressão como direito fundamental passa, portanto, pelo aprofundamento do conceito de “esfera pública”, o local onde a opinião pública se molda, que traduz o ponto mais importante para definição da dimensão tanto defensiva quanto protetiva da liberdade de expressão com base no argumento democrático.

Segundo Habermas, o conceito de esfera pública se traduz por um espaço em que pessoas privadas se reúnem para debater, entre si e mediante razões, uma ampla gama de questões de interesse geral da sociedade.⁵⁸ É uma espécie de determinada cultura em que meios de comunicação, tais como livros, jornais, periódicos, rádios, televisão e — agora — redes sociais, desempenham uma função de agir comunicativo.⁵⁹ Trata-se de um *locus* em que se discutem questões fora do ambiente político, por meio da troca de argumentos racionais⁶⁰.

A função normativa deste espaço intermediário entre sociedade e Estado é, portanto, fornecer um dos pilares da legitimação para o exercício do poder em um regime democrático: a troca de razões entre pessoas livres e iguais.

Este é o fundamento mais sólido quando se pensa na ligação entre liberdade de expressão e democracia. Protege-se, especialmente, os atos discursivos necessários à construção de uma opinião pública, pois eles são um dos pilares da legitimação estatal em um regime de democrático. Se não há liberdade para a discussão de assuntos em geral, as funções políticas e críticas da esfera pública restam substancialmente prejudicadas.

Por outro lado, cabe ao Estado também garantir as circunstâncias ideais necessárias para que esta esfera pública possa cumprir suas funções normativas, quais sejam a construção de um local ideal de fala, que assegure aos cidadãos condições razoáveis para manifestar-se criticamente, mediante um procedimento que cumpra de forma ao menos aproximada os

⁵⁶ POST, Robert. Participatory Democracy and Free Speech. *Virginia Law Review*, [s.l.], v. 97, n. 3, 2011, p. 482.

⁵⁷ *Ibid.*, p. 483.

⁵⁸ HABERMAS, Jürgen. *Mudança estrutural da esfera pública: investigações sobre uma categoria da sociedade burguesa*. Trad. Denilson Luís Werle. São Paulo: Editora Unesp, 2014, p. 21-22.

⁵⁹ VESTING, Thomas. A Mudança da Esfera Pública pela Inteligência Artificial. Trad. Pedro Henrique Ribeiro. In: ABOUD, Georges; CAMPOS, Ricardo; NERY JR., Nelson (Orgs.). *Fake News e Regulação*. 2. ed. São Paulo: Revista dos Tribunais, 2020. p. 193-212.

⁶⁰ CAVALLAZI, Vanessa Wendhausen. *E-democracia deliberativa: a criação de espaços de deliberação em rede para a implementação de direitos sociais*. São Paulo: Juspodium, 2020, p. 69.

requisitos de igualdade, universalidade, racionalidade, reflexividade e reciprocidade.⁶¹ A democracia, portanto, não se restringe à mera vontade da maioria, mas requer a construção e manutenção de condições comunicativas que impliquem tratar o outro como um espelho de si próprio, para que se alcance o melhor argumento.

Estas premissas podem trazer perplexidade se considerarmos que se analisa aqui a necessidade de uma regulação pública das redes sociais, eis que uma das principais críticas objetadas ao conceito de “esfera pública” de Habermas era sua incapacidade de lidar com o número significativo de grupos excluídos do debate público.⁶²

A dificuldade de acesso à esfera pública por grupos tradicionalmente excluídos do debate público foi um problema substancialmente mitigado pela *internet*, que concretiza uma ideia de sociedade aberta “sem distinções entre portas de entradas na rede”⁶³, que substitui, em parte, o tradicional modelo de comunicação de “um para todos” (típico das mídias de massas) para um modelo em que cada indivíduo se torna ao mesmo tempo receptor e transmissor da informação (de todos para todos).⁶⁴

A *internet*, por muito tempo, foi percebida como uma viabilizadora de uma *Ágora Digital*, visto que seria dado a todos o acesso em tempo real a inúmeras notícias e informações ao redor do mundo, com a possibilidade de comentar, influir e opinar sobre os mais diversos assuntos, eliminando as barreiras do mundo físico. Do ponto de vista político, reduz-se substancialmente os custos de comunicação do representante com o representado e abre-se oportunidade para uma infinidade de mecanismos de exercício de democracia direta, como, por exemplo, consultas públicas, petições eletrônicas e iniciativas de *crowdsourcing*.⁶⁵

Se a *internet* promove o empoderamento do indivíduo, reduzindo o poder econômico e político de grandes conglomerados de mídia, superando em definitivo o cenário de escassez informacional que predominou até a primeira metade do século XX, admitir a regulação pública das redes sociais — um dos principais difusores de informação na *internet* — parece desnecessário, além de implicar risco de prejuízo ao acesso à esfera pública.

⁶¹ CAVALLAZI, Vanessa Wendhausen. *E-democracia deliberativa: a criação de espaços de deliberação em rede para a implementação de direitos sociais*. São Paulo: Juspodium, 2020, p. 71.

⁶² Cf. FRASER, Nancy. Rethinking the public sphere: a contribution to the critique of actually existing democracy. *Social Text*, [s.l.], n. 25/26, 1990, p. 56-80.

⁶³ BALDI, Vania. A construção viral da realidade: ciberepopulismos e polarização dos públicos em rede. *Observatorio (OBS*)*, [s.l.], special issue, p. 4-20, 2018.

⁶⁴ Cf. CASTELLS, Manuel. *A Sociedade em rede*. São Paulo: Paz e Terra, 2013. v.1, p. 412.

⁶⁵ SUSSKIND, Jamie. Future Politics. *Living Together in a World Transformed by Tech*. Nova Iorque: Oxford University Press, 2018. pt. 4, p. 211-257.

O raciocínio é, em parte, correto. A *internet*, sem dúvidas, ampliou os espaços de comunicação especialmente para públicos marginalizados, que passaram a atingir e influenciar a esfera pública dominante pautando a mídia tradicional.

Isso não obsta reconhecer que as novas formas de comunicação propiciadas pela *internet* produziram também mazelas. A pretensa autonomia comunicativa nas redes sociais sofre forte influência de um procedimento de curadoria algorítmica pouco transparente, que pautado por um modelo econômico de atenção, elimina o dissenso, conferindo aos seus usuários acesso apenas àquilo que confirma suas pré-compreensões.

De um lado, tal procedimento acirra a polarização, transformando aquele que pensa diferente em inimigo. De outro, torna o conhecimento efêmero, pois mais importa a capacidade de disseminação (ou viralização) do discurso do que a profundidade ou aderência dos argumentos apresentados ao contexto fático subjacente⁶⁶. Como consequência, observa-se uma transição do conceito relacional de esfera pública para um agir comunicativo altamente fragmentado e narcísico, onde a interpretação do mundo parte sempre do “eu”.⁶⁷

O fundamento democrático para regulação da liberdade de expressão está, portanto, intimamente relacionado à manutenção da qualidade do debate público. Se antes, na formação e amadurecimento da esfera pública burguesa, demandava-se a regulação dos grandes conglomerados de mídias para assegurar a ampla participação no debate público de grupos de interesse com pontos de vista distintos, agora, na era da *internet*, “é preciso desenvolver nós sobre os quais permaneça possível ao *homo digitalis* uma observação mútua da subjetividade e a construção de um conhecimento comum”.⁶⁸

⁶⁶ Neste sentido, Gustavo Binenbojm e Caio Mário da Silva Pereira afirmam “(...) num mundo onde a informação perde credibilidade em meio à multiplicidade de fontes e à explosão de informações falsas, os indivíduos se fecham em grupos mais polarizados, a capacidade de conviver e ponderar os diferentes pontos de vistas vai se reduzindo e a esfera pública empobrece, afetando negativamente a capacidade de decisão coletiva”. Trecho do prefácio à segunda edição da obra do jurista norte-americano Owen Fiss, traduzida pelos autores acima mencionados (FISS, Owen. *A ironia da liberdade de expressão: Estado, regulação e diversidade na esfera pública*. Trad. Gustavo Binenbojm e Caio Mario da Silva Pereira Neto. São Paulo: Editora FGV, 2022, p. 29).

⁶⁷ A lógica hiperindividualista das redes sociais, em última instância, inviabiliza o compartilhamento de valores comuns, que são essenciais à construção de mitos como o da democracia. Como bem destaca Harari, a base da evolução do *homo sapiens* se deve, em grande parte, à evolução das nossas capacidades cognitivas linguísticas, que possibilitaram a capacidade de criar realidades imaginadas passíveis de serem compartilhadas entre muitos estranhos (os mitos). Nas palavras do autor, a “cooperação humana em grandes escalas é baseada em mitos”. É a “confiança mútua” nestes mitos, que viabiliza o progresso humano. Cf. Harari (*Homo Sapiens: uma breve história*. Trad. Janaína Marcoantonio. 25. ed. Belo Horizonte: L&PM, 2017, p. 28-48). Em síntese, sem uma base comum de premissas compartilhadas entre indivíduos não há confiança recíproca e a própria vida em sociedade se torna inviável.

⁶⁸ VESTING, Thomas. *A Mudança da Esfera Pública pela Inteligência Artificial*. Trad. Pedro Henrique Ribeiro. In: ABOUD, Georges; CAMPOS, Ricardo; NERY JR., Nelson (Orgs.). *Fake News e Regulação*. 2. ed. São Paulo: Revista dos Tribunais, 2020. p. 193-212.

1.2.3 A proteção da autonomia

A liberdade de expressão também está associada a um dos principais valores substanciais da ordem jurídica constitucional, pois protege a autonomia dos cidadãos, um dos elementos que abrange o princípio da dignidade da pessoa humana (art.1º, inciso III da CRFB).⁶⁹

A princípio, qualquer indivíduo tem o direito fundamental de se expressar e pronunciar suas ideias, opiniões e preferências sem qualquer interferência estatal. Nesta seara, o perfeccionismo e o moralismo não têm vez e manifestações de cunhos variados guardam especial proteção do ordenamento jurídico, cabendo ao Estado uma posição de isenção diante de múltiplas visões de mundo.

Em alguma medida, as razões para proteção com fundamento na autonomia se confundem com os argumentos já apresentados acerca do “livre mercado de ideias”. A diferença é que enquanto neste último se olha para o efeito social oriundo da concessão de uma ampla autonomia a todos os indivíduos (a viabilização do progresso), o enfoque aqui é no direito subjetivo do cidadão de exigir que o Estado não interfira no âmbito de sua vida privada, mais especificamente sobre a sua autonomia de se manifestar, comunicar ou informar. Há, portanto, uma faceta distinta na vinculação da liberdade individual à noção de autonomia, que confere ao direito uma proteção, ainda que a manifestação em si não promova um “bem maior” para sociedade, como a verdade ou a democracia.⁷⁰

De fato, não há como ignorar que esta afirmativa seja tomada como uma premissa quando se aborda a liberdade de expressão. O Estado não deve investir contra determinadas visões de mundo sob o argumento de uma suposta imoralidade ou necessidade de proteção à ordem pública. Assim, posições paternalistas do Estado, cujo objetivo precípua seja proteger o cidadão de si próprio, devem ser evitadas, sob pena de negar aos indivíduos a responsabilidade moral por suas ações e manifestações.⁷¹

⁶⁹ BARROSO, Luís Roberto. Aqui, lá e em todo lugar: A dignidade da pessoa humana no direito contemporâneo e no discurso transnacional. *Revista dos Tribunais*, São Paulo, v. 919, p. 127-196, 2012. Também vendo a autonomia como um elemento integrante do conteúdo mínimo da dignidade da pessoa humana, *vide* Sarmiento (*Dignidade da Pessoa Humana: conteúdo, trajetórias e metodologia*. [S.l.]: [s.n.], 2021. No prelo).

⁷⁰ OSÓRIO, Aline. *Liberdade de expressão e Direito Eleitoral*. Belo Horizonte: Fórum, 2017, p.64.

⁷¹ Cf. DWORKIN, Ronald. Why Speech must be free. In: _____. *Freedom's Law: the Moral Reading of the American Constitution*. Cambridge: Harvard University Press, 1996, p. 200.

Há, contudo, um ponto que, ainda hoje, suscita inúmeros debates quando se sustenta uma autonomia plena do ponto de vista do particular: os danos produzidos pelo denominado “discurso de ódio” (*hate speech*).

O termo discurso de ódio é, por vezes, invocado de forma genérica para definir uma série de discursos e situações de ataque a grupos muito distintos entre si, o que não impede a conjugação de alguns elementos comuns: o preconceito, a discriminação e a intolerância.⁷² Nesta esteira, André Andrade define o discurso odioso como a “manifestação ou expressão, motivada por preconceito ou intolerância, através do qual uma pessoa ou grupo é discriminado, com base em suas características identitárias”⁷³.

A questão que se põe é se autonomia individual é suficiente, por si, para justificar a manifestação de opiniões — vulgarmente referidas como “odiosas” — que estigmatizam grupos ou pessoas com base em aspectos específicos de sua identidade.

A definição da existência ou não desta limitação e sua extensão varia conforme o contexto histórico-cultural de cada país. Os Estados Unidos é, mais uma vez, o exemplo de país mais protetivo a esta espécie de discurso.⁷⁴

No Brasil, ainda que persistam debates sobre a extensão dos limites aos discursos de ódio, é inequívoco que não subsiste uma proteção à autonomia tal como amparada pela jurisprudência norte-americana. Isto se evidencia por passagens expressas do texto constitucional, dentre as quais cabe destacar o repúdio ao racismo, no art. 4º, inciso VIII, elevado à categoria de princípio reitor do Estado Brasileiro nas suas relações com nações vizinhas; o art. 5º, inciso XLI, que prevê a punição — nos termos da lei — a qualquer discriminação atentatória de direitos e liberdades fundamentais; e o art. 5º, XLII, que prevê o racismo como espécie de crime inafiançável e imprescritível nos termos da lei.

No âmbito legal, a Lei nº 7716/89 prevê, em seu art. 20, como crime “praticar, induzir ou incitar a discriminação ou preconceito de raça, cor, etnia, religião ou procedência

⁷² ANDRADE, André. *Liberdade de expressão em tempos de cólera*. Rio de Janeiro: GZ, 2020, p. 150.

⁷³ *Ibid.*, p. 151.

⁷⁴ Um caso emblemático desta proteção remonta à declaração de inconstitucionalidade das normas do Município de Skokie, cujo objetivo precípuo era evitar a realização de uma passeata organizada pelo Partido Nacional-Socialista da América, com pessoas utilizando uniformes e símbolos nazistas. Um detalhe do caso é fundamental: os moradores da cidade eram majoritariamente judeus, muito dos quais fugitivos do holocausto. Apesar da circunstância sensível, a norma foi declarada inconstitucional pela 7º *Court Circuit*, por violação à liberdade de expressão. Posteriormente, a Suprema Corte Norte-Americana recusou-se a reexaminar o mérito do caso. Cf. SCOTUS, *Smith vs. Collin*, 436 U.S. 953 (1978). No Brasil, comentários mais detalhados do julgado podem ser encontrados em artigo de Daniel Sarmiento (A liberdade de expressão e o problema do hate speech. *Revista de Direito do Estado*, Rio de Janeiro: Renovar, n. 4, p. 53-105, out./dez. 2006. Mimeografado).

nacional”⁷⁵. Ainda, o Estado Brasileiro incorporou ao seu ordenamento jurídico, através do Decreto nº 65.810/69, a Convenção Internacional sobre a Eliminação de todas as formas de Discriminação Racial — CIETFDR, que prevê, em seu art. 4º, o dever de seus signatários condenarem qualquer propaganda ou grupo que inspire teorias baseadas na superioridade de raça ou etnia.

No campo jurisprudencial, o Supremo Tribunal Federal consagrou, no caso *Ellwanger*, o conceito de raça como uma construção “político-social”⁷⁶, para defender a aplicabilidade do seu dispositivo à condenação de propagandas e atos antissemitas. Recentemente, a premissa de um conceito socialmente construído de raça foi utilizada pelo Supremo Tribunal Federal para estender a aplicação do art. 20 da Lei nº 7716/89 aos atos de homofobia e transfobia, até que sobrevenha legislação do Congresso Nacional sobre a matéria.⁷⁷

No campo doutrinário, há também fartos fundamentos que observam no discurso de ódio um limite legítimo ao exercício da liberdade de expressão. Segundo Jeremy Waldron, a dignidade humana pressupõe que seja atribuído a todos um *status* social que lhe atribua garantia de respeito pelos demais membros e igualdade. Neste sentido, sustenta ser dever do Estado criar legislações protetivas aos membros das minorias para lhes assegurar igual dignidade aos demais membros da sociedade, o que abarca leis restritivas do discurso de ódio.⁷⁸

Daniel Sarmiento elenca dois grandes motivos para a proibição do discurso odioso. O primeiro é o prejuízo à dignidade do ouvinte. O ataque preconceituoso é redutor da sua autonomia, que se vê diminuído em sua existência por ataques de terceiro. Gera, ainda, prejuízo à sua própria autoestima, mais precisamente por uma perda de possibilidade de entender a si próprio como um ser estimado por suas propriedades e capacidade características⁷⁹. Sustentar

⁷⁵ Com maior enfoque na honra subjetiva, poderia ser citado também o crime de injúria racial constante no art. 140, § 3º do Código Penal.

⁷⁶ Neste sentido, segue trecho de ementa do acórdão prolatado no bojo do HC nº 82.424/RS: (...) Raça humana. Subdivisão. Inexistência. Com a definição e o mapeamento do genoma humano, cientificamente não existem distinções entre os homens, seja pela segmentação da pele, formato dos olhos, altura, pêlos ou por quaisquer outras características físicas, visto que todos se qualificam como espécie humana. Não há diferenças biológicas entre os seres humanos. Na essência são todos iguais. 4. Raça e racismo. A divisão dos seres humanos em raças resulta de um processo de conteúdo meramente político-social. Desse pressuposto origina-se o racismo que, por sua vez, gera a discriminação e o preconceito segregacionista (...).⁷⁷

⁷⁷ Reproduzindo argumentação similar à prolatada no caso *Ellwanger*, o Ministro Relator Celso de Mello destacou no referido julgamento que: “O conceito de racismo, compreendido em sua dimensão social, projeta-se para além de aspectos estritamente biológicos ou fenotípicos, pois resulta, enquanto manifestação de poder, de uma construção de índole histórico-cultural motivada pelo objetivo de justificar a desigualdade e destinada ao controle ideológico, à dominação política, à subjugação social e à negação da alteridade, da dignidade e da humanidade daqueles que, por integrarem grupo vulnerável (LGBTI+) e por não pertencerem ao estamento que detém posição de hegemonia em uma dada estrutura social, são considerados estranhos e diferentes (...)”.

⁷⁸ WALDRON, Jeremy. *The harm of hate speech*. Cambridge: Harvard University Press, 2012, p. 33.

⁷⁹ HONNETH, Axel. *Luta por Reconhecimento: A Gramática Moral dos Conflitos Sociais*. Trad. Luiz Repa. São Paulo: Editora 34, 2003, p. 218 apud SARMENTO, Daniel. *A Liberdade de expressão e o hate speech*. Mimeografado.

a ilegitimidade da limitação ao discurso de ódio é, em síntese, ignorar o prejuízo inestimável que o discurso de ódio produz à autoestima entre as vítimas do segmento mais estigmatizado.⁸⁰

É o que Owen Fiss denomina de “efeito silenciador” (*silencing effect*) do discurso, que não só impede a plena participação das vítimas em várias atividades da sociedade, incluindo o debate público, como também retira a autoridade de suas palavras, como se nada dissessem.⁸¹

Em segundo lugar, o discurso de ódio nos leva a refletir sobre o “paradoxo da tolerância” de Karl Popper e precisamente sobre os limites da própria democracia.⁸² Nas palavras do autor, “se não estivermos preparados para defender uma sociedade tolerante dos ataques dos intolerantes, o resultado será a destruição dos tolerantes e, com eles, da tolerância”.⁸³ No mesmo sentido, Rawls afirma que “um indivíduo só pode denunciar as violações dos princípios que ele mesmo reconhece”, para concluir que “os intolerantes não possuem o direito de reclamar quando lhes é negada uma mesma liberdade”⁸⁴

É firme nestas razões que Lowenstein defende um princípio de “democracia militante” para sustentar que o regime democrático não é um pacto suicida, o que justificaria o uso de instrumentos de restrição à liberdade de expressão e política por parte de movimentos ou indivíduos que integrem grupos cuja orientação se pautar por pensamentos e expressões calcados pelo preconceito e intolerância.⁸⁵

Sem embargo da eterna discussão sobre quais elementos do discurso, em concreto, justificam seu enquadramento na categoria do *hate speech*, tarefa que tem cabido precipuamente ao Poder Judiciário, por meio de uso da técnica de ponderação⁸⁶, é inequívoco que o Brasil reconhece, em abstrato, atos de racismo, transfobia, homofobia e antissemitismo como limites legítimos ao exercício da autonomia de quem discursa, conforme evidenciado pelos diplomas legislativos e constitucionais citados.

No que tange ao nosso objeto de estudo, os parâmetros para governança dos discursos nas redes sociais, esta consideração é importante, pois todas as redes sociais são gerenciadas

⁸⁰ SARMENTO, Daniel. A liberdade de expressão e o problema do hate speech. *Revista de Direito do Estado*, Rio de Janeiro: Renovar, n. 4, p. 53-105, out./dez. 2006. Mimeografado.

⁸¹ FISS, Owen. *A Ironia da Liberdade de Expressão: Estado, Regulação e Diversidade na Esfera Pública*. Trad. Gustavo Binenbojm e Caio Mário da Silva Pereira Neto. Rio de Janeiro: Renovar, 2005, p. 55.

⁸² A observação é de Daniel Sarmiento (A liberdade de expressão e o problema do hate speech. *Revista de Direito do Estado*, Rio de Janeiro: Renovar, n. 4, p. 53-105, out./dez. 2006. Mimeografado).

⁸³ POPPER, Karl. *A Sociedade Aberta e Seus Inimigos*. Trad. Milton Amado. Belo Horizonte: Ed. Itatiaia; São Paulo: Ed. da Universidade de São Paulo, 1974, p. 118.

⁸⁴ RAWLS, John. *A Theory of Justice*. Cambridge: Harvard University Press, 1999, p. 190 (tradução livre).

⁸⁵ Cf. PONTES, João Gabriel Madeira. *Democracia militante em tempos de crise*. Rio de Janeiro: Lumen Juris, 2020.

⁸⁶ Neste sentido, veja-se artigo já mencionado do professor Daniel Sarmiento, que indica parâmetros para o uso da proporcionalidade na definição dos discursos que podem caracterizar-se como *hate speech*.

por companhias estrangeiras, que nem sempre terão valores comunitários e termos de uso em conformidade com os parâmetros jurídicos nacionais.⁸⁷ A conformidade à legislação nacional é, portanto, um elemento a ser considerado quando se debate a regulação do funcionamento das redes sociais.

1.3 Os riscos da regulação pública

A regulação estatal da liberdade de expressão e de canais essenciais à formação da esfera pública pode, com efeito, justificar-se à luz das razões que sustentam a própria natureza fundamental da liberdade de expressão, seja para viabilizar o “livre mercado de ideias”, seja para assegurar condições que contribuam para a melhor qualidade do debate na esfera pública para fins de fortalecimento do regime democrático; sem prejuízo de poder ainda ser necessária para afastar da esfera discursiva determinados conteúdos de caráter odioso, que possam causar danos substanciais à dignidade de terceiros e à coletividade.

A experiência histórica acerca da imposição de limites à liberdade de expressão por parte do Estado recomenda, contudo, algumas cautelas na definição do espaço regulatório legítimo da intervenção estatal. E, aqui, há dois riscos genericamente atribuídos a qualquer pretensa intervenção regulatória com impactos no exercício da liberdade de expressão: o risco do autoritarismo estatal e, ainda, o *chilling effect* (efeito resfriador) produzido por intervenções do Estado sobre o debate público.

1.3.1 A Censura Estatal e o Autoritarismo

Quando se pensa em censura, o primeiro grande desafio é definir exatamente o que a caracteriza, especialmente delimitando o que difere a “censura” de um legítimo limite à liberdade de expressão.

⁸⁷ Felizmente, tal fato não tem sido a regra. O propósito de expansão global destas empresas tem, na verdade, levado tais plataformas a relegarem a tradição norte-americana para se adequar aos parâmetros de outros países. Talvez o principal exemplo, hoje, de rede social de alcance significativo resistente à criação de mecanismos de combate ao ódio em suas plataformas seja o *Telegram*.

A liberdade de expressão inclui a liberdade de expressar um ponto de vista e ainda de receber quaisquer dessas expressões. A censura é compreendida em essência como uma interferência por supressão de quaisquer dessas expressões, ou ainda o impedimento à divulgação de tais expressões por parte de um censor, dotado de poder (de punir ou financeiro), que o faz com base na “convicção de que será ruim se as pessoas acessarem o conteúdo daquela expressão”.⁸⁸ Por regra, a censura partirá de um órgão governamental, mas nada obsta que decorra de entidades privadas quando “tenham controle sobre uma modalidade inteira de comunicação”⁸⁹, que lhes permita restringir certas categorias de expressão. A diferença entre a censura e uma limitação legítima seria normativa, pois na última há boas razões para sustentar tal restrição.

Definir quais seriam as boas razões para restringir a esfera discursiva traduz o receio para muitos de que não haja argumento hábil a distinguir uma censura de uma limitação legítima da expressão. Neste sentido, é comum se apresentar como contraponto a intervenções estatais sobre fontes de discurso um argumento de “ladeira escorregadia”, segundo o qual a falta de um critério claro e objetivo para indicar as hipóteses possíveis à restrição de liberdade de expressão levará a sociedade a uma situação na qual a censura será a regra.

Ainda que a “ladeira escorregadia” seja usualmente referenciada como uma falácia⁹⁰, é certo que algumas experiências históricas implicam considerar tais razões com maior seriedade no campo da liberdade de expressão. Como bem explica o Ministro Luís Roberto Barroso, “em todos os tempos e em todos os lugares, a censura jamais se apresenta como mecanismo de intolerância, de prepotência ou de outras perversões ocultas. Ao contrário, como regra, ela destrói em nome da segurança, da moral, da família e dos bons costumes. Na prática, todavia, oscila entre o arbítrio, o capricho, o preconceito e o ridículo”.⁹¹

De fato, são muitos os episódios de censura desde a formação do Estado moderno. A Alemanha nazista, por exemplo, queimou livros científicos e literários para “proteção da nação” e do Estado, e a relação dos livros judeus proibidos pela Gestapo chegou a 12.400 títulos e 149 autores.⁹²

⁸⁸ MATHIESEN, Kay. Censura e Acesso à Expressão. In: SAMPAIO, José Adércio Leite (Coord.). *Liberdade de Expressão no Século XXI*. Belo Horizonte: Del Rey, 2016, p. 5-27.

⁸⁹ *Ibid.*

⁹⁰ VOLOKH, Eugene. The Mechanisms of Slippery Slope. *Harvard Law Review*. Vol.116. nº4. pp.1026-1137.2003.

⁹¹ BARROSO, Luís Roberto. Liberdade de expressão, censura e controle da programação na Constituição de 1988. In: _____. *Temas de Direito Constitucional*. Rio de Janeiro: Renovar, 2001. Tomo 1, p. 345-346.

⁹² NOVINSKY, Anitta. Os regimes totalitários e a censura. In: TUCCI, Maria Luiza (Org.). *Minorias Silenciadas: História da Censura no Brasil*. 2. ed. São Paulo: Editora da Universidade de São Paulo, 2020, p. 27-37.

No Brasil, em um passado recente observa-se exemplos trágicos dos seus efeitos sobre a imprensa e a sociedade civil. No Estado Novo, de Getúlio Vargas, o Departamento de Imprensa e Propaganda tinha por objetivo “fazer a censura do teatro, do cinema, das funções recreativas e esportivas, da radiodifusão, da literatura social, da política e da imprensa”⁹³.

Mais recentemente, a censura generalizada se deu durante a ditadura militar, entre 1964 e 1985, especialmente durante a vigência do Ato Institucional nº 5. Na imprensa escrita, os jornais eram submetidos à censura prévia, apreendiam-se jornais e revistas, boicotava-se a publicidade daqueles que eram percebidos como ameaças pelo governo. Ao todo, cerca de 500 filmes, 450 peças, 200 livros e mais de 500 letras de música sofreram vetos.⁹⁴

Em todos estes casos, o receio de um dano, usualmente atrelados a atividades subversivas, e o apelo ao patriotismo consistiram no principal argumento para cerceamento da palavra crítica, o que torna evidente que não é absurdo cogitar a escalada do autoritarismo, por meio da censura, como um fato possível.

O anseio de dominação da esfera discursiva é uma pretensão de qualquer governante com inclinações autoritárias. Em uma democracia liberal, no entanto, há inúmeros contrapontos ao exercício do poder majoritário, que permitem distinguir um limite legítimo de um ato abusivo de censura, evitando um declínio ao autoritarismo temido pelos defensores da tese da “ladeira escorregadia”.

Guinadas ao autoritarismo ocorrem principalmente em duas circunstâncias. O mais usual são os golpes de Estado, que implodem por completo sistemas democráticos levando à banalização de episódios de censura e perseguição a opositores. Aqui, a censura é mais uma consequência da derrubada do regime democrático do que uma causa, uma vez que a superação da democracia invalida, por completo, o sistema de direitos fundamentais.

A ciência política tem dado ênfase, ainda, ao processo de lenta erosão de grades de proteção da democracia até um ponto de não retorno⁹⁵. A polarização extremada associada à rejeição da política e a perda de credibilidade das instituições leva à ascensão de candidatos

⁹³ BARROSO, Luís Roberto. *Sem data venia: um olhar sobre o Brasil e o Mundo*. [S.l.]: História Real, 2020, p. 112.

⁹⁴ *Ibid.*, loc. cit.

⁹⁵ Neste sentido, Levitsky e Ziblatt (*Como as Democracias Morrem*. Trad. Renato Aguiar. Rio de Janeiro: Zahar, 2018).

com tendências autoritárias⁹⁶, que, especialmente quando legitimados pela reeleição⁹⁷, podem vir a perseguir e censurar opositores. Mesmo neste caso, a instituição da censura ocorre após um processo inicial de fragilização de instituições liberais de controle,⁹⁸ que pavimentam o caminho para leis posteriores que criem mecanismos supressores das liberdades individuais.

A banalização da censura é, portanto, mais uma sequela da fragilização dos mecanismos de contraponto ao exercício do poder majoritário do que sua causa. Em outras palavras, o temor do autoritarismo é algo que sempre irá persistir mesmo em democracias saudáveis, mas é possível encontrar inúmeros pontos de paradas — institucionais, formais e materiais — que evitem um deslizamento completo da imposição de alguns limites substanciais ao exercício da liberdade de expressão para um cenário de naturalização da censura.

Quais seriam estes pontos de parada? Em primeiro lugar, um Poder Judiciário independente, que forneça a salvaguarda institucional necessária a coibir abusos dos demais poderes na elaboração de legislações restritivas ao discurso. Aqui, subsiste o papel do Poder Judiciário como sentinela contra os riscos da maioria⁹⁹ e aquele a quem cabe, em último caso, perquirir as razões que fundamentam a imposição de um limite legítimo à liberdade de expressão.¹⁰⁰

Ainda que os limites da atuação do Poder Judiciário seja um tema sempre atual nos debates acadêmicos no Brasil e no mundo, sua atuação contramajoritária costuma ser um ponto de consenso nas sociedades ocidentais. O afastamento do Poder Judiciário do processo eleitoral, a inércia judicial, a dialética processual (que favorece o uso da racionalidade), a transparência do procedimento judicial, bem como a *expertise* argumentativa dos tribunais¹⁰¹, conferem um

⁹⁶ CASTELLS, Manuel. *A ruptura: a crise da democracia liberal*. Rio de Janeiro: Zahar, 2018, cap. 2. Em sentido similar, falando ao contexto brasileiro, ver Souza Neto (*A Democracia em Crise no Brasil: valores constitucionais, antagonismo político e dinâmica institucional*. São Paulo: Contracorrente, 2020).

⁹⁷ Cláudio Pereira Souza Neto argumenta que a reeleição do candidato pretensamente autocrático é um provável ponto de não retorno para conversão da democracia em autocracia (SOUZA NETO, Cláudio Pereira. *A Democracia em Crise no Brasil: valores constitucionais, antagonismo político e dinâmica institucional*. São Paulo: Contracorrente, 2020, p. 31).

⁹⁸ Sobre a importância dos mecanismos liberais de controle como forma de proteção do povo contra si mesmo, cf. MOUNK, Yascha. *O povo contra a democracia: porque nossa liberdade corre perigo e como salvá-la*. Trad. Cássio de Arantes Lete e Débora Landsberg. São Paulo: Companhia das Letras, 2019.

⁹⁹ BARROSO, Luís Roberto. *A Judicialização da Vida e o Papel do Supremo Tribunal Federal*. Belo Horizonte: Fórum, 2021, p. 156.

¹⁰⁰ RAWLS, John. *A Theory of Justice*. Cambridge: Harvard University Press, 1999, p. 225-233.

¹⁰¹ Frise-se que há uma extensa dogmática constitucional apta a orientar inclusive o processo de tomada de decisões difíceis no caso de conflitos entre direitos fundamentais, usualmente referenciada como técnica da ponderação, e que tem no princípio constitucional da proporcionalidade a medida para equacionamento entre estes conflitos. Como nosso enfoque é a limitação decorrente de lei e não propriamente os critérios que pautam a atuação do Poder Judiciário, não aprofundaremos no tema. Para leitura, indica-se: BARCELLOS, Ana Paula de. *Ponderação, Racionalidade e Atividade Jurisdicional*. Rio de Janeiro: Renovar, 2005. Com uma visão mais crítica, atentando que os parâmetros de ponderação usualmente utilizados em conflitos relacionados à liberdade de expressão não

porto seguro para correção de eventuais excessos dos demais poderes na imposição de limites à liberdade de expressão.

O segundo ponto de parada são os limites impostos formalmente pela própria Constituição da República. Embora a liberdade de expressão não tenha salvaguarda absoluta, há uma preocupação sabida da Carta maior de assegurar que o pleno gozo da liberdade, intelectual, artística e científica seja exercido independentemente de *censura ou licença prévia* (art. 5º, inciso XIV) e de *qualquer censura de natureza política, ideológica ou artística* (art. 220, § 2º da Constituição Federal). O objetivo da norma constitucional é claro. Evitar episódios como os que remontam ao Estado Novo da Era Vargas ou ao período da Ditadura Militar, nos quais órgãos executivos com atribuição de poder de polícia fiscalizavam e estabeleciam quais conteúdos poderiam ou não circular na esfera pública, imbuído de propósitos ideológicos ou políticos.

Isto não impede, entretanto, que uma norma venha a impor limites à liberdade de expressão, ou regular canais de mídias, e de que tais parâmetros formais, ainda que tenham uma redação mais objetiva, não sejam suficientes *per se* para avaliar a constitucionalidade da proposta. De um lado, tanto a Constituição da República quanto os inúmeros tratados internacionais dos quais o Brasil é signatário preveem a possibilidade de limitar a liberdade de expressão com base em termos abstratos, cuja significação pode variar a depender do contexto (veja o exemplo dos princípios elencados no art. 221 da CRFB). De outro, mesmo a vedação à censura de caráter ideológico ou político requer algum critério para diferenciar o argumento razoável daquele de cunho radical estritamente partidário.

E, aqui, subsiste o terceiro ponto de parada. É preciso que os limites à liberdade de expressão estejam amparados por argumentos materialmente idôneos, isto é, que ultrapassem o filtro da razão pública.

O termo “razão pública” foi cunhado por Rawls com o propósito de justificar quais argumentos possibilitam e ao mesmo tempo limitam o exercício do poder em um contexto plural, onde múltiplas “doutrinas abrangentes” disputam as concepções morais, religiosas e filosóficas diversas sobre as formas de se viver. Neste contexto, as razões passíveis de debate no espaço público seriam aquelas capazes de serem aceitas por todos aqueles de boa-fé, independentemente das concepções privadas sobre vida, sociedade e política de cada um. Trata-

são suficientes para salvaguardar a proteção necessária a este direito a nível jurisprudencial, confira-se: ALMEIDA, Guilherme da Franca Couto Fernandes de; LEITE, Fábio Carvalho; HANNIKAINEN, Ivar Allan Rodriguez. Liberdade de Expressão e Direito à honra: Medindo Atitudes e Prevendo Decisões. *Espaço Jurídico Journal of Law*, Joaçaba, v. 21, n. 1, p. 191-216, jan./jun. 2020.

se de um corolário do dever de civilidade, que demanda o reconhecimento de todos como seres dignos de igual respeito e consideração, eis que o poder político, em uma democracia, é compartilhado por membros de uma sociedade em um projeto de autogoverno coletivo.¹⁰²

De forma mais direta, razões que ultrapassam o filtro da razão pública são aquelas construídas por meio de argumentos recíprocos e universalizáveis, no qual eventual divergência ou polêmica fique constricta a uma “margem de segurança”, que permita que o argumento dos vencedores, ainda que não queridos, sejam ao menos racionalmente aceitos pelos vencidos.¹⁰³

No caso da liberdade de expressão, este cuidado deve ser redobrado, porquanto se trata de usar o filtro da razão pública para, paradoxalmente, restringir o próprio espaço de debate, tornando absolutamente excepcional medidas restritivas sobre o conteúdo. Em todo o caso, um dever é imposto ao Estado, como regulador: o ônus de apresentar justificativas razoáveis para a imposição de uma decisão regulatória, conferindo-lhe ampla publicidade. Nesta seara, a inclusão de restrições ao discurso por meio de leis — submetidas a um amplo escrutínio — tem preferência face a outros mecanismos estatais, como decretos, atos administrativos e até mesmo decisões judiciais, especialmente quando de caráter monocrático.

De fato, em matéria de liberdade de expressão estamos “necessariamente na encosta, gostemos ou não, e a tarefa é sempre decidir o quão alto ou baixo nós escolhemos ir, e não se nós deveríamos descer totalmente”.¹⁰⁴

O que impede que escorreguemos de legítimas medidas restritivas à liberdade de expressão para a censura generalizada é, portanto, a garantia institucional de um Poder Judiciário independente que possa avaliar com imparcialidade as razões apresentadas pelos demais poderes no exercício da ação regulatória, e, ainda, a observância, por parte do ente regulador, das vedações formais trazidas pela Constituição, quais sejam, o impedimento à exigência de licença ou censura prévia e a vedação a qualquer controle de ordem artística, ideológica ou política sobre a liberdade de se expressar.

1.3.2 O *chilling effect*

¹⁰² RAWLS, John. *O Liberalismo Político*. [S.l.]: Presença, 1997, p. 218-219.

¹⁰³ MENDONÇA, José Vicente Santos de. *Direito Constitucional Econômico: a intervenção do Estado na economia à luz da razão pública e do pragmatismo*. 2. ed. Belo Horizonte: Fórum, 2018, p. 121.

¹⁰⁴ VAN MILL, David. Free speech. In: ZALTA, Edward (Ed.). *The Stanford Encyclopedia of Philosophy*. [S.l.]: [s.n.], 2002. Disponível em: <http://plato.stanford.edu/archives/win2002/entries/freedom-speech/>. Acesso em: 20 dez. 2022.

Há, no entanto, um outro temor decorrente de medidas restritivas ou regulatórias sobre a liberdade de expressão: o suposto efeito resfriador que estas produzem ao debate público, mais comumente referido como *chilling effect*.

A expressão foi desenvolvida pela jurisprudência norte-americana, havendo alguma disputa quanto a sua origem histórica. Suneal Bedi conecta o surgimento do termo a um julgado de 1952 da Suprema Corte Norte Americana (*Wieman x Updegraff*). No caso, o Estado de Oklahoma exigiu um “juramento de lealdade” de seus servidores, afirmando que não defendiam a derrubada violenta do governo dos EUA e negando envolvimento direto ou indireto com o apoio de qualquer agência, partido, organização, associação ou grupo que tenha sido oficialmente reconhecido como “comunista” pelo Procurador Geral dos Estados Unidos. Vários professores do Estado se recusaram a prestar tal depoimento e Paul W. Updegraff, um cidadão de Oklahoma, ajuizou ação no Tribunal Distrital do Condado de Oklahoma para impedir que o Estado pagasse salários a esses indivíduos. A Corte invalidou a norma por violar a décima quarta emenda. Em um dos votos, o *Justice* Felix Frankfurter afirmou que “*tal restrição injustificada... tem uma tendência inconfundível de resfriar o livre espírito que todos os professores devem cultivar, especialmente...*”.¹⁰⁵

Frederick Schauer destaca, no entanto, que a referência a um *chilling effect* (efeito resfriador) — e não propriamente ao uso do verbo “*chill*” isoladamente — é posterior a este caso, remontando a precedente de 1963 da Suprema Corte Norte-Americana¹⁰⁶. A partir daí, o conceito se desenvolveu de um argumento emocional para um método relevante na identificação de quais discursos poderiam ou não ser objeto de regulação governamental.¹⁰⁷

No Brasil, a expressão, embora não seja de uso comum pela doutrina¹⁰⁸, já fora evocada em várias oportunidades pelo Supremo Tribunal Federal como argumento de reforço para invalidação de normas e atos de caráter excessivamente genérico relacionadas à liberdade

¹⁰⁵ BEDI, Suneal. The Mith of Chilling Effect. *Harvard Journal of Law & Technology*, [s.l.], v. 35, n. 1, p. 267-307, 2021. Mimeografado.

¹⁰⁶ SCOTUS, *Gibson v. Florida Legis. Investigation Comm.*, 372 U.S. 539, 556-557 (1963). Nesta decisão, a Suprema Corte disse: “*While, of course, all legitimate organizations are the beneficiaries of these protections, they are all the more essential here, where the challenged privacy is that of persons espousing beliefs already unpopular with their neighbors and the deterrent and “chilling” effect on the free exercise of constitutionally enshrined rights of free speech, expression, and association is consequently the more immediate and substantial.*”

¹⁰⁷ SCHAUER, Frederick. Fear, Risk and the First Amendment: Unraveling the Chilling Effect. *Boston University Law Review*, Boston, v. 58, 1978, p. 687.

¹⁰⁸ À título de exceção, o tema é abordado com maior profundidade por FARAH, André. *Liberdade de expressão e remoção de conteúdo*. Rio de Janeiro: Lumen Juris, 2018.

acadêmica e de expressão de servidores públicos¹⁰⁹, assim como da liberdade de reunião.¹¹⁰ A expressão é ainda comumente referenciada, no STF, para justificar imunidades parlamentares.¹¹¹

Mas, afinal, no que consiste o efeito resfriador (*chilling effect*)? A construção da doutrina é composta por dois objetivos. O primeiro relaciona-se com a inevitabilidade do erro, preocupação que já era destacada desde os tempos de Mill. Por mais perfeito que seja um sistema jurídico sempre haverá erros na aplicação de uma norma, o que implicará vez ou outra a limitação indevida de um discurso legítimo¹¹². Um dos objetivos da doutrina do *chilling effect* é reduzir, tanto quanto possível, essa margem de erro. O segundo é mitigar os efeitos do viés de aversão ao risco, qual seja, a suposição de que a combinação de alguns fatores na elaboração da norma (vagueza e excessiva abstração, especialmente quando associada a sanções civis e criminais onerosas) tende a provocar um excesso de dissuasão e silenciamento em alguns indivíduos, pois somos avessos ao risco¹¹³.

Perceba-se que a preocupação da doutrina do *chilling effect* é substancialmente distinta da abordada anteriormente quanto ao risco de naturalização da censura. Enquanto nesta há o receio de que a ampliação da regulação estatal sobre o discurso possa promover o autoritarismo e perseguições políticas, naquela a preocupação é com uma espécie de externalidade negativa da regulação pública: o risco de que mesmo diante de atos normativos produzidos de boa-fé

¹⁰⁹ O argumento fora apresentado pelo Ministro Luís Roberto Barroso nas ADI's 5537, 5580 e 6038/AL, que julgou inconstitucional leis do Estado de Alagoas, que impunham ao corpo docente dos colégios estaduais o impedimento de induzir aos alunos opiniões político-partidárias, filosóficas e religiosas. Na visão do Ministro, “o nível de generalidade com o que as muitas vedações previstas pela Lei 7.800/2016 foram formuladas gera um risco de aplicação seletiva e parcial das normas (chilling effect)(…), por meio da qual será possível imputar todo tipo de infrações aos professores que não partilhem da visão dominante em uma determinada escola ou que sejam menos simpáticos à sua direção”. Mais recentemente, a referência ao *chilling effect* surgiu como argumento de reforço no voto de diversos ministros (mais precisamente, Ministros Gilmar Mendes, Luiz Fux e Rosa Weber), no julgamento da ADPF nº 722/DF (Min. Relatora Carmen Lúcia, publicado em 22/10/2020), no qual se concedeu medida cautelar para suspender todo e qualquer ato do Ministério da Justiça e Segurança Pública de produção ou compartilhamento de informações sobre a vida pessoal, as escolhas pessoais e políticas, as práticas cívicas de cidadãos, servidores públicos federais, estaduais e municipais identificados como integrantes de movimento político antifascista.

¹¹⁰ O precedente clássico é a ADPF nº 187/DF, que julgou a legalidade das denominadas “marchas da maconha”. Dentre inúmeros argumentos, o Ministro Relator Celso de Mello destaca a importância do livre mercado de ideias para viabilizar a força democrática do dissenso e o risco do *chilling effect* produzidos por normas limitadoras do debate público. O argumento fora novamente utilizado na ADI nº 5582/MS, (Ministro Relator Luiz Fux, Tribunal Pleno, publicado em 26/08/2020). No precedente se declarou a inconstitucionalidade de lei estadual que criava uma série de restrições para o exercício do direito de reunião e associação no espaço territorial onde instalado o centro político e administrativo do Governo do Mato Grosso do Sul, sob pena de cominação de sanções pecuniárias elevadas. Em seu voto, a Ministra Rosa Weber chamou atenção ao *chilling effect* por força do efeito inibitório das sanções criadas pela norma impugnada.

¹¹¹ A título de exemplo, RE. 600.063/SP, Relator para o acórdão o Ministro Roberto Barroso, Plenário, DJe 15.5.2015 e RE 1232884/SP, Min. Relatora Carmen Lúcia, Plenário, DJe: 14/10/2019.

¹¹² SCHAUER, Frederick. Fear, Risk and the First Amendment: Unraveling the Chilling Effect. *Boston University Law Review*, Boston, v. 58, 1978, p. 697.

¹¹³ *Ibid.*, p. 701.

possa se coibir, incidentalmente e de forma excessiva, discursos lícitos. Neste sentido, Schauer sustenta a existência de um princípio do “dano comparativo” para reforçar que a permissão equivocada da livre circulação de um discurso danoso sempre será menos prejudicial do que uma restrição indevida da liberdade de expressão.¹¹⁴

Para ficarmos em um exemplo concreto, vale referência ao multicitado caso *New York Times vs Sullivan*, julgado pela Suprema Corte Norte-Americana.¹¹⁵ No precedente, fora cunhada a chamada doutrina da *actual malice*, no qual se estabeleceu que uma imputação falsa feita a um funcionário público deve, em regra, gozar de proteção, salvo se ficar evidenciado que a informação falsa foi difundida com conhecimento de sua real ou provável falsidade ou com temerário desinteresse a respeito de se a dita informação era verdadeira ou não.

Na visão de Schauer, o precedente se preocupa mais com um efeito resfriador sobre o discurso legitimamente protegido pela primeira emenda do que com efeito danoso de um discurso difamatório. A ideia é evitar que uma opinião legítima ou uma informação verdadeira seja indevidamente punida, e não punir a difamação, pois o dano à liberdade de expressão é sempre comparativamente maior.¹¹⁶ Neste sentido, o papel da doutrina do *chilling effect* é criar uma zona de amortecimento — identificável por parâmetros objetivos — onde eventuais excessos discursivos serão tolerados (no caso, o discurso que se pauta em uma verdade subjetiva ainda que posteriormente se verifique como falso).

Com efeito, se pudéssemos resumir a doutrina do *chilling effect* em uma formulação poderíamos afirmar que esta seria: considerando a inevitabilidade de erro do aparato estatal e a aversão ao risco dos indivíduos, o Poder Público deve reduzir, tanto quanto possível, a abstração, vagueza e generalidade na elaboração de leis, atos normativos ou decisões judiciais, criando parâmetros objetivos para justificar a restrição ao exercício da liberdade de expressão.

Há algumas críticas a tal doutrina. Duas são as principais: (i) a ausência de comprovação empírica de um efeito resfriador provocado por regulações, mesmo quando dotadas de maior amplitude e vagueza; (ii) a contestação quanto ao argumento do “dano comparativo” trazido por Schauer.

¹¹⁴ SCHAUER, Frederick. Fear, Risk and the First Amendment: Unraveling the Chilling Effect. *Boston University Law Review*, Boston, v. 58, 1978, p. 704.

¹¹⁵ O caso em tela diz respeito a uma ação ajuizada por funcionário público criticado por alegadas informações difamatórias da imprensa. No bojo do processo, comprovou-se que várias das alegações contidas no jornal eram imprecisas. Vale frisar que, na época, o entendimento prevalecente era de que qualquer informação falsa ou imprecisa que viesse a ferir a honra ou imagem de alguém deveria ser punida.

¹¹⁶ *Ibid.*, p. 708

A primeira crítica, com a qual não se alinha, questiona se de fato existe um efeito resfriador decorrente de regulações estatais. Afirma-se que não há evidência empírica de *chilling effect*, pois não há um conjunto sólido de pesquisas que ateste sua existência.

O problema principal de tais pesquisas é que elas buscam apenas compreender se os indivíduos alegariam que seriam afetados por uma certa censura hipotética¹¹⁷

Tais pesquisas, não raro, direcionam a resposta, e o resultado pela existência ou inexistência do efeito resfriador depende da forma como o pesquisador irá conduzir a investigação¹¹⁸. São pouquíssimos os estudos que impõem de fato uma situação de censura para, então, verificar como as pessoas se comportam nestas situações.¹¹⁹

A ausência de um conjunto de estudos mais sólidos empíricos capazes de comprovar o efeito resfriador (e sua extensão) não significa, no entanto, que ele não exista, apenas que é um dado de difícil comprovação.¹²⁰

Há alguns bons motivos que nos levam a divergir desta crítica e militar a favor da doutrina do *chilling effect*. Primeiro, normas excessivamente genéricas e vagas são algo a ser evitado em qualquer área do direito, e não é diferente no campo da liberdade de expressão. Ademais, o viés da “aversão ao risco” é um elemento empírico comprovável que, no mínimo, denota a plausibilidade da teoria. Isso não significa que pesquisas empíricas sobre o tema sejam inúteis. Em verdade, o aspecto mais interessante e útil destas pesquisas seria definir em que medida esse efeito intimidador/resfriador se reproduz em determinados contextos discursivos, ao invés de investigar se este efeito existe ou não. Tais estudos poderão, em momento futuro, viabilizar respostas regulatórias mais adequadas a cada espécie de discurso e, portanto, proporcionais por parte do Poder Público.

A segunda crítica, que nos parece parcialmente procedente, deriva da constatação de Schauer e da interpretação consolidada na jurisprudência norte-americana de que um dano à

¹¹⁷ Por exemplo: se houvesse uma regulação governamental que punisse a “intenção de assediar outra pessoa”, isso impactaria na sua forma de escrever ou falar algo *online*? E a pergunta é seguida de uma série de respostas que vão de “nenhum impacto” a “muito provavelmente”.¹¹⁷ Há diversos estudos que seguem esta linha com objetivo de medir o *chilling effect*. A respeito, confira-se Penney (Internet Surveillance, Regulation, and Chilling Effects Online: A Comparative Case Study. *Internet Policy Review*, [s.l.], v. 6, n. 2, mai. 2017. Mimeografado).

¹¹⁸ KENDRICK, Leslie. Speech, Intent and the Chilling Effect. *William & Mary Law Review*, [s.l.], v. 54, n. 5, p. 1633-1691, 2013.

¹¹⁹ Um dos poucos estudos que aplica essa modelagem empírica desvendou que o *chilling effect* de fato existe, mas nem sempre provoca a intimidação da forma como prevista pela doutrina. Em alguns casos, este efeito seria positivo, já que pode implicar modificações sutis na forma e no tom da expressão — tornando-as mais civilizadas — sem propriamente alterar seu conteúdo (BEDI, Suneal. The Mith of Chilling Effect. *Harvard Journal of Law & Technology*, [s.l.], v. 35, n. 1, p. 267-307, 2021. Mimeografado).

¹²⁰ Schauer chega a afirmar, ironicamente, que o *chilling effect* é “provavelmente improvável” (“*most likely unprovable*”, no original) (SCHAUER, Frederick. Fear, Risk and the First Amendment: Unraveling the Chilling Effect. *Boston University Law Review*, Boston, v. 58, 1978, p. 730).

liberdade de expressão será sempre mais grave que o dano produzido a outro bem jurídico. Vale a indagação: será mesmo? A afirmativa parece fazer sentido em um primeiro momento, mas, quando reduzimos o nível de sua abstração, percebe-se que nem sempre é assim.

Sob o ponto de vista dogmático, há um considerável número de constitucionalistas que concordam com a atribuição de alguma preferência axiológica para a liberdade de expressão, por sua estrita vinculação com a autonomia individual (corolário da dignidade da pessoa humana), e com a própria ideia da democracia, mas são poucos os que confiam a tal direito uma prevalência absoluta.

É possível encontrar exemplos práticos de casos em que o risco de “falsos positivos” para violação à liberdade de expressão é assumido como razoável, porque há um consenso social estabelecido de que certas expressões são graves o suficiente para justificar restrições mais rigorosas.

Veja-se um exemplo do Marco Civil da *Internet*, norma sobre a qual se debruçará nos próximos capítulos.

O art. 19 — regra de ouro da responsabilização das aplicações de conteúdo — prevê a possibilidade de responsabilizar o provedor de conteúdo somente se, após decisão judicial específica, o conteúdo indicado como danoso não for retirado. Qual é o raciocínio por trás da norma? O Poder Judiciário funciona como filtro para evitar um excesso de bloqueios por parte dos provedores de conteúdo lícito, impedindo que a mera provocação de particulares incomodados com alguma publicação possa gerar alguma obrigação jurídica e produzir o efeito resfriador do debate pela *internet*. Essa regra é, entretanto, flexibilizada no caso de divulgação de material com cena de nudez, sem autorização dos seus participantes (art. 21 da mesma lei).

Neste caso, a mera provocação do particular que se sentir prejudicado — observados alguns parâmetros formais mínimos — é suficiente para justificar a responsabilidade do provedor. O legislador assume aqui o risco de “falsos positivos” que possam levar à restrição de alguma manifestação lícita de nudez *online*, dada uma percepção social compartilhada de que os efeitos produzidos pelo fenômeno conhecido como “pornografia de vingança” (*revenge porn*) são graves o suficiente para justificar eventuais restrições indevidas à liberdade de divulgar eventuais conteúdos lícitos. A preocupação quanto ao *chilling effect* existente na regra do art. 19 é, portanto, deixada de forma consciente em um segundo plano no art. 21.¹²¹

¹²¹ Outra exceção — certamente mais polêmica — é a previsão constante do art. 19, § 2º do Marco Civil da *Internet*, que afasta as regras do *caput* em relação a discussões que envolvam direitos autorais, que deverão ser regulamentadas por lei específica. Como essa lei ainda não existe, a mera notificação na via administrativa já é suficiente para levar à responsabilização do intermediário que é negligente no dever de retirar o material violador de direito autoral de sua página.

Este exemplo, embora não afaste a premissa de que por regra a liberdade de expressão deverá prevalecer, indica que a ideia de uma preferência absoluta, a que alude Schauer, não se sustenta e que, portanto, o receio de *chilling effect* não é um impedimento em absoluto a leis que restrinjam a autonomia discursiva.¹²²

Há, ainda, quem veja a importação da doutrina do *chilling effect* com alguma inquietação, pois ela só faria sentido em um sistema jurídico que adotasse a liberdade de expressão como um valor preferencial se comparado a outros direitos fundamentais. Tal construção, certamente, faz sentido no contexto norte-americano, que trabalha sob uma lógica de categorização. Busca-se definir qual discurso é ou não abarcado pela primeira emenda. Aquele por ela protegido gozaria de uma proteção absoluta. A técnica seria incompatível, contudo, com uma ideia de verificação *ad hoc* do valor prevalecente no caso concreto, mediante uso da ponderação. Se a prevalência da liberdade de expressão é absoluta, não há o que ponderar.

Esta crítica tem sido respondida pelo Supremo Tribunal Federal por um caminho do meio, à luz do que é denominado de primazia *prima facie* da liberdade de expressão, uma versão *soft* da visão americana de que a liberdade de expressão seria um valor preferencial (*preferred value*).¹²³ Nesta linha, a liberdade de expressão — especialmente a informativa — gozaria de uma prevalência axiológica em princípio, o que lhe conferiria um maior peso em abstrato na partida do juízo valorativo, mas que poderia vir a ceder em face de circunstâncias específicas do caso concreto que tornassem inequívoco o risco da não limitação na circulação do discurso¹²⁴.

¹²² Um ponto mais sensível neste embate — e que fora abordado quando se avaliou os impactos negativos dos discursos de ódio — é o fato de que o efeito resfriador (decorrente da insegurança propiciada pelo risco de mais falsos positivos) promovido pela regulação pública pode, por vezes, justificar-se por representar um risco menos grave do que o efeito silenciador produzido pelo próprio discurso (expressão cunhada por Owen Fiss para rotular o impacto à participação na esfera pública produzido por discursos discriminatórios). Neste caso, legislações mais restritivas ao discurso de ódio — que por vezes gozam de grau considerável de abstração com previsão de sanções criminais — poderiam vir a ser compreendidas como constitucionais em contextos de perseguição histórica sofrida por determinados grupos minoritários, ainda que suas expressões possam suscitar maiores divergências e insegurança jurídica no momento da sua interpretação.

¹²³ Este tem sido o entendimento do STF, que desde a ADPF 130/DF tem assentado uma primazia *prima facie* da liberdade de expressão a exigir maior rigor na análise de constitucionalidade de normas que versem sobre o tema. O aspecto preferencial da liberdade de expressão é reforçado em outros precedentes, tais como a ADI nº 3.741 (que afastou proibição de divulgar pesquisas eleitorais a menos de quinze dias do pleito); ADPF nº 187 (que reconheceu a licitude da proclamada “Marcha da Maconha”); ADI nº 4815 (que confirmou a constitucionalidade de biografias não autorizadas); ADI nº 2566 (julgou inconstitucional a proibição de “proselitismo” por rádios comunitárias); ADI nº 4451 (que tratou da proibição de uso de humor e sátira durante período eleitoral) e RE 1010606/RJ (que tratou do “direito ao esquecimento”).

¹²⁴ O efeito prático de se reconhecer a primazia *prima facie* da liberdade de expressão é algo muito discutível. O Supremo Tribunal Federal tem apontado, no geral, três efeitos: (I) um escrutínio mais rigoroso de constitucionalidade de leis e medidas que impliquem cerceamento à liberdade de expressão; (II) a vedação à

Não se nega, entretanto, o dever de o Estado buscar redações mais objetivas para leis que incidam, direta ou indiretamente, sobre a liberdade de expressão, a fim de reduzir o risco de um efeito resfriador no debate público. Uma margem mais ampla de discricionariedade — e, portanto, a assunção de um maior risco de erro — pode vir a ser aceita, mas somente de forma muito excepcional, no caso de haver uma percepção geral e razoavelmente justificável em um dado contexto social acerca do caráter excessivamente danoso de algumas espécies pontuais de discurso.

Não obstante, o aspecto mais interessante de considerarmos sobre o *chilling effect* é um fato que não poderia ter sido imaginado por Schauer quando escreveu seu artigo seminal sobre o tema em 1979: o de que grandes conglomerados que passariam a deter o controle sobre canais discursivos que, embora privados, produzam substancial impacto sobre a esfera pública discursiva. Neste sentido, a definição por empresas privadas de quais discursos devem ou não circular nas redes sociais é vista por muitos como um novo elemento produtor do *chilling effect*, pois aqui também estão presentes a inevitabilidade do erro e o viés de aversão ao risco.¹²⁵

Assim como na regulação pública, as plataformas de redes sociais precisam não só elaborar como também aplicar as normas de seus termos de uso para assegurar o bom funcionamento dos canais discursivos que administram, o que inegavelmente levará a eventuais erros — por excesso ou negligência —, e poderá traduzir em alguma espécie de *chilling effect* se a execução de tais normas não ocorrer de forma transparente e coerente, fato que se agrava pela imensidão dos pedidos que tais plataformas analisam diariamente e a inviabilidade de uso exclusivo da mão de obra humana para avaliar a adequada aplicação destas regras, tornando necessário a recorribilidade a mecanismos de inteligência artificial, poucos transparentes e muitas vezes incapazes de fazer, ao menos no estágio atual da tecnologia, análises contextuais

censura prévia; (III) um ônus argumentativo e probatório diferenciado daquele que alega violação a um direito da personalidade em detrimento da liberdade de expressão (as observações constam do voto do Ministro Luís Roberto Barroso na ADI 4815/DF). Assim, o principal efeito concreto seria imputar o ônus argumentativo e da prova em desfavor daquele que pretende restringir a liberdade de expressão, pois a vedação à censura prévia já decorreria de um comando constitucional expresso, sendo absolutamente desnecessária uma construção teórica exclusivamente para tal finalidade. Há inúmeros trabalhos doutrinários e empíricos que atestam, no entanto, a inutilidade destes precedentes do Supremo Tribunal Federal, visto que são usualmente ignorados pelos tribunais de segunda instância. Neste sentido: HARTMANN, Ivar Alberto Martins. A liberdade de expressão nas decisões de primeira instância do TJ-RJ. *Opin. Jur.*, Fortaleza, ano 18, n. 27, p. 13-44, jan./abr. 2020; HARTMANN, Ivar Alberto Martins. A crise dos precedentes no supremo: o caso dos precedentes sobre liberdade de expressão. *Revista Estudos Institucionais*, [s.l.], v. 6, n. 1, p. 109-128, jan./abr. 2020.

¹²⁵ A título de exemplo, cf. DOUEK, Evelyn, Governing Online Speech: From 'Posts-As-Trumps' to Proportionality and Probability. *Colum. L. Rev.*, [s.l.], v. 121, n. 759, ago. 2020. Disponível em: <https://ssrn.com/abstract=3679607>. Acesso em: 20 dez. 2022; MARDER, Ben et al. The extended chilling effect of Facebook: The cold reality of ubiquitous social networking. *Computers in Human Behavior*, [s.l.], v. 60, 2016.

com a sofisticação que um raciocínio pautado por juízos de proporcionalidade muitas vezes exige¹²⁶.

Compreender como o *chilling effect* decorrente da autorregulação privada das plataformas opera — se mais, menos intenso ou de forma equivalente à regulação estatal — e sua relação com o potencial efeito resfriador proveniente de um ato normativo de natureza pública é um novo ponto de estudo para os debates relacionados à liberdade de expressão na *internet* e nas redes sociais, que merece ser aprofundado.

Há, contudo, uma inclinação para se sustentar a tese de que a regulação pública de tais canais pode vir a se justificar como medida adequada para restringir o efeito resfriador da atuação destas plataformas privadas, trazendo tanto quanto possível maior objetividade, transparência e segurança à aplicação de suas próprias regras.

1.4 Liberdade de expressão e regulação de arquiteturas

Ao iniciar este capítulo, indicou-se duas formas de regulação com impacto sobre o exercício da liberdade de expressão. A primeira seria a direta, aquela incidente sobre o conteúdo, que se tornaria ilícito dentro de certas circunstâncias enunciadas pela norma, que justificariam a incidência de uma sanção.

A outra é a que incide sobre canais de comunicação, mas não operando diretamente sobre seu conteúdo, mas sim sobre a “arquitetura” de programação daquele canal de comunicação.

A nível doutrinário, é comum ilustrar este ponto com o emblemático julgamento do caso *Red Lion Broadcasting Co. vs F.C.C.*, em 1969¹²⁷, no qual a Suprema Corte Americana analisou a exigência imposta pela Comissão Federal de Comunicações Americana, que determinava que

¹²⁶ O parâmetro da proporcionalidade é, inclusive, adotado oficialmente pelo Comitê de Supervisão do Facebook para ponderar eventuais conflitos entre liberdade de expressão e outros valores indicados como fundamentais pelas normas comunitárias da empresa (autenticidade, segurança, privacidade e dignidade). Informação em GOVERNANCE. *Oversight Board*, [s.l.], [2018]. Disponível em: <https://oversightboard.com/governance/>. Acesso em 08 mar. 2022.

¹²⁷ ESTADOS UNIDOS DA AMÉRICA. *Red Lion Broadcasting Co., Inc v. FCC*, 395, U.S. 367, 9 de junho de 1969. Disponível em: <https://supreme.justia.com/cases/federal/us/395/367>. Acesso em: 9 abr. 2022. Posteriormente, a Suprema Corte Norte-Americana rejeitou a extensão da *fairness doctrine* aos jornais impressos, visto que não haveria aqui o cenário de escassez das licenças de frequências (cf. Id. *Miami Herald Publishing Co v. Tornillo*, 418, U.S. 241, 25 de junho de 1974. Disponível em: <https://supreme.justia.com/cases/federal/us/418/241/#tab-opinion-1950903>. Acesso em: 09 abr. 2022). Em 1987, contudo, a própria FCC revogou a *fairness doctrine*, sob o argumento de que tal doutrina não era mais necessária ante o aumento da competição dos meios de comunicação em massa e por implicar indireto silenciamento do discurso, já que as emissoras passaram a cobrir conteúdos menos controvertidos.

rádios e emissoras de televisão licenciadas discutissem assuntos de interesse público e, ao fazê-lo, assegurassem uma justa cobertura para cada lado da discussão (*fairness doctrine*). A exigência foi considerada constitucional pela Suprema Corte à época sob o argumento de que dado o número limitado de frequências era imperioso uma regulação central para evitar uma “cacofonia de vozes conflitantes”.

Apesar dos méritos da *fairness doctrine* e sua apresentação como uma doutrina pró-regulação estatal, seus argumentos não parece contrariar a lógica da neutralidade que embasa a concepção libertária norte-americana sobre a liberdade de expressão, visto que busca proteger uma divulgação equilibrada de múltiplos pontos de vista, assumindo a neutralidade quanto ao conteúdo como premissa.

Esta visão, repita-se mais uma vez, parece estar aquém da concepção de liberdade de expressão (democrática) adotada por nossa Constituição, que admite não só a regulação sobre a arquitetura dos canais de comunicação, mas que estas sejam orientadas a promoção de certos valores substantivos em detrimento de outros (como se extrai do art. 221 da Constituição Federal).

É possível antecipar que o Supremo Tribunal Federal já prolatou algumas decisões assentando a constitucionalidade de leis e regulamentos que limitem a autonomia de agentes privados da área de comunicação, ao exigir a apresentação de alertas indicativos de conteúdo sensível ou, ainda, fomentando certos tipos de discurso que traduzam algum interesse social relevante à luz da Constituição Federal. Veja-se alguns exemplos.

Na ADI 2404/DF, o Supremo Tribunal Federal analisou a constitucionalidade do art. 254 do Estatuto da Criança e do Adolescente, que tipificava como infração administrativa a transmissão, via rádio ou televisão, de espetáculo em horário diverso do autorizado ou sem aviso de sua classificação.

O Supremo Tribunal Federal declarou a inconstitucionalidade parcial do dispositivo para vetar, tão somente, a tipificação da infração de transmissão de programa em horário diverso do autorizado, porquanto não há nada a ser “autorizado” pelo Poder Público. Tal determinação violaria o art. 5º, inciso IX, que consagra a vedação de exigência de censura prévia. Nada obstante, a Suprema Corte manteve como infração administrativa a transmissão de espetáculo sem aviso da sua classificação indicativa etária. Tal norma não seria limitadora de qualquer espécie de discurso ou forma de expressão, mas antes um alerta informativo que pretende orientar os adultos acerca do melhor programa a ser assistido pelas crianças ou adolescentes sob sua guarda. A pretensão legítima da norma é “*divulgar, para toda a sociedade, a importância de se fazer uma escolha refletida acerca da programação ofertada ao público*”.

*infanto-juvenil sem, contudo, restringir qualquer espécie de discurso*¹²⁸”, não implicando, portanto, qualquer restrição indevida à autonomia da audiência para escolher o melhor programa a ser assistido conforme seu juízo crítico.

No Recurso Extraordinário nº 1070522/PE, o Supremo Tribunal declarou a compatibilidade com a Constituição de regra consagrada no art. 38, “h” da Lei nº 4717/62 (Código Geral de Telecomunicações) e do art. 16 do Decreto nº 52.795/63 (decreto que regulamenta os serviços de radiodifusão), no ponto em que estabelecem como exigência editalícia à concessão do serviço de radiodifusão e difusão de sons e imagens, a observância de percentuais mínimos de exibição de programação especial de produção local. No caso, o voto do Ministro Relator Luiz Fux destaca a legitimidade da medida por estar em conformidade com os princípios constitucionais estabelecidos no art. 221 da Constituição da República e, ainda, por entender que a referida medida não implica mecanismo de censura de conteúdo, mas antes o incentivo a uma “arquitetura de escolha”, que prestigia a programação da localidade, reforçando laços de pertencimento e o desenvolvimento cultural, econômico e social na região. Neste sentido, transcreve-se trecho do voto, que sintetiza os argumentos do Ministro:

Quando o Poder Público aumenta a oferta de programas locais – por meio de um percentual mínimo de exibição, porém sem qualquer vinculação prévia sobre o modo de inserção na grade programação (i. e. formato, horário de exibição) –, ele cria um nudge, isto é, uma política pública de difusão da cultura que se beneficia do viés cognitivo, quanto à predileção por determinadas emissoras ou quanto à escuta de rádio em horários específicos. Na situação sob apreço, o interesse socialmente desejável é conquistar uma audiência cativa para os programas de rádio oriundos do mesmo Município onde se situa o ouvinte, em prol da inserção na comunidade política; do reforço aos laços de identificação e de pertencimento; da movimentação da economia local, dentre outros interesses. Não obstante, a inobservância dessa finalidade social pelas ações individuais não gera punição ao agente, razão pela qual a medida é norma de incentivo (grifo nosso).¹²⁹

Embora o uso do termo *nudge*¹³⁰ pelo ilustre Ministro pareça impreciso, visto que a norma de percentual mínimo restringe, ainda que indiretamente, os conteúdos que podem ser disponibilizados à audiência — indo além do mero “cutucão” —, o precedente indica a receptividade do Supremo Tribunal Federal a regulações que busquem privilegiar certas formas de expressão indicadas como positivas pela Constituição, distinguindo esta prática regulatória

¹²⁸ Trecho do voto do Min. Dias Toffoli, relator do acórdão (ADI 2404/DF, Tribunal Pleno, Ministro Dias Toffoli, Dje: 31/08/2016).

¹²⁹ Trecho do voto do Ministro Luiz Fux, relator do acórdão (RE 1070522/PE, Tribunal Pleno, Min. Rel. Luiz Fux, Dje: 04/03/2021).

¹³⁰ O termo *nudge* refere-se a mecanismos regulatórios de criação de arquiteturas de escolhas que dão um “empurrãozinho” para orientar os particulares à tomada de decisões em um determinado sentido, sem privar o indivíduo de sua liberdade de escolha.

da censura. Veja-se, portanto, um indicativo de que o Supremo Tribunal Federal se afasta da perspectiva libertária de que todos os pontos de vista merecem igual tratamento, mesmo quando trata de regulação das arquiteturas dos canais de comunicação.

A jurisprudência do Supremo traduz, assim, um parâmetro considerado como tolerável para a intervenção estatal em matéria de liberdade de expressão: a possibilidade de uma regulação orientada a influenciar a “arquitetura de escolha” construída pelos agentes privados,¹³¹ elemento potencialmente útil para a construção de resposta regulatória às redes sociais.

1.5 Parâmetros para a regulação pública da liberdade de expressão

A identificação dos elementos capazes de justificar a regulação pública de infraestruturas privadas essenciais à formação da esfera pública, bem como dos riscos inerentes à liberdade de expressão em decorrência destas regulações, permite sistematizar alguns parâmetros para legitimar a intervenção estatal no campo da comunicação.

Nesta esteira, finaliza-se este capítulo com algumas linhas bem gerais sobre os parâmetros a serem observados na elaboração de leis que gozem desta pretensão:

- (I) O Estado tem o dever de fundamentar seus atos regulatórios sob o filtro da razão pública, viabilizando a ampla avaliação dos argumentos apresentados e sempre buscando articular suas premissas em um ou mais dos fundamentos justificadores da proteção ao direito fundamental à liberdade de expressão;
- (II) A regulação pública não pode se converter em medida de censura prévia (art. 5º, IX da CF), tampouco censura por razões não públicas (ideológicas, políticas ou artísticas) — art. 220, § 2º da CF.
- (III) A proibição de circulação de determinados conteúdos serão absolutamente excepcionais, justificando-se tão somente quando o Estado comprovar que há

¹³¹ Ainda que sem fazer referência expressa ao termo “arquitetura de escolhas”, o STF chancelou argumento similar no julgamento da ADI 4923/DF, que julgou a proporcionalidade da política de cotas de conteúdo nacional para os serviços de acesso condicionado, conceito criado pelo art. 5º, XXIII da Lei nº 12.485/2011 (e.g., serviços de televisão por assinatura). O tema foi analisado em conjunto com diversos pontos impugnados da referida lei. No que interessa ao tema desta pesquisa, fora destacado no referido precedente a possibilidade de o Poder Público interferir lícitamente na autonomia privada das empresas prestadoras de serviços de acesso condicionado para impor alguns parâmetros na definição do conteúdo a ser veiculado, fomentando certos tipos de discurso e desestimulando outros. Notadamente, fora estabelecido um percentual mínimo de programas nacionais e limites máximos para promoção comercial a fim de garantir a devida proteção à cultura nacional e ao consumidor. À época se considerou que a indicação de parâmetros mínimos ou limites máximos seriam ferramentas proporcionais à adequação da programação a interesses sociais relevantes à luz da Constituição, sem que isto implicasse ofensa à liberdade de expressão e à autonomia privada. Em sentido contrário, o art. 25 da norma fora declarado inconstitucional por conferir uma preferência absoluta e, portanto, desproporcional a empresas de publicidade nacional. (ADI 4923/DF, Tribunal Pleno, Min. Rel. Luiz Fux, Dje: 08/11/2017).

consensos solidamente estabelecidos e justificáveis acerca dos riscos graves de dano que certas espécies de discurso podem produzir à democracia e à sociedade como um todo, e a inexistência de meios menos onerosos capazes de lidar com esta hipótese de dano, hipótese na qual caberá ao Poder Judiciário realizar um escrutínio mais robusto sobre tais razões.

(IV) Os atos regulatórios passíveis de restrição à liberdade de comunicar devem, tanto quanto possível, pautarem-se por critérios objetivos e transparentes, evitando o uso de expressões vagas ou ambíguas, de modo a mitigar os riscos de efeito resfriador.

(V) Deve se dar preferência às regulações que orientem a “arquitetura de escolhas” apresentadas por agentes privados da área de comunicação – por meio do fomento a certos tipos de conteúdo e expressões ou disponibilização de um espectro mais amplo de informação ao destinatário do conteúdo — em detrimento de medidas que proíbam, peremptoriamente, determinados discursos.

É importante destacar, no entanto, que as premissas consideradas acerca da regulação de arquiteturas analisadas acima tiveram por base, especialmente, precedentes ligados a serviços de radiodifusão e difusão de sons e imagens, que não se confundem com a atividade econômica dos provedores de redes sociais. Há diferenças substanciais entre as atividades, razão pela qual não se defende uma analogia regulatória para circunstâncias distintas.

Decerto, o ponto que as aproxima é o fato de serem canais privados e essenciais para a circulação de opiniões e ideias na esfera pública, exercendo profunda influência sobre a cultura da sociedade e o exercício do poder político, mas há diferenças substanciais a serem pontuadas. Primeiro, porque os serviços de radiodifusão sonora e de sons e imagens, embora prestado por particulares, estão sujeitos a um regime de concessão pública¹³² o que, inegavelmente, confere maior ingerência ao poder estatal¹³³ se comparado às atividades de provedores que operam redes sociais, espécie de atividade econômica em sentido estrito que independe de qualquer espécie de concessão. A autonomia privada é mais ampla no segundo caso. Segundo, porque nas redes sociais o controle de conteúdo ocorre, por vezes, posteriormente à divulgação da expressão ou comunicação por meio de um processo substancialmente mais complexo (conhecido como “moderação”), quase sempre orientado por inteligência artificial, atividade que inexistente no caso dos veículos de mídia tradicional onde o controle é de ordem editorial e usualmente realizado por seres humanos, na definição daquilo que será ou não publicado. Logo, os riscos decorrentes de cada procedimento são distintos.

Tais diferenças, embora não invalidem o uso dos precedentes acima como orientadores de um parâmetro quanto a modelos regulatórios menos invasivos, reforçam que uma resposta

¹³² O regime de concessão é justificado pela escassez de cabos físicos e ondas de radiofrequência, o que impõe uma limitação ao número de empresas aptas a explorar os serviços de radiodifusão sonora e de sons e imagens.

¹³³ Neste sentido, RE 1.026.923/DF, Redator do Acórdão: Min. Alexandre de Moraes, Tribunal Pleno, DJe de 24/02/2021. O precedente em questão reconheceu a constitucionalidade na obrigatoriedade de retransmissão do programa “A Voz do Brasil”, pautado no vetusto argumento de supremacia do interesse público. Tal fundamento é dificilmente sustentável para viabilizar a restrição a uma atividade econômica em sentido estrito.

eficiente no campo das redes sociais requer também compreender como se opera seu funcionamento e quais são as principais contribuições e riscos que tais aplicações da *internet* geram para o exercício da liberdade de expressão.

Com efeito, é preciso analisar com maiores minúcias os aspectos atinentes ao ecossistema informacional da *internet* e, mais precisamente, as redes sociais, para melhor entender por que devemos regulá-la e como podemos fazer isso sem violar as premissas gerais acima apresentadas. Esta será a tarefa dos próximos capítulos.

2 A LIBERDADE DE EXPRESSÃO NO SÉCULO XXI: NOVOS DESAFIOS E RISCOS

2.1 A liberdade de expressão nas camadas da *internet*: da neutralidade à imunidade dos intermediários

Muito já foi dito sobre o desenvolvimento da *internet*, de modo que não se irá aprofundar os aspectos técnicos evolutivos desde sua origem na década de 1960, a partir de um projeto de pesquisa militar norte-americano,¹³⁴ até sua privatização na década de 90, que possibilitou seu uso para finalidades diversas como pesquisas, comércio e informação, tornando-se um elemento fundamental para o dia a dia dos indivíduos.

Interessa, em especial, saber como o tema liberdade de expressão e *internet* se entrecruza. Isto é, dimensionar o impacto da *internet* e, em um segundo momento, das redes sociais, sobre o exercício da liberdade de expressão. Neste ponto cabe rememorar a maior força da perspectiva libertária da liberdade de expressão no procedimento de expansão e desenvolvimento da *internet*.

O exemplo mais emblemático é dado no discurso de John Perry Barlow, em 1996, na Conferência de Davos, a denominada “Declaração de Independência do Ciberespaço”¹³⁵. Segundo o famoso discurso, a *internet* não seria passível de regulação, pois o ciberespaço seria um universo distinto do mundo físico, que desconheceria a autoridade estatal, tampouco as fronteiras convencionadas entre os Estados. O ciberespaço seria assim um espaço excepcional infenso a controles externos, cabendo a regulação do seu funcionamento aos seus próprios agentes.

O discurso é, em grande medida, influenciado pelo sentimento antiestatal que tomou o mundo com a extinção da União Soviética e a ascensão do movimento neoliberal na Europa e nos Estados Unidos a partir da década de 80 do século passado, mas traduz também parte das dificuldades a serem superadas para viabilizar a regulação de um sistema informacional privado amplamente descentralizado e, aparentemente, anárquico como a *internet*.

¹³⁴ Para aprofundamento, ver Keller (*Regulação Nacional de Serviços na Internet: Exceção, Legitimidade e o Papel do Estado*. Rio de Janeiro: Lumen Juris, 2019, p. 27-61).

¹³⁵ BARLOW, John Perry. A Declaration of Independence of Cyberspace. *Duke Law and Technology Review*, [s.l.], v. 18, n. 5-7, 2019.

O trabalho mais crítico a esta visão é de Lawrence Lessig¹³⁶, que desmistifica a ideia de que a *internet* seria um espaço absolutamente livre dos influxos de qualquer espécie de regulação.¹³⁷ Ao revés, o autor destaca que particulares podem ter suas condutas influenciadas não só por normas estatais, mas também por normas sociais, pelo mercado e, em especial, por arquiteturas. Os códigos e algoritmos, que garantem a operabilidade da rede mundial de computadores são as arquiteturas da *internet*. Tais códigos não necessariamente levariam a uma ampla liberdade do indivíduo face às amarras do poder estatal como preconizado por Barlow e demais *ciberlibertários*, já que a forma como são estruturados poderão produzir resultados bons ou ruins ao definir o que é permitido ou não no ciberespaço.

Em síntese, uma proposta puramente autorregulatória não necessariamente produzirá um cenário de maior liberdade na *internet*. Com efeito, a mesma arquitetura que pode impulsionar a liberdade em todas as suas acepções (de expressão, econômica e artística, por exemplo), pode ser utilizada como instrumento de controle, vigilância e ferramenta impulsionadora de ataques odiosos e desinformação. Neste caso, a regulação estatal pode ser necessária para redirecionar as arquiteturas da *internet* (códigos e algoritmos) a bons resultados. Nos dizeres de Lessig, “nenhum Direito, por vezes, produz códigos ruins”.¹³⁸

Isto não significa, no entanto, invalidar o mérito da iniciativa privada na construção das arquiteturas atuais da *internet*, que possibilitou não só sua expansão, mas uma verdadeira revolução no campo econômico e das liberdades individuais¹³⁹. De fato, é possível dizer que as arquiteturas que impulsionaram o crescimento da *internet* tornaram factível o ideal de “realização de uma cultura verdadeiramente democrática”¹⁴⁰, mediante um processo de revolução digital, que significou a “criação e disseminação de uma tecnologia que tornou

¹³⁶ A regulação por arquitetura consta originariamente de trabalho desenvolvido por Joel R. Reidenberg, embora o trabalho de Lessig seja o mais conhecido sobre o tema. Cf. REINDENBERG, Joel R. *Lex informatica: the formulation of information policy rules through technology*. *Texas Law Review*, Texas, v. 76, n. 3, 1998, p. 553-584, 1998.

¹³⁷ LESSIG, Lawrence. *Code and Others Laws of Cyberspace, version 2.0*. Nova Iorque: Basic Books, 2006, p. 1-9.

¹³⁸ *Ibid.*, p. 267. No original: “No law, in other words, sometimes produces bad code”.

¹³⁹ Nas palavras de Luís Roberto Barroso: “A Revolução digital de fato transformou profundamente a maneira como se realiza uma pesquisa, fazem-se compras de mercadorias, reserva-se um voo ou ouve-se música, para citar alguns exemplos. A sociedade contemporânea vive sob a égide de um novo vocabulário, uma nova semântica e uma nova gramática. A linguagem dos nossos dias inclui um conjunto de termos recém incorporados, sem os quais, no entanto, já não saberíamos mais viver. Para citar alguns: Google, Windows, Mac, Whatsapp, Telegram, Uber, Dropbox, Skype, Facetime, Facebook, Twitter, Instagram, Waze, Spotify, Amazon, Google maps, Google translator, iTunes, Netflix, YouTube (...)”. Cf. BARROSO, Luís Roberto. *Revolução tecnológica, crise da democracia e mudança climática: limites do direito num mundo em transformação*. *Revista Estudos Institucionais*, [s.l.], v. 5, n. 3, p. 1234-1313, 2019. Mimeografado.

¹⁴⁰ BALKIN, Jack M. *Digital Speech and Democratic Culture: A Theory of Freedom of Expression for the Information Society*. *N.Y.U. L. Rev.*, [s.l.], v. 79, n. 1, 2004, p. 39.

simples copiar, modificar, anotar, agrupar, transmitir e distribuir arquivos, por meio de seu armazenamento na forma digital”.¹⁴¹

Neste contexto, os agentes reconhecidos como “intermediários” da *internet* tiveram papel essencial no fomento à inovação na *internet*.

Segundo a Organização para Cooperação e Desenvolvimento Econômico (OCDE), os intermediários da *internet* são os agentes que reúnem e facilitam substancialmente a transação entre terceiros na *internet*, por meio da viabilização do “acesso, hospedagem, transmissão e indexação de conteúdo, produtos e serviços originados de terceiros na *internet* e de serviços oriundos da *internet* a terceiros.”¹⁴² Os intermediários não decidem sobre os dados, conteúdos, produtos ou serviços que passam por suas redes ou servidores, mas são o motor deste movimento.

O reconhecimento do positivo impacto econômico e cultural destes intermediários para a sociedade é fruto da consolidação de alguns princípios basilares que orientaram o processo de construção e expansão da *internet*, tendo por base um fim último: manter a *internet* como um ambiente seguro para um constante processo de inovação tecnológica, cultural e econômica.

Keller aponta duas características técnicas da *internet* que possibilitaram esta diretriz: a abertura da *internet* e o princípio da neutralidade das redes.¹⁴³ A abertura da *internet* significa que o ciberespaço depende da manutenção de padrões técnicos facilmente operáveis, que viabilizem a comunicação entre as redes que compõe a *internet*¹⁴⁴. A troca livre de informação entre estruturas é a condição operacional, que possibilita o reconhecimento da *internet* como ferramenta fomentadora da liberdade de expressão e informação.

Esta abertura decorre de um princípio conhecido como “neutralidade das redes”. A lógica por trás da neutralidade é de que provedores de acesso às redes não podem “discriminar conteúdos e serviços que se desenvolvem através de suas redes”¹⁴⁵, sob pena de se tornarem os mais poderosos censores do mundo.

¹⁴¹ BALKIN, Jack M. Digital Speech and Democratic Culture: A Theory of Freedom of Expression for the Information Society. *N.Y.U. L. Rev.*, [s.l.], v. 79, n. 1, 2004, p. 6.

¹⁴² ORGANIZAÇÃO PARA A COOPERAÇÃO E DESENVOLVIMENTO ECONÔMICO. *The Economic and Social Role of Internet Intermediaries*. [S.l.]: OECD, 2010, p. 9.

¹⁴³ KELLER, Clara Iglesias. *Regulação Nacional de Serviços na Internet: Exceção, Legitimidade e o Papel do Estado*. Rio de Janeiro: Lumen Juris, 2019, p. 40-45.

¹⁴⁴ Parâmetro seguido no art. 4º, inciso IV do Marco Civil da *Internet*.

¹⁴⁵ BALKIN, Jack M. O Futuro da Liberdade de Expressão na Era Digital. In: SAMPAIO, José Adércio Leite (Coord.). *Liberdade de Expressão no Século XXI*. Belo Horizonte: Del Rey, 2016, p. 351-374.

Para compreender adequadamente, o escopo da neutralidade utiliza a abordagem da *internet* por camadas¹⁴⁶. A *internet* é composta, simplificadamente, por duas camadas.¹⁴⁷ Uma é a camada física, onde se situa a infraestrutura física necessária ao funcionamento da *internet* (tais como os mecanismos de acesso à banda larga) e os padrões técnicos exclusivos da *internet* (os protocolos padrões da rede mundial de computadores conhecido como *http* — *Hypertext Transfer Protocol* — e *html* — *Hypertext Markup language*); e, por fim, há a camada de conteúdo, onde operam os aplicativos com os quais os usuários finais interagem diretamente (redes sociais, *blogs* e aplicativos de serviços diversos, por exemplo) e por meio do qual são prestados algum tipo de serviço ou informação a terceiros.¹⁴⁸

Quando se pensa em neutralidade das redes, busca-se garantir principalmente a livre circulação de dados na camada física da *internet*, isto é, os intermediários que viabilizam a conexão à *internet*¹⁴⁹. Por esse princípio, ressalvadas razões de caráter técnico-operacional, não é possível que os provedores de acesso à *internet* (integrantes da camada física) discriminem pacotes de dados de determinados aplicativos, sob pena de violação à liberdade de expressão, à livre-atividade econômica e à isonomia¹⁵⁰. Seu objetivo é ampliar o acesso à *internet* ao maior número de vozes possíveis. No Brasil, tal princípio é positivado pelos art. 3º, inciso IV¹⁵¹ e art. 9º do Marco Civil da *Internet*.¹⁵²

Este esclarecimento é relevante, pois permite superar uma confusão muito comum quando se aborda a temática dos limites ao “poder de governança” do discurso, no âmbito das redes sociais. Os limites do poder de governança de plataformas de conteúdo não se confundem com o debate sobre a neutralidade da *internet*.

¹⁴⁶ Cf. KELLER, Clara Iglesias. *Regulação Nacional de Serviços na Internet: Exceção, Legitimidade e o Papel do Estado*. Rio de Janeiro: Lumen Juris, 2019, p. 36-37

¹⁴⁷ Opta-se por esta abordagem por ser a que melhor dialoga com o Marco Civil da *Internet* e sua distinção entre “conexões de *internet*” e “aplicações da *internet*”. Há, no entanto, classificações mais minuciosas que mencionam até sete camadas na *internet*. Cf. *Ibid.*, p. 36.

¹⁴⁸ *Ibid.*, p. 37

¹⁴⁹ Art. 5º, inciso V da Lei nº 12.965/14.

¹⁵⁰ O tema da neutralidade guarda alguma polêmica em alguns países. Provedores de acesso à banda larga e grandes operadoras defendem que eventual cobrança diferenciada para o acesso de alguns aplicativos seria legítima, tendo em vista a maior/menor sobrecarga que algumas aplicações de *internet* podem acarretar para suas infraestruturas. Veja-se, portanto, que o debate embora relacionada à camada física pode repercutir na camada de conteúdo. A regra, embora ainda adotada na maior parte do mundo, caiu nos Estados Unidos em 2018, por força de decisão da Comissão Federal de Comunicações (FCC). Cf. EUA mudam regras da internet e favorecem grandes operadoras. *Tilt Uol*, [s.l.], 14 dez. 2017. Disponível em: <https://www.uol.com.br/tilt/noticias/redacao/2017/12/14/estados-unidos-decidem-pelo-fim-da-neutralidade-da-rede.htm>. Acesso em: 18 mar. 2022.

¹⁵¹ Lei nº 12.465/2017 – Art. 3º A disciplina do uso da *internet* no Brasil tem os seguintes princípios:

(...)

IV-preservação e garantia da neutralidade de rede;

¹⁵² Lei nº 12.965/2017 – Art. 9º O responsável pela transmissão, comutação ou roteamento tem o dever de tratar de forma isonômica quaisquer pacotes de dados, sem distinção por conteúdo, origem e destino, serviço, terminal ou aplicação.

O intermediário que opera na camada de conteúdo — denominada de “aplicações de *internet*” pela Lei nº 12.965/2014¹⁵³ — submete-se, portanto, à lógica diversa, podendo se ancorar na sua autonomia privada para dispor quais tipos de discursos, comportamentos e expressões serão aceitos, restringidos ou excluídos à luz de seus termos de uso e normas comunitárias aos quais os usuários aderem voluntariamente, sem que isto implique violação à ideia de neutralidade da *internet*.

A ampla autonomia conferida às aplicações de conteúdo da *internet* permite a expansão significativa destes serviços e isto se revela, em especial, pela atual dependência da sociedade contemporânea das redes sociais (*facebook, instagram, youtube, tik tok, twitter, telegram* etc.), plataformas que se apresentam como serviços de comunicação que buscam conectar pessoas, possibilitando o compartilhamento de informações de caráter pessoal, profissional, comercial ou de interesse público¹⁵⁴.

O debate sobre a intervenção regulatória estatal sobre a esfera de atuação destas aplicações decorre desta dependência. Modificações nos termos de uso ou falta de critérios claros na aplicação de suas regras podem produzir impacto resfriador sobre a esfera pública. Por outro lado, experencia-se um aumento significativo de discursos de ódio, assédio, *bullying* e desinformação no âmbito destas plataformas.

Nos Estados Unidos, construiu-se uma regra basilar para garantir a proteção da autonomia destes agentes econômicos de pretensões regulatórias estatais, a denominada regra da “imunidade dos intermediários”, consagrada na Seção 230 do *Communication Decency Act* (CDA). A importância desta regra é tamanha que a *Electronic Frontier Foundation* (EFF) a define como a mais importante norma para a proteção da liberdade no século XXI, vulgarmente a referenciando como as vinte e seis palavras que fundaram a *internet* (o número faz referência à quantidade de palavras que compõem a norma).¹⁵⁵

A regra se tornou, ainda, um produto de exportação norte-americano, pois fora consagrada como a primeira cláusula dos Princípios de Manila sobre a responsabilidade dos

¹⁵³ Neste sentido, dispõe o inciso VII do art.5º da Lei:

Art. 5º - Para os efeitos desta Lei, considera-se:

VII - aplicações de *internet*: o conjunto de funcionalidades que podem ser acessadas por meio de um terminal conectado à *internet*

¹⁵⁴ A relevância pública dos debates nas redes sociais é tamanha que está em discussão no Supremo Tribunal Federal se o Presidente da República pode bloquear indiscriminadamente perfis de terceiros após debate de natureza política na plataforma. O processo está concluso pendente a análise do Ministro Relator Alexandre de Moraes (MS 36648/DF).

¹⁵⁵ A EFF é a mais importante das associações privadas que atuam para a preservação dos direitos fundamentais — em especial, a liberdade de expressão e privacidade, no âmbito da *internet*. A informação sobre a importância fundamental da cláusula 230 do CDA está disponível em SECTION 230 of the Communications Decency Act. EFF, [s.l.], [2015]. Disponível em: <https://www.eff.org/issues/cda230>. Acesso em: 18 mar. 2022.

intermediários¹⁵⁶, conjunto de parâmetros criados na conferência *RightsCon*, realizada em 2015 nas Filipinas, com vistas ao estabelecimento de garantias e parâmetros quanto à responsabilidade de intermediários e liberdade de expressão nas redes, tendo também notória influência sobre o processo que culminou na edição do atual art. 19 do Marco Civil da *Internet*. Saber a história da regra da “imunidade dos intermediários” é, portanto, fundamental para desvendar as motivações que inspiraram o art. 19 do Marco Civil da *Internet*.

2.2 A Seção 230 do CDA

Compreender o alcance da regra da imunidade dos intermediários implica conhecer antes a história da Seção 230 do CDA.

O nascimento desta regra decorre de dois precedentes da Suprema Corte Norte-Americana. O primeiro caso, julgado em 1991, ficou conhecido como *Cubby Inc. vs Compuserve, Inc.* No precedente, a Suprema Corte declarou a ausência de responsabilidade da empresa *Compuserve*, tendo em vista que o intermediário nunca havia revisto qualquer conteúdo postado no fórum online. A atividade da empresa seria, portanto, de um mero distribuidor, um canal de comunicação entre as partes, que não atuava como *publisher*. Quatro anos depois, no julgamento do caso *Stratton Oakmont, Inc. vs Prodigy Services, Co.*, a Suprema Corte Norte-Americana entendeu em sentido diverso. A Corte asseverou que a empresa *Prodigy* era responsável por conteúdo difamatório, pois fazia uso de um software automático na avaliação dos conteúdos e possuía linhas gerais de conduta, que refletiam uma escolha consciente desta empresa de ter um controle editorial sobre aquilo que era postado, assumindo o risco por eventuais equívocos.¹⁵⁷

Em síntese, a orientação da Suprema Corte Norte-Americana criou um desincentivo à construção de arquiteturas responsivas à segurança no ambiente *online*, pois toda empresa que se prestasse a promover qualquer controle sobre aquilo que é postado, a fim de tornar a página

¹⁵⁶ Segundo tal cláusula, os intermediários devem ser protegidos por lei da responsabilização por conteúdos produzidos por terceiros. Embora indique tal regra como opção preferencial, a cláusula segunda dos Princípios de Manila admite alguma flexibilização deste parâmetro, admitindo a exclusão de conteúdo com fundamento em ordem judicial (opção que prevaleceu no Brasil). Informação disponível em PRINCÍPIOS de Manila sobre a responsabilidade dos intermediários. [S.l.]: EFF, 2015. Disponível em: https://www.eff.org/files/2015/07/02/manila_principles_1.0_pt.pdf. Acesso em: 18 mar. 2022.

¹⁵⁷ O histórico jurisprudencial que precedeu à criação da Seção 230 do CDA é detalhado por Kate Klonick em artigo seminal sobre o poder de governança das redes sociais (KLONICK, Kate. The New Governors: The People, Rules, and Processes Governing Online Speech. *Harv. L. Rev.*, [s.l.], v. 131, n. 1598, mar. 2017).

ou aplicativo mais seguro, poderia ser responsabilizada por eventuais casos de negligência ou excessos.

A reação legislativa ao último precedente veio em poucos meses. No início de 1996, foi votado no Congresso Norte-Americano uma lei que buscava regulamentar a obscenidade *online*, tornando ilegal o envio ou a apresentação deste conteúdo a menores (o CDA). Na sua redação final, fora incluída uma emenda, que declarava o seguinte: “*No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider*”¹⁵⁸. A cláusula ainda consubstancia, em uma de suas passagens o princípio do “bom samaritano”, que estende a imunidade às plataformas que optem por moderar, *de boa-fé*, conteúdos que considerem obscenos e ofensivos.¹⁵⁹

A despeito das inúmeras críticas que o CDA sofreu e do fato de muitas de suas disposições terem sido declaradas inconstitucionais pela Suprema Corte Norte-Americana, a Seção 230 não só permaneceu de pé como teve sua constitucionalidade certificada pelo tribunal.

Na visão da Suprema Corte, a Seção 230 não só era importante para tornar a *internet* um ambiente mais seguro, por meio da autorização para moderação de boa-fé (conforme cláusula do bom samaritano), como também era importante para evitar o *chilling effect*, ao afastar o risco de responsabilização por conteúdo ofensivo postado por terceiros. Na visão do Congresso Norte-Americano, admitir a responsabilização destas empresas por qualquer conteúdo produzido por terceiros geraria um natural incentivo à remoção de qualquer conteúdo.¹⁶⁰

Como já se antecipou, a cláusula da imunidade dos intermediários se tornou uma espécie de regra de ouro da liberdade de expressão no âmbito da regulação da responsabilidade dos intermediários que operam na camada de conteúdo das redes. Mais especificamente, as principais redes sociais do mercado exerceram (e ainda exercem)¹⁶¹ forte *lobby* para sua intacta

¹⁵⁸ Em tradução livre: “nenhum provedor ou usuário de um serviço de computador interativo deve ser tratado como editor de qualquer informação fornecida por terceiros”.

¹⁵⁹ No original, “*No provider or user of an interactive computer service shall be held liable on account of - any action voluntarily - taken in good faith to restrict access to or availability of material that the provider or user considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally protected*”. A íntegra da cláusula pode ser conferida no seguinte link: <http://www.columbia.edu/~mr2651/e-commerce3/2nd/statutes/CommunicationsDecencyAct.pdf>. Acesso em 20 mar. 2022.

¹⁶⁰ KLONICK, Kate. The New Governors: The People, Rules, and Processes Governing Online Speech. *Harv. L. Rev.*, [s.l.], v. 131, n. 1598, mar. 2017, p. 1607.

¹⁶¹ Neste sentido, cumpre destacar campanhas coordenadas das principais empresas gerenciadoras de plataformas das redes sociais contra projeto de lei em trâmite no Congresso Nacional, que impõe obrigações de transparência às plataformas (PL nº 2630/2020). Informação disponível em Paiva (PL das Fake News pode tornar plataformas menos seguras, critica Google. *Jota*, [s.l.], 11 mar. 2022. Disponível em: <https://www.jota.info/coberturas-especiais/liberdade-de-expressao/pl-das-fake-news-pode-tornar-plataformas-menos-seguras-critica-google-11032022>”. Acesso em: 21 mar. 2022.

manutenção ao redor do mundo. Com efeito, o dispositivo revelou-se um importante instrumento para que as aplicações privadas pudessem adaptar seus termos de uso às necessidades globais dos consumidores/usuários, garantindo-lhes um espaço considerável para a inovação e crescimento econômico sem que temessem qualquer tipo de efeito repressor do Estado sobre suas atividades, considerando inclusive a restritiva jurisprudência norte-americana sobre limites à liberdade de expressão.¹⁶²

Em termos simples, a Seção 230 do CDA consagrou a autorregulação como o *default* em matéria de regulação da liberdade de expressão nas redes sociais¹⁶³. Se há alguém que deva restringir o discurso no âmbito da *internet*, para torná-las um ambiente minimamente seguro, este alguém são as próprias empresas gerenciadoras das plataformas (os intermediários da camada de conteúdo) a quem cabe estabelecer os limites do aceitável por meio de seus termos de uso e demais normas comunitárias.

2.3 A imunidade dos intermediários no Brasil: a construção do art. 19 do Marco Civil da *Internet*

Até a vigência da Lei nº 12.965/14 — o denominado Marco Civil da *Internet* — o entendimento sobre os limites da responsabilização dos provedores da camada de conteúdo da *internet* sobre postagens realizadas por terceiros era tema bastante disputado na doutrina e jurisprudência nacionais.

Dois eram os entendimentos com maior proeminência na matéria. Uma primeira corrente defendia que a responsabilização do provedor de *internet* demanda uma mera notificação extrajudicial daquele que teve seu direito violado por uma postagem de terceiro. Trata-se da teoria do *notice and takedown*, inspirada na controversa legislação norte-americana de proteção dos direitos autorais (*Digital Millenium Copyright Act – DMCA*).¹⁶⁴ Para os

¹⁶² Neste sentido, vale a lembrança de Kate Klonick (tradução livre): “Reconhecer plataformas online como atores estatais e, portanto, obrigá-las a observar os direitos da primeira emenda de seus usuários, não somente conflitaria explicitamente com os propósitos da § 230, mas também criaria uma *internet* que ninguém quer. Plataformas não poderiam mais remover conteúdo violento ou obsceno (...) quase todo discurso seria expressamente permitido e protegido tornando significativamente piores problemas correntes como discurso de ódio, *bullying* e terrorismo (...)” (KLONICK, Kate. Op. cit., p. 1659).

¹⁶³ A construção da regra em 1996, vale dizer, é fortemente influenciada pelo apogeu do movimento ciberlibertário, cuja orientação fora exemplificada pela carta de Joe Perry Barlow.

¹⁶⁴ O DMCA é uma das exceções pontuais ao modelo de irresponsabilidade dos intermediários adotados pela legislação americana, por meio da cláusula 230 do CDA. Trata-se de uma hipótese de responsabilização especial de natureza subjetiva.

defensores desta linha, a prática possibilitaria uma atuação mais proativa das plataformas das redes sociais, viabilizando uma resposta equilibrada entre os temor de se impor a tais plataformas um dever geral de monitoramento e, de outro, possibilitar uma reparação integral dos danos, considerando os riscos decorrentes da própria atividade. Somente assim, seria conferida proteção integral ao consumidor à luz do art. 5º, inciso XXXII da Constituição Federal. Sua principal vantagem era evitar a necessidade de se recorrer a via judicial para assegurar a devida proteção de direitos na *internet*, pela velocidade da abrangência com que as violações se propagam.¹⁶⁵

Não é exagero afirmar que, apesar de algumas decisões do Superior Tribunal de Justiça oscilarem quanto à adoção desta teoria¹⁶⁶, havia uma quantidade substancial de precedentes daquela Corte que apontavam essa orientação como a mais adequada ao menos até o advento do Marco Civil da *Internet*.¹⁶⁷

Uma corrente adversa, fortemente inspirada pela cláusula 230 do CDA e pelo movimento de consolidação universal da regra da imunidade dos intermediários entendia, por outro lado, que a responsabilização de provedores da camada de conteúdo por eventual notificação extrajudicial não respondida poderia vir a produzir um indesejado efeito resfriador (*chilling effect*) do debate público na *internet*, o que iria de encontro com os fins que inspiraram a construção da arquitetura deste ecossistema informacional: a criação de um livre espaço para debate e criação, fundamental para a transformação da cultura democrática.

Assim, o mero temor de que qualquer notificação apresentada por um usuário, descontente com alguma publicação possa implicar responsabilização civil da plataforma junto ao Poder Judiciário, produziria o indesejado efeito de bloqueio de uma quantidade substancial de conteúdo por parte destes agentes privados, transformando tais canais em verdadeiros

¹⁶⁵ Por todos, cf. SCHREIBER, Anderson. Liberdade de expressão e Tecnologia. In: SCHREIBER, Anderson; MORAES, Bruno Terra de; TEFFÉ, Chiara Spadaccini (Coords.). *Direito e Mídia. Tecnologia e Liberdade de Expressão*. São Paulo: Foco, 2020, p. 1-29.

¹⁶⁶ Brasil. STJ. REsp 1.308.830/RS; Recurso Especial 2011/0257434-5; Relator(a): Ministra Nancy Andrighi; Órgão Julgador: Terceira Turma; Data do Julgamento: 08/05/2012; Data da Publicação/Fonte: DJe 19/06/2012.

¹⁶⁷ Brasil. STJ. REsp 1.323.754-RJ, Rel. Min. Nancy Andrighi, julgado em 19/6/2012; AREsp 983.194/SP; Agravo em Recurso Especial 2016/0242777-4; Relator(a): Ministro Marco Aurélio Bellizze; Decisão monocrática; Data da Publicação: 03/10/2016. No mesmo sentido, ver AREsp 982.469/RJ; Agravo em Recurso Especial 2016/0241619-7; Relator(a): Ministro Marco Aurélio Bellizze; Decisão monocrática; Data da Publicação: 07/10/2016. Confirmando a adoção do sistema de notificação extrajudicial, ver AREsp 893.764/RJ; Agravo em Recurso Especial 2016/0082068-2; Relator(a): Ministro Marco Buzzi; Decisão monocrática; Data da Publicação: 27/04/2016. Alguns precedentes traziam uma distinção entre provedores de conteúdo e de pesquisas. Neste último caso, como o provedor limitava-se a indexar e facilitar o acesso às páginas da rede mundial de computadores, entendia-se pela isenção de responsabilidade por não lhe caber o papel de “filragem de conteúdo”. Apontando nesta direção, REsp 1.316.921-RJ, Rel. Min. Nancy Andrighi, julgado em 26/6/2012.

censores. Condicionar, portanto, a responsabilização da plataforma a uma prévia ordem judicial reduziria o risco de *overblocking*.¹⁶⁸

Esta tese acabou se tornando vitoriosa após a edição da Lei nº 12.965/ 2014, tendo em vista o disposto no art. 19 da referida lei¹⁶⁹. O otimismo com o potencial inovador da *internet* e o receio de captura de silenciamento do espaço digital foi a tônica nos debates do Marco Civil da *Internet*, reconhecido internacionalmente como um *case* de sucesso, por se tratar de uma lei construída por meio de uma plataforma *online* aberta para contribuições e sugestões da sociedade civil. Uma catarse em meio ao movimento de encanto, que via na *internet* e suas aplicações instrumentos promotores da democracia e do livre debate.¹⁷⁰

A opção legislativa fora respaldada pelo Superior Tribunal de Justiça, que passara a condicionar, após a vigência do Marco Civil da *Internet*, a responsabilidade das plataformas pela retirada de conteúdo das aplicações de *internet* a prévia e específica ordem judicial.¹⁷¹ No âmbito das instâncias ordinárias há, contudo, ainda resistência na aplicação da regra, o que levou o tema à análise do Supremo Tribunal Federal¹⁷². A matéria, embora já tenha repercussão geral reconhecida, resta pendente de julgamento pela Corte Suprema.¹⁷³

Com efeito, embora o diploma legislativo não tenha impedido os provedores de *internet* de suprimirem conteúdo que viole suas políticas de uso, cria uma estrutura normativa que

¹⁶⁸ Cf. LEMOS, Ronaldo; SOUZA, Carlos Affonso de. A Bill of Rights for the Brazilian Internet ('Marco Civil'): A Multistakeholder Policymaking Case. In: URS, Gasser; SCHULZ, Wolfgang (Orgs.). *Governance of Online Intermediaries: observations from a Series of National Case Studies*. Cambridge, MA: Berkman Center Research Publication, 2015. Também nesta linha: Id. Marco Civil da Internet: construção e aplicação. Juiz de Fora: Editar, 2016.

¹⁶⁹ Art. 19. Com o intuito de assegurar a liberdade de expressão e impedir a censura, o provedor de aplicações de *internet* somente poderá ser responsabilizado civilmente por danos decorrentes de conteúdo gerado por terceiros se, após ordem judicial específica, não tomar as providências para, no âmbito e nos limites técnicos do seu serviço e dentro do prazo assinalado, tornar indisponível o conteúdo apontado como infringente, ressalvadas as disposições legais em contrário.

¹⁷⁰ Para mais detalhes sobre o procedimento de aprovação da lei, confira-se: TEFFÉ, Chiara Antonia Spadaccini. 2015. A Responsabilidade Civil do Provedor de Aplicações de Internet Pelos Danos Decorrentes do Conteúdo Gerado por Terceiros, de Acordo com o Marco Civil da Internet. *Revista Fórum De Direito Civil*, Belo Horizonte, ano 4, n. 10, set./dez. 2015. Disponível em: <https://www.editoraforum.com.br/wp-content/uploads/2015/12/A-responsabilidade-civil-do-provedor-de-aplicacoes-de-internet.pdf>. Acesso em: 26 abr. 2022.

¹⁷¹ Brasil. STJ. REsp 1.641.133/MG; Recurso Especial 2016/0218229-7; Relator(a): Ministra Nancy Andrighi; Órgão Julgador: Terceira Turma; Data do Julgamento: 20/06/2017; Data da Publicação/Fonte: DJe 01/08/2017. Com o mesmo teor, ver REsp 1.531.653/RS; Recurso Especial 2015/0108398-4; Relator(a): Ministra Nancy Andrighi; Órgão Julgador: Terceira Turma; Data do Julgamento: 13/06/2017; Data da Publicação/Fonte: DJe 01/08/2017; REsp 1.629.255/MG; Recurso Especial 2016/0257036-4; Relator(a): Ministra Nancy Andrighi; Órgão Julgador: Terceira Turma; Data do Julgamento: 22/08/2017; Data da Publicação/Fonte: DJe 25/08/2017; e REsp 1.642.997/RJ; Recurso Especial 2016/0272263-4; Relator(a): Ministra Nancy Andrighi; Órgão Julgador: Terceira Turma; Data do Julgamento: 12/09/2017; Data da Publicação/Fonte: DJe 15/09/2017.

¹⁷² A respeito, confira-se Fábio Carvalho Leite (Por que juízes não aplicam o art. 19 do marco civil da Internet? *PLEB*, [s.l.], 22 jul. 2019. Disponível em: <https://www.plebuc.science/post/por-que-ju%C3%ADzes-n%C3%A3o-aplicam-art-19-do-marco-civil-da-internet>. Acesso em: 26 abr. 2022.

¹⁷³ O tema está pendente de análise em dois recursos extraordinários. Um que trata de situação anterior à vigência da lei (RE nº 1.057.258/MG) e outro, posterior, no qual se suscita a declaração de inconstitucionalidade do art. 19 do Marco Civil da *Internet* (RE 1.037.396/SP-RG, Min. Rel. Dias Toffoli).

claramente desincentiva qualquer tipo de controle nas mídias digitais, visto que somente serão responsabilizados por eventual conteúdo nocivo não retirado se, após ordem judicial, não adotarem as providências necessárias.

Mais do que avaliar a constitucionalidade do art. 19 do Marco Civil da *Internet* ou ainda compará-la com a orientação jurisprudencial predominante até o advento do MCI, é imperioso destacar que tais regras — ao conferir protagonismo ao Poder Judiciário — se apresentam como insuficientes para lidar com os desafios que a livre circulação de informações e pensamentos pelas redes sociais tem suscitado na contemporaneidade.

Não se coaduna aqui com a tese de que o modelo brasileiro seria substancialmente distinto do sistema de irresponsabilidade dos intermediários adotado nos Estados Unidos, aproximando-se de um regime corregulatório. Como se verá com maior profundidade adiante, a “corregulação” pressupõe algo mais do que a possibilidade de intervenção judicial, qual seja “a conjugação da autorregulação privada destes agentes com alguma vinculação à legislação primária (a instituição de uma base principiológica, a supervisão estatal de órgão co ou autorregulador ou, ainda, a previsão de mecanismos de transparência e *accountability*, que vinculem agentes privados)”.¹⁷⁴

Ademais, a adoção integral do modelo norte-americano — de irresponsabilidade plena — nem poderia ser adotada por aqui, pois retiraria, em absoluto, do usuário prejudicado a possibilidade de recorrer ao Poder Judiciário contra o provedor, em clara afronta ao art. 5º, inciso XXXV da Constituição Federal, que consagra o direito de acesso à justiça. Com efeito, é razoável dizer que a norma do art. 19 do MCI consagra o regime autorregulatório possível à luz do ordenamento jurídico nacional.

Um regime verdadeiramente *corregulatório* — pelo menos na acepção mais estrita que se adota neste trabalho¹⁷⁵ — parte da premissa de que seu papel é orientar a autorregulação privada, especialmente no que tange a construção das arquiteturas das redes sociais e na fixação de critérios para exercício da moderação em parâmetros aceitáveis à luz da Constituição, possibilitando o melhor funcionamento do mercado de ideias e da arena democrática no espaço digital.

Isto não é possível por meio da judicialização, na forma do art. 19 do Marco Civil da *Internet*. Como bem explica Keller, há duas limitações centrais nesta abordagem que mais aproximam o modelo brasileiro de um sistema puro de autorregulação privada do que um

¹⁷⁴ KELLER, Clara Iglesias. *Regulação Nacional de Serviços na Internet: Exceção, Legitimidade e o Papel do Estado*. Rio de Janeiro: Lumen Juris, 2019, p. 189.

¹⁷⁵ O ponto será aprofundado no próximo capítulo.

modelo híbrido de construção de política pública.¹⁷⁶ A principal é a seletividade da atuação do Poder Judiciário.

A quantidade descomunal de pedidos de remoção e questões relacionadas ao debate sobre a liberdade de expressão nas redes sociais não é compatível com a estrutura que o Judiciário dispõe para realizar tais análises. Assim, não só se gera prejuízo à isonomia, pois os casos que tendem a ganhar destaque costumam ter maior relação com pessoas com maior poder aquisitivo (não raro, políticos);¹⁷⁷ mas, principalmente, fornece escrutínio público somente sobre uma quantidade muito limitada da esfera de regulação de conteúdo *online*, deixando muito espaço para uma regulação privada exercida ao arbítrio da empresa que administra a rede social.

Segundo, embora o Poder Judiciário detenha a legitimidade necessária para aferir com independência e imparcialidade eventuais conflitos relacionados aos limites externos à liberdade de expressão, possui capacidade institucional limitada para aferir se as redes sociais estão adotando dentro “dos limites técnicos do seu serviço” as medidas adequadas para conferir maior transparência à sua política de curadoria e moderação de conteúdo. Esta incapacidade para atuar de forma sistêmica fica evidente, quando se observa que o Supremo Tribunal Federal se tem, não raro, socorrido de mecanismos pouco ortodoxos para viabilizar soluções a distorções no exercício da liberdade de expressão nestas redes sociais. Esta atuação, além de muitas vezes deficiente, tem causado imenso desgaste na imagem de imparcialidade do Tribunal.¹⁷⁸

Com efeito, ainda que o art. 19 do Marco Civil da *Internet* tenha louváveis méritos ao buscar evitar um resfriamento do debate público no âmbito da *internet* e das redes sociais, conferindo ao judiciário o papel de árbitro destes conflitos e prestigiando a primazia *prima facie* da liberdade de expressão (conforme jurisprudência do STF), é certo que seu escopo se revela

¹⁷⁶ KELLER, Clara Iglesias. Policy by judicialisation: the institutional framework for intermediary liability in Brazil. *International Review of Law, Computers and Technology*, [s.l.], 13 jul. 2020. DOI: <https://doi.org/10.1080/13600869.2020.1792035>. Acesso em: 26 abr. 2022.

¹⁷⁷ Levantamento do GEPI-FGV relativo ao ano de 2014 indica não só altos percentuais de participação dos Tribunais Eleitorais na regulação da liberdade de expressão, como taxas de deferimento de pedidos de indenização e remoção de conteúdo realizados nessa eleição sempre acima de 50%, o que tem estimulado, inclusive, a tática recorrente de políticos de mobilizarem direitos da personalidade, mesmo fora de períodos eleitorais, para pleitear indenizações e remover discursos críticos — especialmente da *internet* —, numa busca por manutenção contínua da sua reputação. Cf. CRUZ, Francisco Brito. *Novo Jogo Velhas Regras: Democracia e Direito na era da nova propaganda política e das fake news*. Belo Horizonte: Casa do Direito, 2020, p. 344.

¹⁷⁸ Tome-se, por exemplo, o enorme ônus argumentativo assumido pelo Supremo Tribunal Federal na condução do famigerado inquérito das *fake news*, principal mecanismo existente hoje de escrutínio público sobre as políticas de combate à desinformação e discursos de ódio adotadas pelas redes sociais. O tema é abordado com maior profundidade no último capítulo.

insuficiente a nível sistêmico, pois o Judiciário é incapaz, por si só, de orientar minimamente as plataformas privadas quanto aos parâmetros procedimentais e materiais que devem pautar sua atuação.

É necessário, portanto, um agir regulatório consciente para que se tenha um melhor endereçamento dos novos desafios impostos ao exercício da liberdade de expressão em uma esfera pública predominantemente virtual. Antes, contudo, é preciso definir qual seria o escopo desta regulação pública. Isto é, o que torna a moderação destas redes insuficiente? E qual seria a devida medida de uma regulação pública neste campo a fim de evitar que caminhemos de um contexto de abuso e distorções no exercício da liberdade de expressão nas redes para um cenário de censura e autoritarismo estatal?

Para enfrentar tal questão, recorre-se à teoria de Balkin, que concebe a liberdade de expressão no século XXI como uma relação triangular, no qual interagem sociedade civil, Estado e os intermediários da *internet*.

2.4 A teoria triangular da liberdade de expressão

A doutrina clássica liberal foi cunhada sob uma visão dicotômica que opõe, de um lado, o Estado e, de outro lado, a sociedade civil, que depende, em grande parte, dos meios de comunicação de massa para a construção de uma opinião pública. Neste cenário, a esfera pública é um espaço fortemente influenciado pela visão estabelecida por grandes grupos de mídia. O acesso a estes canais comunicativos e a capacidade de influenciar a opinião pública é obtido, geralmente, por meio de grupos e organizações formais, tais como sindicatos, universidades e partidos políticos, havendo pouco espaço para manifestações individuais.

No século XXI, a esfera pública se modifica. Cada vez mais, a formação da opinião pública é dependente dos espaços construídos por infraestruturas privadas de acesso, protagonistas de uma esfera pública virtual e altamente fragmentada, no qual cada indivíduo pode contribuir de forma única para um fluxo de informações acelerado e, por vezes, amplificado, no ecossistema informacional privado da *internet*.¹⁷⁹

Há uma diferença substancial no funcionamento destas novas infraestruturas privadas se comparada aos tradicionais canais de mídia, que se consolidaram ao longo do século XX

¹⁷⁹ VESTING, Thomas. A Mudança da Esfera Pública pela Inteligência Artificial. Trad. Pedro Henrique Ribeiro. In: ABOUD, Georges; CAMPOS, Ricardo; NERY JR., Nelson (Orgs.). *Fake News e Regulação*. 2. ed. São Paulo: Revista dos Tribunais, 2020. p. 193-212.

(emissoras de rádio e televisão, jornais e revistas). Ainda que lá já houvesse a migração para um modelo de *economia da atenção* — ante a maior facilidade de difusão de opiniões e informações por todo —, o público em geral possuía um papel passivo, sendo mera audiência desta mídia, mais do que um criador de conteúdo, que tem livre acesso a estes canais para se comunicar com terceiros.¹⁸⁰ A definição daquilo que seria ou não publicado exigia um controle editorial da própria empresa, conforme critérios que julgasse apropriado “para uma audiência imaginada de adultos médios e suas famílias”.¹⁸¹

As redes sociais, por sua vez, caracterizam-se pela livre possibilidade de todos produzirem seu conteúdo, tornando-se ao mesmo tempo comunicador e audiência. A rede social não publica seu próprio conteúdo — como a mídia tradicional —, mas sim publica o conteúdo de qualquer um que postar em suas plataformas. Há também um processo de curadoria de conteúdo nas redes sociais, mas este é significativamente mais complexo, sutil e dispendioso do que o controle editorial realizado pelas mídias tradicionais, o que requer a construção de verdadeiras burocracias administrativas e tecnológicas aptas a viabilizar este controle. Isso se deve, em parte, pela própria opção normativa de conferir ampla autonomia gerencial a estes intermediários, mas decorre principalmente da maior complexidade de se disciplinar a liberdade discursiva em um ambiente que apresenta características como: (I) um incomensurável fluxo de informações e conteúdo, (II) cuja velocidade de disseminação pode ser de segundos, (III) muitas vezes impulsionados por robôs e perfis falsos, e que exige (IV) a necessidade de análises sofisticadas, uma vez que o caráter global destas plataformas pode implicar soluções regulatórias diferenciadas a depender do contexto cultural e linguístico subjacente. Tal cenário demandou a construção de um aparato regulatório e tecnológico interno muito mais sofisticado que o controle editorial feito pelas principais mídias de massa.

Este é o elemento distintivo na atuação destas plataformas privadas. Pela primeira vez, há burocracias não estatais aptas a monitorar e condicionar o exercício da liberdade discursiva

¹⁸⁰ Esta, aliás, é uma crítica trazida por parte da doutrina para justificar a necessidade de democratização dos meios de comunicação de massa. Neste sentido, afirma Fábio Konder Comparato: “(...) A lógica da atividade empresarial, no sistema capitalista de produção, funda-se na lucratividade, não na defesa da pessoa humana. Uma organização econômica voltada para a produção do lucro e sua ulterior partilha entre capitalistas e empresários não pode, pois, apresentar-se como titular de direitos inerentes à dignidade da pessoa humana. Ora, a liberdade de expressão não se confunde com liberdade de exploração empresarial nem é, de modo algum, garantida por ela. Constitui, pois, uma aberração que os grandes conglomerados do setor de comunicação de massa invoquem esse direito fundamental à liberdade de expressão, para estabelecer um verdadeiro oligopólio nos mercados, de forma a exercer, com segurança, isto é, sem controle social ou popular, uma influência dominante sobre a opinião pública” (COMPARATO, Fábio Konder. *A Democratização dos Meios de Comunicação de Massa*. In: GRAU, Eros Roberto; GUERRA FILHO, Willis Santiago (Orgs.). *Direito Constitucional: Estudos em Homenagem a Paulo Bonavides*. São Paulo: Malheiros, 2001, p. 149-166).

¹⁸¹ BALKIN, Jack. How to regulate social media. *Journal of Free Speech Law*, [s.l.], v. 1, n. 71, nov. 2019. Mimeografado.

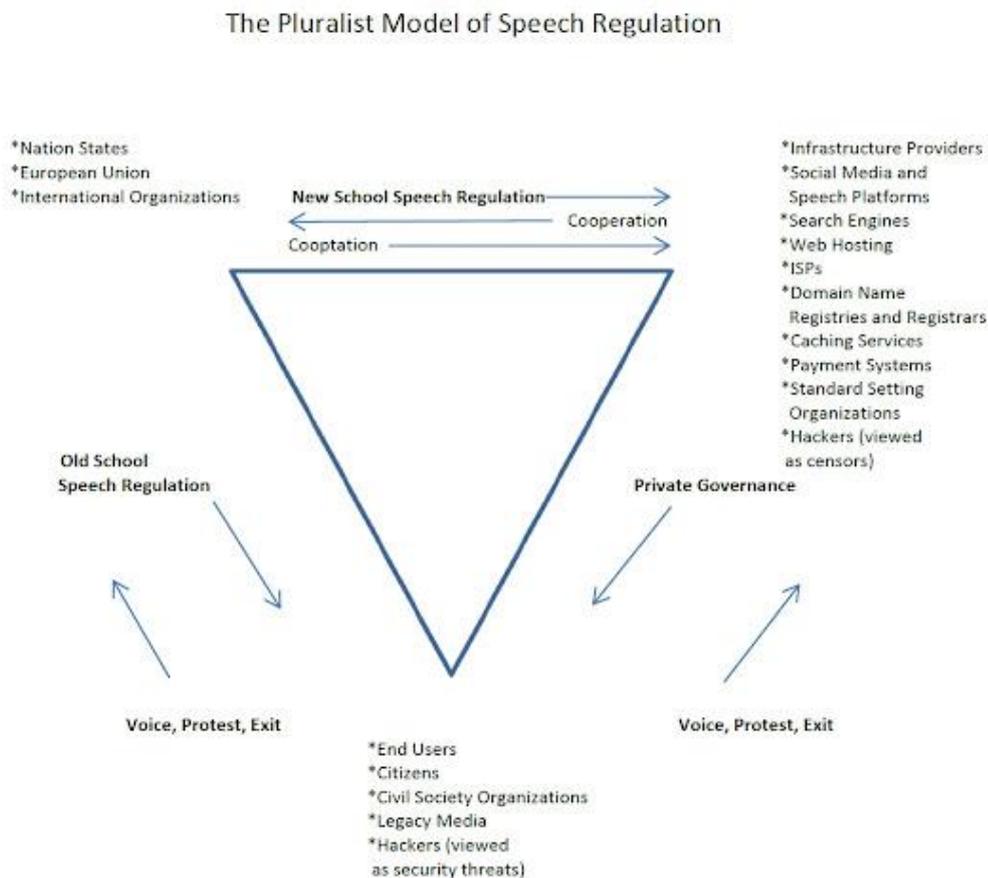
de forma ampla na sociedade. Daí diversos autores mencionarem um “poder de governança” das plataformas privadas de redes sociais, que seria composto por um órgão central; um corpo estabelecido de regras (os termos de uso e normas comunitárias); mecanismos de aplicação destas regras tanto de forma prévia quanto supressiva (moderação *ex ante* e *ex post*); procedimento para julgamento de alegações de violação aos seus termos de uso e, ainda, valores democráticos declarados (usualmente nas normas comunitárias ou no próprio termo de uso), que podem ser modificados ou atualizados em razão de *inputs* externos, tais como pressões de usuários-cidadãos, do Estado ou de grupos de terceiros interessados¹⁸².

Jack Balkin descreve, assim, os intermediários da *internet* como elemento central de uma nova teoria da liberdade de expressão.¹⁸³ A liberdade de expressão não deve mais somente ser analisada da perspectiva relacional entre Estado e sociedade civil. Há um terceiro sujeito que passa a influenciar nessa configuração dada sua ampla autonomia e capacidade de monitoramento sobre os discursos em sua esfera, moldando em boa medida o debate público: os intermediários digitais, dentre os quais se destacam as redes sociais, estruturas privadas da *internet* capazes de governar discursos e monitorar expressões e comportamento por meio de aparato tecnológico e burocrático próprio. Estes sujeitos se relacionariam com a sociedade civil e Estado em uma relação jurídica triangular, no qual todos são capazes de se influenciar mutuamente.

¹⁸² KLONICK, Kate. The New Governors: The People, Rules, and Processes Governing Online Speech. *Harv. L. Rev.*, [s.l.], v. 131, n. 1598, mar. 2017, p. 1662.

¹⁸³ BALKIN, Jack. Free Speech in the Algorithmic Society: Big Data, Private Governance and New School Speech Regulation. *UC Davis Law Review*, Davis, v. 51, n. 3, p. 1149-1210, fev. 2018. Mimeografado.

Figura 1 – O desenho triangular da liberdade de expressão



Fonte: J. Balkin, 2018.

Esta nova teoria, por certo, não rompe em definitivo com todos os paradigmas trazidos pela teoria clássica da liberdade de expressão, fundada em um modelo dual de oposição entre Estado e sociedade, pois tanto as mídias tradicionais quanto outros grupos coletivos de interesse da sociedade civil ainda detêm capacidade de influir na esfera pública. O debate, contudo, certamente se tornou mais complexo, já que não só há novos instrumentos de censura à disposição do Estado, como a própria atividade prestada por empresas como Google ou Meta pode vir a viciar o processo de formação de opinião de seus usuários, produzindo efeitos sociais indesejáveis.

A importância da forma como os códigos e algoritmos destas plataformas influenciam o comportamento de seus usuários passa a ser o centro do debate regulatório das redes sociais. O debate da liberdade de expressão no século XXI demanda, portanto, entender a forma como tais estruturas privadas de tecnologias influenciam a sociedade civil e, ainda, como estas ferramentas privadas podem ser capturadas pelo Estado para fins de censura.

A explicação sobre tais riscos dividir-se-á em duas partes.

A primeira, relacionada ao temor de cooptação/captura destas infraestruturas privadas pelo Estado. Tais riscos remetem, em grande medida, ao receio de inclinação ao autoritarismo e de produção de efeito resfriador, abordados no capítulo primeiro. A dinâmica de como tal controle é exercido é nova, no entanto, pois as intervenções regulatórias potencialmente perigosas não necessariamente decorrerão diretamente do exercício do poder de polícia estatal. Mesmo a censura do Estado é capaz de operar por meio de agentes privados. Governantes podem, por vezes, recorrer ao aparato tecnológico das redes sociais para monitoramento com o propósito de privilegiar o seu próprio discurso.

Em um segundo momento, serão abordados os riscos que decorrem da própria atividade das empresas que gerenciam estas aplicações da *internet*, isto é, a forma como influenciam e capturam a opinião da sociedade civil por meio do seu poder de governança. São riscos inteiramente novos que decorrem da percepção de que o que é comentado ou debatido *online* produz consequências substanciais no mundo *offline*. Neste ponto, a arquitetura que as plataformas utilizam para amplificar ou reduzir o alcance de certos discursos e, ainda, a falta de transparência na prática da moderação ganham maior destaque.

2.5 Estados e redes sociais: compreendendo os novos tipos de censura

2.5.1 A censura colateral

Já são há muito conhecidos os mecanismos tradicionais dos quais o Estado se vale, no modelo dual, para impor restrições às liberdades de expressão, manifestação, pensamento, informação ou artística da sociedade civil. Já abordados no capítulo primeiro, temos como principais exemplos o condicionamento do exercício da liberdade de expressão a uma prévia autorização de uma autoridade governamental (fenômeno conhecido como “censura prévia”); o exercício do poder de polícia por órgãos governamentais que, não raro, apreendiam livros¹⁸⁴,

¹⁸⁴ Infelizmente, tal prática, embora bastante residual em um contexto democrático, acaba por vezes ocorrendo. Tome-se, por exemplo, decisão do ex-prefeito do Rio de Janeiro Marcelo Crivella, em 2019, que emitiu ordem administrativa para recolhimento de livros que tratam do tema da homotransexualidade no evento da Bienal/RJ, contando inclusive com chancela do Tribunal de Justiça do Estado do Rio de Janeiro. A ordem foi corretamente

jornais e revistas; e, em casos extremos, perseguições, torturas e assassinatos de jornalistas e artistas, que ocorreram com a leniência do Estado¹⁸⁵.

Mesmo sob a égide do Estado Democrático de Direito, o Supremo Tribunal Federal reconheceu que a censura poderia abranger não só tais práticas de controle, como atos manifestamente abusivos e desproporcionais oriundos dos Poderes Legislativo e Judiciário.¹⁸⁶

No contexto atual, o controle estatal sobre a liberdade de expressão pode se manifestar de novas formas que merecem ser conhecidas para se compreender as novas ameaças à liberdade de expressão no século XXI.

Uma delas é a censura colateral, que Balkin define numa acepção estritamente descritiva como quando o “Estado mira uma entidade A para regular o discurso de uma entidade B”¹⁸⁷. O objetivo deste tipo de censura é se valer do aparato regulatório público ou de pressões informais para induzir a empresa privada ao controle de um discurso em nome do Estado.

Alguns exemplos podem ser mencionados para melhor ilustração desta nova forma de cooptação pelo Estado da infraestrutura privada de intermediários da camada de conteúdo.

Primeiro, o caso mais evidente. Pouco após o início do conflito armado entre Rússia e Ucrânia, no final de fevereiro, o parlamento russo aprovou uma lei que tornara crime chamar de “guerra” o conflito entre os países ao invés de utilizar a expressão “operação militar especial”. Como consequência, não só inúmeros *blogs* jornalísticos foram bloqueados, como a página de dois importantes intermediários da camada de conteúdo (*facebook* e *twitter*).¹⁸⁸ O caráter manifestamente arbitrário da medida permite configurá-la como hipótese de censura e, no caso, seria colateral, pois pretende transferir o ônus fiscalizatório da medida aos intermediários privados. Uma vez que tais plataformas se recusaram a seguir a lei, acabaram sendo sancionadas com o bloqueio.

revertida no bojo da Suspensão de Segurança nº 1248/RJ, julgada pelo Ministro Dias Toffoli, que considerou a circunstância como uma hipótese de censura.

¹⁸⁵ O caso mais famoso é o do jornalista Vladimir Herzog. Para maiores informações a respeito, confira-se: CORTE INTERAMERICANA DE DIREITOS HUMANOS. *Caso Herzog e outros vs. Brasil*. Autores: Centro pela Justiça e o Direito Internacional e outros. Juiz Presidente: Eduardo Ferrer Mac-Gregor Poisot, 15 de março de 2018. Disponível em: https://www.corteidh.or.cr/docs/casos/articulos/seriec_353_por.pdf. Acesso em: 28 mar. 2022.

¹⁸⁶ Como bem destacou o Ministro Celso de Mello, no seu voto à ADPF nº 130, o conceito de censura “não se limita a atos de cerceamento da liberdade de informação emanados do Poder Executivo, abrangendo também aqueles provenientes dos Poderes Legislativo e do próprio Judiciário”. STF, ADPF 130, rel. Min. Carlos Britto, Pleno, Dje: 05.06.2009 apud GIACHETTA, André Zonaro. Atuação e Responsabilidade dos Provedores Diante das Fake News e da Desinformação. In: RAIS, Diogo (Coord.). *Fake News a conexão entre desinformação e o direito*. 2. ed. São Paulo: Thomsom Reuters Brasil, 2020.

¹⁸⁷ BALKIN, Jack. Free Speech in the Algorithmic Society: Big Data, Private Governance and New School Speech Regulation. *UC Davis Law Review*, Davis, v. 51, n. 3, p. 1149-1210, fev. 2018. Mimeografado.

¹⁸⁸ Informação disponível em: RÚSSIA criminaliza cobertura da guerra na Ucrânia e bane BBC, Twitter e Facebook. *Época Negócios*, [s.l.], 4 mar. 2022. Disponível em: <https://epocanegocios.globo.com/Mundo/noticia/2022/03/epoca-negocios-russia-criminaliza-cobertura-da-guerra-na-ucrania-e-bane-bbc-twitter-e-facebook.html>. Acesso em: 28 mar. 2022.

Mais interessante, contudo, é indagar em que medida a censura colateral é um potencial produtor de efeito resfriador (*chilling effect*) no debate sobre as redes sociais. Veja-se, portanto, que não se fala aqui de casos clássicos de “censura” em que se busca atribuir à rede social o papel de suprimir um conteúdo ou expressão em nome do Estado, mas avaliar em que circunstâncias a criação de novos conceitos jurídicos pode, por via indireta, levar a um resfriamento injustificado do debate social, quando tais elementos não são passíveis de identificação mais objetiva. Balkin costuma citar dois exemplos problemáticos sobre o tema.

O primeiro relaciona-se com a regulamentação do que se convencionou denominar “direito ao esquecimento”. A existência de um direito ao esquecimento fora reconhecida — pela primeira vez — na União Europeia, que ordenou ao Google que criasse um sistema de burocracia privada para decidir em primeira instância se um artigo específico pode ser desindexado dos resultados de pesquisa, de modo a proteger a intimidade e vida privada de um particular.¹⁸⁹

Outro exemplo é a lei alemã para aplicação da lei na *internet* (o *NetzDG*), uma lei que prevê a responsabilidade das plataformas de redes sociais para promoverem bloqueios de conteúdos considerados ilegais à luz da legislação alemã sobre a liberdade de expressão, observadas determinadas garantias procedimentais.¹⁹⁰ A avaliação do cumprimento da legislação caberá a um órgão administrativo estatal que poderá impor multas administrativas significativas em caso de descumprimento sistemático de seus comandos.

Balkin destaca que estes tipos de regulações trazem três consequências problemáticas¹⁹¹.

A primeira é um efeito indireto que tais legislações produzem sobre a capacidade de governança destas plataformas. Legislações ou decisões nacionais são limitadas pelo aspecto da soberania, não podendo transcender o território de um dado país. No entanto, a pressão que as burocracias estatais podem fazer sobre estas plataformas, com ameaças de multas ou bloqueios, acaba por provocar mudanças nos termos de uso destas comunidades, permitindo que visões locais sobre os limites à liberdade de se expressar, especialmente quando decorrentes de mercados mais atraentes economicamente para tais plataformas, acabem por se universalizar.¹⁹²

¹⁸⁹ Os detalhes do *Leading Case* — Google Spain S.L, Google Inc. y Agencia Española de Protección de Datos (AEPD), Mario Costeja González (C-131/12) - podem ser conferidos em: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A62012CJ0131>. Acesso em: 28 mar. 2022.

¹⁹⁰ Os méritos e pontos problemáticos da legislação alemã serão aprofundados em capítulo próprio.

¹⁹¹ BALKIN, Jack. Free Speech is a triangle. *Columbia Law Review*, [s.l.], v. 118, p. 2011-2056, mai. 2018, p. 2028.

¹⁹² A decisão monocrática recentemente prolatada pelo Ministro Alexandre de Moraes (Petição 9935/DF, Min. Alexandre de Moraes julgado em 17/03/2022), a pedido da Polícia Federal, serve de bom exemplo de como

A segunda é o risco de que tais regulações, ainda que aparentemente razoáveis, possam dar azo a um efeito resfriador excessivo quando sua aplicação pelas plataformas se dê de forma pouco transparente, sem que seja dado ao usuário a possibilidade de compreender e de contestar porque determinado conteúdo fora bloqueado (temática que se relaciona com os riscos de “censura privada”, que será analisada adiante).

A terceira é o risco de que normas muito vagas e genéricas criem nestas empresas um temor de penalização administrativa, podendo produzir um excesso de bloqueios — gerando prejuízo substancial ao livre debate — e, em casos mais graves, que a infraestrutura possa ser manipulada como um mecanismo de vigilância e coerção por parte dos Estados (como demonstrado no exemplo russo).

Tais consequências, entretanto, precisam ser sopesadas.¹⁹³ A intervenção do Estado sobre a esfera de responsabilidade destas plataformas só será ilegítima se não houver boas razões que a sustentem, ou quando a utilização de novos conceitos jurídicos puder trazer insegurança à sua observância pelas plataformas.

A consideração é importante, porquanto, na esteira das lições de Lessig, a forma mais eficiente de garantir a observância do direito fundamental à liberdade de expressão e demais direitos fundamentais no âmbito da *internet* é por meio da orientação das arquiteturas da rede utilizada pelos agentes privados a fins legítimos, já que o Estado não tem como, por si só, verificar eventuais atos atentatórios à legislação nacional praticados no âmbito destas plataformas de redes sociais, quanto mais da *internet*.

Em síntese, não há como fugir desta espécie de estratégia regulatória indireta. O efeito censor “colateral”, contudo, só irá advir se tal regulação estiver atrelada à ameaça de uma sanção ao intermediário da camada de conteúdo por descumprimento de uma norma ambígua

ameaças estatais podem ser um instrumento eficiente para viabilizar a conformação de plataformas de redes sociais à legislação local. No caso, a decisão que ameaçou a suspensão em âmbito nacional da plataforma *Telegram* fora eficiente não só para garantir a assunção de uma série de compromissos de combate à criminalidade por parte da empresa, como também viabilizou a aceitação pela empresa de um termo de adesão ao Programa Permanente de Enfrentamento à Desinformação no âmbito da Justiça Eleitoral, após meses de tentativas infrutíferas de contato por parte do Tribunal Superior Eleitoral. Após o atendimento das determinações do Poder Judiciário, a decisão fora revogada. Informação disponível em Fraga (Telegram assina acordo de adesão a programa do TSE. *Agência Brasil*, [s.l.], 25 mar. 2022. Disponível em: <https://agenciabrasil.ebc.com.br/justica/noticia/2022-03/telegram-assina-acordo-de-adesao-programa-do-tse>. Acesso em: 28 mar. 2022).

¹⁹³ Essa conclusão parece em conformidade com as próprias lições de Jack Balkin. Embora o autor defina a censura colateral, em um primeiro momento, como a forma de intervenção do Estado sobre a capacidade de governança dos intermediários para limitar alguma espécie de discurso, em outra parte do seu texto destaca a necessidade de regulação estatal como um necessário “contrapeso ao desenvolvimento de tecnologias de controle privado e vigilância” (BALKIN, Jack. *Free Speech in the Algorithmic Society: Big Data, Private Governance and New School Speech Regulation*. *UC Davis Law Review*, Davis, v. 51, n. 3, p. 1149-1210, fev. 2018. Mimeografado, p. 1218). Em síntese, o que o preocupa é a restrição regulatória que indevidamente restringe a capacidade discursiva dos usuários na *internet*, em especial nas redes sociais.

ou genérica, ampliando o risco de ocorrência do que os especialistas do direito digital denominam de *overblocking*.

No julgamento do Supremo Tribunal Federal sobre o direito ao esquecimento, por exemplo, embora o mote não fosse a responsabilidade dos intermediários digitais para desindexação de conteúdos, é evidenciado a todo tempo pela corte a preocupação com o potencial efeito censor decorrente de um direito genérico de ser esquecido.

Segundo a corte, o direito a ser esquecido pode ser compreendido como a “pretensão apta a impedir a divulgação, seja em plataformas tradicionais, seja em plataformas virtuais, de dados ou fatos verídicos e lícitamente obtidos, mas que em razão da passagem do tempo, teriam se tornado descontextualizados ou destituídos de interesse público relevante”¹⁹⁴. Este termo amplamente genérico, na visão da corte, não poderia ser evocado como subterfúgio para anular o direito à liberdade de expressão, o que não obsta que situações específicas possam ser consideradas como restrições razoáveis, desde que prevista em lei.¹⁹⁵ É preciso, contudo, que estejam muito bem delimitadas. Há, portanto, uma clara preocupação de que a categoria jurídica do direito ao esquecimento acabe se tornando um instrumento de censura colateral, e a resposta para este temor é a busca por definição mais estrita de hipóteses de sua incidência, sempre fundada em um amplo debate legislativo.

Conquanto o ponto da dissertação seja as plataformas de redes sociais, a dificuldade na delimitação do direito ao esquecimento é uma das perplexidades também encontradas quando se pensa na construção de um conceito de “desinformação” (vulgarmente referenciada como *fake news*), tendo em vista o esforço contínuo em se buscar significantes precisos, de modo a evitar que uma possível proposta regulatória sobre o tema possa se converter em medida

¹⁹⁴ Trecho do voto do Ministro Relator Dias Toffoli, no julgamento do RE 1010606/RJ, Tribunal Pleno, Julgado em 11/02/2021. p.32.

¹⁹⁵ Tal referência é expressa na ementa do acórdão supracitado (item 5, p.3). Preocupação similar consta da Declaración Conjunta Sobre la Independencia y la Diversidad de los Medios de Comunicación en la Era Digital, documento assinado em conjunto pela Organização das Nações Unidas (ONU), Organização dos Estados Americanos (OEA); Organização para Segurança e Cooperação na Europa (OSCE); Organização Africana sobre Direitos Humanos e dos Povos (CADHP), com a assunção de uma série de compromissos no âmbito de proteção da liberdade de expressão na *internet*, dentre os quais se destaca os limites à configuração de um direito ao esquecimento: *a eliminación o desindexación de contenido en línea de acuerdo con el llamado “derecho a ser olvidado” suscita importantes preocupaciones en relación con la libertad de expresión. Si tales medidas están previstas en el ordenamiento jurídico, los Estados deben asegurar que se dispongan por ley en términos claros y específicos, que se apliquen solamente en los casos en que el peticionario demuestre la existencia de perjuicios sustanciales para su privacidad que superen todo interés relacionado con la libertad de expresión, que estén supeditadas a garantías apropiadas del debido proceso y que se apliquen de una manera que, tanto desde el punto de vista procesal como sustantivo, respete plenamente el derecho a la libertad de expresión*”. Informação disponível em: ORGANIZAÇÃO DOS ESTADOS AMERICANOS et al. *Declaración conjunta sobre la independencia y la diversidad de los medios de comunicación en la era digital*. [S.l.]: [s.n.], 2018. Disponível em: https://www.ohchr.org/sites/default/files/Documents/Issues/Opinion/JointDeclaration2May2018_SP.pdf. Acesso em: 29 mar. 2022.

censura. Como se observará no capítulo terceiro, a construção de um conceito jurídico preciso de “desinformação” é fundamental para impedir que o justo debate regulatório seja capturado por governos de inclinações autoritárias, tornando-se uma ferramenta de censura colateral.

2.5.2 A censura reversa

Mais peculiar é a denominada técnica da censura reversa. Diferentemente do fenômeno da censura colateral, em que medidas regulatórias e ações estatais podem ser articuladas para tensionar os intermediários a suprimirem discursos em nome do Estado na *internet*; na censura reversa, agentes políticos se valem de uma retórica de proteção à liberdade de expressão e da estrutura algorítmica das redes sociais para disseminar mentiras e promover ataques contra terceiros que façam as vezes de opositores.

Se na censura colateral busca-se reduzir as possibilidades de discurso, na censura reversa o foco é na seleção do discurso que alcança o ouvinte. É o uso do aparato do Estado para inundar a esfera pública com uma quantidade imensa de informações para reduzir o alcance do discurso que lhe é desfavorável. Quando politicamente motivada, pode envolver a disseminação de desinformação ou ataques coordenados à reputação de críticos (*trolling* ou *cyberbulling*¹⁹⁶) para distrair ou desacreditar algum conteúdo ou notícia prejudicial ao governo de ocasião.¹⁹⁷

O uso deste instrumento por governos que não guardam tanto apreço pela democracia, como China, Rússia e Turquia¹⁹⁸, comprovam como este mecanismo tem sido eficiente na interdição de debates que possam suscitar críticas e questionamento aos detentores do poder.

A título de exemplo, o Governo chinês tem fabricado, em média, cerca de 448 milhões de *posts* em diversas redes sociais ao longo do ano. Este conteúdo tem sido produzido de duas formas distintas: mediante contratação de um exército de pessoas ou mediante uso de robôs (*bots*) que alimentam redes artificiais de popularidade, contribuindo para o impulsionamento

¹⁹⁶ Segundo Schreiber, *cyberbullying* é a “prática de intimidação sistemática a alguém por meio da *internet* ou tecnologias relacionadas. O *cyberbullying* consiste, em suma, na utilização do espaço cibernético para intimidar e hostilizar uma pessoa de modo continuado” (SCHREIBER, Anderson. *Cyberbullying: Responsabilidade civil efeitos na família. Carta Forense*, [s.l.], [2018]. Disponível em: www.cartaforense.com.br. Acesso em: 28 mar. 2022).

¹⁹⁷ WU, Tim. Is the First Amendment Obsolete? *Michigan Law Review*, [s.l.], v. 117, p. 547, 2018.

¹⁹⁸ *Ibid.*

de informações favoráveis ao governo.¹⁹⁹ Agências estatais e empresas russas, não raro, também têm sido acusadas de promover igual prática para desestabilizar eleições ao redor do mundo²⁰⁰ e desacreditar o uso de vacinas produzidas no Ocidente²⁰¹.

É curioso perceber como países com pouca ou nenhuma tradição democrática recorrem cada vez mais a estes expedientes de impulsionamento artificial de conteúdos falaciosos, odiosos e atrozes, quando têm à sua disposição ferramentas mais tradicionais de censura.

O baixo custo para inserção de informações nas redes sociais torna escassa a atenção que se direciona ao imenso fluxo de informações com qual nos confrontamos diariamente, portanto, se o intuito do Estado é calar um opositor, vale mais direcionar ataques sistemáticos a esta pessoa ou divulgar informações fraudulentas que coloquem em xeque a veracidade de uma acusação do que propriamente calá-lo. O que irá determinar o êxito da tentativa de “censura reversa”, neste caso, é a popularidade da informação, impulsionada por uma arquitetura algorítmica que viabilize a viralização do conteúdo fabricado.

Com efeito, mesmo para um governo pouco democrático, a tática se torna mais interessante, pois lhe permite alcançar um efeito similar ao da censura tradicional sem que lhe seja imputada a acusação de violação aos direitos humanos perante a comunidade política local e internacional, visto que o caráter censor da prática não é evidente.

A produção de falsidades ou ataques a terceiros por canais de governo não é uma novidade do século XXI. Regimes totalitários, não raro, valem-se do seu controle sobre canais de comunicação para disseminar mentiras. No âmbito nacional, o plano *Cohen* talvez seja o principal exemplar de falsificações da política nacional, tendo sido o elemento justificador da instauração do Estado Novo. Em todo caso, essas táticas vinham como prática subsequente à prévia perseguição de canais privados de imprensa e grupos de oposição. Primeiro, buscava-se calar a mídia independente, depois iniciavam-se as mentiras. Na era da esfera pública virtual, embora métodos tradicionais de censura ainda possam caminhar em conjunto com táticas de censura reversa (vide casos de China e Rússia), não há necessariamente tal correlação. É possível provocar instabilidades em regimes democráticos tão somente por meio da produção

¹⁹⁹ KING, Gary; PAN, Jennifer; ROBERTS, Margaret E. How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, Not Engaged Argument. *Am. Pol. Sci. Rev.*, [s.l.], v. 111, n. 484, 2017.

²⁰⁰ A título de exemplo: AGENTE da CIA diz que Putin interferiu nas eleições americanas de 2016. *Exame*, [s.l.], 10 set. 2019. Disponível em: <https://exame.com/mundo/russia-confirma-que-empregou-espiao-dos-eua-em-gabinete-presidencial/>. Acesso em: 16 dez. 2021. De forma mais detalhada, confira-se: GRAHAM, David A. The Age of Reverse Censorship. *The Atlantic*, [s.l.], 26 jun. 2018. Disponível em: <https://www.theatlantic.com/politics/archive/2018/06/is-the-first-amendment-obsolete/563762/>. Acesso em: 16 dez. 2021.

²⁰¹ Informação disponível em: AGÊNCIA tentou contratar youtubers para desacreditar vacinas contra covid. *Exame*, [s.l.], 26 jul. 2021. Disponível em: <https://exame.com/tecnologia/agencia-tentou-contratar-youtubers-para-desacreditar-vacinas-contracovid/>. Acesso em: 16 dez. 2021.

desenfreada de (des)informação. Normas que restrinjam substancialmente o poder de moderação das plataformas privadas de redes sociais podem, portanto, ser evocadas como subterfúgios para permitir que governos de ocasião possam livremente usar de tal estratégia censora, viralizando conteúdos falaciosos ou destruindo a reputação de seus rivais.

Um exemplo recente sobre o tema pode ser extraído da própria experiência nacional. No momento de maior tensão política e instabilidade democrática que o país passara nos últimos vinte anos, o Presidente Jair Bolsonaro concebeu a Medida Provisória nº 1068/2021, posteriormente substituída pelo Projeto de Lei nº 3227/2021, ainda em trâmite no Congresso Nacional.

A Medida Provisória nº 1068/2021, editada às vésperas do feriado da Independência, em um momento de tensão política, no qual o Supremo Tribunal Federal avançava nas investigações em curso no inquérito das *fake news*, que vinculam servidores e aliados do atual governo a uma campanha coordenada de desinformação e ataques às instituições, agentes políticos e juízes. As manifestações, no Sete de Setembro, foram uma resposta a tais investigações e traziam como pautas não só a defesa por um Poder Executivo forte, mas ataques a inúmeras autoridades de instituições democráticas, como o Poder Legislativo e o Supremo Tribunal Federal, havendo ainda a defesa explícita de uma intervenção militar.²⁰²

Em uma observação mais ampla, o movimento da edição da medida provisória se deu de forma calculada, visto que meses antes de sua edição o Presidente Bolsonaro já mencionava um decreto para regular as redes sociais e evitar que ele e seus aliados fossem “cerceados”²⁰³ de forma arbitrária por estas plataformas. A reclamação do presidente não se deu por acaso, já que ao longo do último ano teve diversas publicações etiquetadas ou excluídas pelas redes sociais.²⁰⁴ Do contexto, é possível perceber um descontentamento do governo com a forma

²⁰² CONTEXTO: as manifestações do 7 de Setembro. *GI*, [s.l.], 7 set. 2021. Disponível em: <https://g1.globo.com/politica/noticia/2021/09/07/contexto-as-manifestacoes-do-7-de-setembro.ghtml>. Acesso em: 17 dez. 2021.

²⁰³ CARVALHO, Lucas; MILITÃO, Eduardo. Bolsonaro cita decreto para regular internet, mas assessores negam. *Tilt Uol*, [s.l.], 5 mai. 2021. Disponível em: <https://www.uol.com.br/tilt/noticias/redacao/2021/05/05/bolsonaro-fala-em-regulamentar-marco-civil-da-internet-somos-cerceados.htm>. Acesso em: 17 dez. 2021.

²⁰⁴ A título de exemplo: BOLSONARO compartilha conteúdo falso, e Instagram oculta postagem do presidente. *GI*, [s.l.], 12 mai. 2020. Disponível em: <https://g1.globo.com/economia/tecnologia/noticia/2020/05/12/bolsonaro-compartilha-conteudo-falso-e-instagram-oculta-postagem-do-presidente.ghtml>. Acesso em: 20 dez. 2022; “CARNAVAL da esquerda”, diz Carlos após Instagram marcar posts como falsos. *Poder 360*, [s.l.], 8 nov. 2021. Disponível em: <https://www.poder360.com.br/brasil/carnaval-da-esquerda-diz-carlos-apos-instagram-marcars-posts-como-falsos/>. Acesso em: 20 dez. 2022; FACEBOOK remove live em que Bolsonaro relaciona vacina à Aids. *Tecmundo*, [s.l.], 25 out. 2021. Disponível em: <https://www.tecmundo.com.br/redes-sociais/227472-facebook-remove-live-bolsonaro-relaciona-vacina-aids.htm>. Acesso em: 20 dez. 2022. Frise-se que há acusações de insuficiência dessas políticas de moderação: RIBEIRO, Amanda. Falas de Bolsonaro sobre ‘tratamento precoce’ ferem regras de redes sociais, mas seguem no ar. *Aos Fatos*, [s.l.], 28 jun. 2021. Disponível em: <https://www.aosfatos.org/noticias/falas-de-bolsonaro-sobre-tratamento-precoce-ferem-regras-de-redes-sociais-mas-seguem-no-ar/>. Acesso em: 17 dez. 2021.

como as redes sociais têm exercido essa moderação, alegando um viés ideológico na abordagem.²⁰⁵

A Medida Provisória nº 1068/2021 teve como principal objetivo regular “os direitos e garantias dos usuários nas redes sociais”. Para tanto, insere alguns conceitos novos ao Marco Civil da *Internet* (Lei nº 12.965/2021), tais como “redes sociais” e “moderação em redes sociais”, e acresce uma Seção II ao Capítulo II (Dos Direitos e Garantias dos Usuários) da Lei para tratar especificamente dos direitos e garantias dos usuários de redes sociais.

Estes direitos e garantias são estabelecidos ao longo de quatro artigos (arts. 8-A ao 8-D) com inúmeros incisos e parágrafos. Dentre estes artigos, destacam-se em especial os arts. 8-B e 8-C, que restringem as possibilidades de exclusão, suspensão ou bloqueio de conta, perfil ou conteúdo gerado a um grupo de circunstâncias específicas, que configurariam “justa causa” para o exercício da moderação pelas redes sociais. Fora dessas hipóteses, o exercício da moderação seria ilícito.

A reação à Medida Provisória foi imediata, dando ensejo a inúmeras críticas no meio acadêmico e à profusão de ações diretas de inconstitucionalidade ajuizadas, em sua maioria, por partidos de oposição.²⁰⁶ Os argumentos foram vários. Indo da violação à liberdade econômica e autonomia das empresas, passando pela ausência de urgência no trato da matéria até a impossibilidade de se dispor sobre liberdades fundamentais e direitos políticos por medida provisória.²⁰⁷ Após, o Presidente do Senado Federal devolveu a medida provisória, o que levou ao fim da tramitação dos processos no STF, mas sem encerrar a questão, visto que o Projeto de Lei nº 3227/2021, de igual teor, passou a tramitar no Congresso Nacional.

O principal desafio desta norma é identificar como ela é capaz de produzir censura, pois o fundamento que a embasa é a “proteção à liberdade de expressão, comunicação e manifestação de pensamento” (vide *caput* dos novos arts. 8-B e 8-C, conforme projeto de lei).

O conceito de censura reversa torna-se fundamental, portanto, para correlacionar a restrição às possibilidades de moderação das redes privadas como um elemento censor. As campanhas coordenadas de “desinformação” configuram risco à democracia e ao livre exercício

²⁰⁵ Reportagem da Folha de São Paulo comprova, contudo, que as abordagens do Facebook também recaíram sobre o Presidente Venezuelano Nicolás Maduro e aliados (SOPRANA, Paula. Ação do Facebook contra rede ligada aos Bolsonaros também atingiu esquerda latina. *Folha de S. Paulo*, [s.l.], 10 jul. 2020. Disponível em: <https://www1.folha.uol.com.br/poder/2020/07/acao-do-facebook-contra-rede-ligada-aos-bolsonaros-tambem-atingiu-esquerda-latina.shtml>. Acesso em: 17 dez. 2021).

²⁰⁶ ADI 6996 (PDT), ADI 6995 (Partido Novo), ADI 6994 (PT), ADI 6993 (PSDB), ADI 6992 (Solidariedade) e ADI 6991 (PSB).

²⁰⁷ Os dois últimos argumentos foram acolhidos na decisão liminar da Ministra Rosa Weber (MC-ADI 6991/DF). O mérito acabou não sendo julgado pelo Plenário, haja vista a devolução da medida provisória pelo Presidente do Senado Federal.

da liberdade de expressão, eis que tal como qualquer espécie de censura, tem por objetivo silenciar opositores e críticos do governo, valendo-se não da óbvia restrição de conteúdo, mas da amplificação impulsionada por robôs e aliados de conteúdos que sejam favoráveis ao governo, ainda que isto implique difusão de mentiras e incentivos ao ódio e à polarização.

O revelador da lista constante nos arts. 8-B e 8-C é identificar, portanto, aquilo que dela não consta para se projetar os propósitos censores da norma. Não configuram justa causa para moderação conteúdos, por exemplos, declarações que configurem discriminação ou preconceito em razão de raça, cor, etnia, religião, procedência nacional (crime de racismo identificado no art. 20 da Lei nº 7716/89), homofobia ou transfobia (também reconhecido como crime pelo Supremo Tribunal Federal). Não há também menções a campanhas de desinformação e de ataque às instituições democráticas.²⁰⁸ Trata-se, portanto, de omissão que permite, por vias oblíquas, a circulação de boa parte dos conteúdos publicados quase que diariamente por aliados do atual governo.

Há ainda um ponto mais sensível bem lembrado pelo professor Carlos Affonso. O projeto de Lei nº 3327/2021 inclui um novo artigo (art. 109-B) na Lei de Direitos Autorais (Lei nº 9610/98), para prever a possibilidade de aplicação de sanção à rede social que tornar conteúdo protegido por direitos autorais indisponível sem “justa causa”. O esdrúxulo dispositivo cria, portanto, um artifício para constranger as redes sociais a seguirem o rol de conteúdos preconizados nos arts. 8-B e 8-C da legislação, por meio da ameaça do exercício do poder sancionatório por um órgão do Poder Executivo, desprovido de autonomia técnica. Nos dizeres do Carlos Affonso Pereira de Souza,²⁰⁹ a regra constrói o fundamento para a criação de um “Ministério da Mentira”. A respeito, confira-se:

(...) Como não existe nenhuma “justa causa” na MP para a remoção de conteúdos que são *fake news*, que mentem sobre temas tão distintos como tratamento precoce ou fraude de urnas eletrônicas, esse artigo — a partir de uma visão exótica dos direitos autorais — dá ao autor da publicação o direito de requerer ao “órgão responsável” (provavelmente a Secretaria Nacional de Cultura) que puna a rede social por ter removido seu conteúdo. As penas incluem advertências, suspensões e multas que podem chegar até dez por cento do faturamento da rede social no País (...) A MP inova ao permitir que, sob o pretexto de um direito autoral à mentira, à desinformação e ao ataque, o autor da publicação que foi excluída possa acionar o Governo para que a rede social seja punida e o conteúdo republicado. Esse artigo verdadeiramente cria o

²⁰⁸ Vale lembrar: quando da aprovação da Lei do Estado Democrático de Direito (Lei nº 14.197/2021), o Presidente Jair Bolsonaro vetou tipo penal que previa como crime a comunicação enganosa em massa identificada como a promoção ou financiamento de campanha ou iniciativa para se disseminar fatos que se sabem inverídicos e que sejam capazes de comprometer a higidez do processo eleitoral.

²⁰⁹ Opinião publicada em SOUZA, Carlos Affonso de. Análise: inconstitucional, MP de Bolsonaro cria ‘Ministério da Mentira’. *Tilt Uol*, [s.l.], 7 set. 2021. Disponível em: <https://www.uol.com.br/tilt/colunas/carlos-affonso-de-souza/2021/09/07/medida-provisoria-de-bolsonaro-sobre-redes-sociais-e-inconstitucional.htm>. Acesso em: 28 mar. 2022.

Ministério da Desinformação (ou o Ministério da Mentira), com amplos poderes para punir quem cerceia o discurso mentiroso e para fazer com que a mentira, o ataque e a desinformação possam ser vistos por todos.

O objetivo do Projeto de Lei nº 3227/2021, portanto, é focar quais conteúdos podem ou não justificar a moderação promovida pelas plataformas, à luz da visão do governo de ocasião, o que eleva de sobremaneira o risco de uso das suas normas para fins de legitimação da censura reversa.

A elaboração do referido projeto de lei, em um momento de quase ruptura institucional, evidencia a limitação da perspectiva libertária da liberdade de expressão, especialmente na era das redes sociais. Mais informação nem sempre significa maior liberdade de expressão. Ao contrário, a manipulação da arquitetura algorítmica das redes sociais, somada a um processo pouco transparente de moderação cria a tempestade perfeita para que teorias conspiratórias, mentiras e discursos de ódio sejam promovidos, de forma viral, por governos que fazem do diversionismo e da intolerância seu principal projeto de poder.

2.6 Redes sociais e esfera pública: os novos desafios

2.6.1 As externalidades algorítmicas

O conceito de “externalidades algorítmicas” (*algorithmic nuisance*) é cunhado por Balkin para se referir ao efeito colateral decorrente do uso massivo de dados e algoritmos para realização de julgamento sobre a identidade e personalidade das pessoas, o que pode vir a afetar suas oportunidades e vulnerabilidades²¹⁰. Diz-se que são “externalidades”, pois tais julgamentos não afetam somente os usuários do serviço, mas também terceiros que não possuem relação direta com as empresas de tecnologia, que usam estes instrumentos para a tomada de decisões. Os algoritmos podem produzir muitas externalidades negativas, mas as principais decorrem do seu uso para fins discriminatórios ou manipulatórios.²¹¹

²¹⁰ BALKIN, Jack. Free Speech in the Algorithmic Society: Big Data, Private Governance and New School Speech Regulation. *UC Davis Law Review*, Davis, v. 51, n. 3, p. 1149-1210, fev. 2018. Mimeografado, p. 1163.

²¹¹ *Ibid.*, p. 1165.

O fenômeno da discriminação algorítmica decorre uma coleta de dados imperfeita ou de preconceitos — conscientes ou inconscientes — dos próprios programadores que espelham, na formulação dos algoritmos, padrões discriminatórios já existentes na sociedade, que não raro podem ser utilizados como instrumentos fomentadores de condições sociais e materiais desiguais.²¹² A superação da externalidade requer regulações que reduzam os vieses algorítmicos²¹³, permitindo não só respostas que melhor retratem a realidade, mas também que possibilitem sua transformação em direção a um padrão social constitucionalmente adequado. Este fenômeno, embora de extrema importância, relaciona-se mais a questões afeitas à isonomia e foge ao objetivo deste trabalho aprofundar o tema.

A manipulação algorítmica, por sua vez, ocorre por meio da exploração, de forma insidiosa, de vieses e vulnerabilidades dos usuários e é matéria diretamente relacionada ao debate sobre a liberdade de expressão no âmbito das redes sociais, já que é elemento que compromete o ideal de um mercado livre de ideias. Aqui, importa analisar como a arquitetura algorítmica da rede contribui para gerar prejuízos à sociedade e à democracia.

O termo “algoritmo” é um tanto quanto abstrato, e embora seja comumente referenciado no campo do direito, quando do estudo de novas tecnologias, são poucos os que o conceituam. Pela clareza e didatismo da exposição, utiliza-se aqui as palavras de Harari, que define os algoritmos como:

um conjunto metódico de passos que pode ser usado na realização de cálculos, na resolução de problemas e na tomada de decisões. Não se trata de um cálculo específico, mas do método empregado quando se fazem cálculos. Por exemplo, quando se quer calcular a média entre dois números, pode-se usar um algoritmo simples. Primeiro passo: obtenha a soma dos dois números. Segundo passo: divida a soma por 2.²¹⁴

O algoritmo, portanto, é uma sequência de instruções minuciosamente codificada pelos criadores de um programa ou aplicativo para cumprir um determinado objetivo. Quanto mais dados houver, maiores serão os elementos aptos a orientar o sequenciamento destas instruções.

No estágio tecnológico atual, houve um salto de qualidade no funcionamento destes algoritmos a partir da concepção de uma inteligência artificial, isto é, algoritmos que não apenas se limitam a reproduzir instruções, mas se aperfeiçoam ao longo do tempo, com a experiência,

²¹² Cf. BAROCAS, Solon; SELBST, Andrew D. Big Data's Disparate Impact. *Calif. L. Rev.*, [s.l.], v. 104, n. 671, 2016.

²¹³ Neste sentido, cf. SUNSTEIN, Cass. Governing By Algorithm? No Noise and (Potentially) Less Bias. Preliminary Draft. *Duke Law Journal*, [s.l.], set. 2021.

²¹⁴ HARARI, Yuval Noah. *Homo Deus: uma breve história do amanhã*. Trad. Paulo Geiger. São Paulo: Companhia das Letras, 2016, p. 91.

produzindo resultados que não eram originariamente esperados²¹⁵. Esta técnica, conhecida como aprendizado da máquina, ou *machine learning*, tem como objetivo precípuo a construção de resolução para problemas pelo computador ou aplicativo, por conta própria, a partir da descrição apresentada por seus programadores.²¹⁶

Sem os algoritmos não há como os intermediários da camada de conteúdo da *internet* funcionarem, pois são eles que permitem a execução de uma série de aplicativos que se tornaram essenciais na sociedade contemporânea, como, por exemplo, serviços de transporte individual e coletivo, de entrega de alimentos e mercadorias, de *streaming* de vídeo e, em especial, redes sociais.

Se os algoritmos podem ser sucintamente descritos como uma sequência de códigos orientados a um fim idealizado por um grupo de programadores, a melhor forma de compreendê-lo é partir das premissas de quais seriam as finalidades declaradas de uma rede social.

Usualmente as redes sociais se apresentam como desenvolvedora de tecnologias e serviços para que “as pessoas conectem umas às outras”,²¹⁷ disponibilizando “um fórum para que as pessoas se informem e inspirem-se em outros usuários”²¹⁸. Assim, a missão principal dos seus algoritmos seria, a princípio, fomentar a comunicação e troca de conteúdos e ideias entre seus múltiplos usuários. Nada obstante, há um aspecto nesta atividade de comunicação que, por vezes, passa despercebido e que é essencial para que a maior parte destas empresas continuem a operar de forma lucrativa.

Como seus serviços são gratuitos aos usuários, a lucratividade destas empresas usualmente depende da venda do espaço destas redes para divulgação de anúncios e serviços de outras empresas²¹⁹. Estes anúncios, diferentemente do que ocorre na mídia televisiva ou imprensa, são direcionados por meio de uma estratégia de negócios inovadora, denominada de microdirecionamento (*microtargeting*).

²¹⁵ O termo “inteligência artificial” é definido pela Comissão Europeia como um conceito que se aplica aos “sistemas que apresentam um comportamento inteligente, analisando o seu ambiente e tomando medidas — com um determinado nível de autonomia — para atingir objetivos específicos” (UNIÃO EUROPEIA. *Comunicação da Comissão ao Parlamento Europeu, ao Conselho, ao Comitê Econômico e Social Europeu e a ao Comitê das Regiões. Inteligência Artificial para a Europa*. [S.l.]: UE, 2018, p.1. Disponível em: <http://data.consilium.europa.eu/doc/documents/ST-8507-2018-INIT/pt/pdf>. Acesso em: 11 abr. 2022).

²¹⁶ MEDON, Filipe. *Inteligência Artificial e Responsabilidade Civil*. Salvador: Juspodium, 2020, p. 92.

²¹⁷ A referência consta do primeiro parágrafo dos termos de serviço da empresa Meta, responsável pelas redes sociais Facebook e Instagram. Informação disponível em: <https://pt-br.facebook.com/terms/>. Acesso em 11 abr. 2022.

²¹⁸ É o que consta dos termos de serviço do Youtube. Informação disponível em: <https://www.youtube.com/static?gl=BR&template=terms&hl=pt>. Acesso em: 11 abr. 2022.

²¹⁹ Conforme descrito nos termos de serviço da Meta. Informação disponível em: <https://pt-br.facebook.com/terms/>. Acesso em: 11 abr. 2022.

A prática consistem no encaminhamento de mensagens publicitárias e informações moldadas ao perfil dos usuários. O sucesso do empreendimento de direcionamento publicitário requer uma captura maciça de dados²²⁰ para fins de construção de perfis e de arquiteturas que mantenham o usuário engajado às redes sociais, capturando sua atenção e influenciando comportamentos.²²¹

O microdirecionamento não é algo ruim. Ele produz efeitos econômicos positivos, pois possibilita às empresas que contratam os serviços fornecidos por estas plataformas um acesso específico à sua audiência com um custo mais baixo do que teria para divulgar publicidade nos canais tradicionais da imprensa. Ademais, é uma necessidade em um contexto de aumento exponencial dos elementos informativos à nossa disposição. Segundo o economista Herbert Simon, “em um mundo rico em informação, sua riqueza significa a escassez de algo mais: a escassez de tudo que esta informação consome. O que a informação consome é bastante óbvio: a atenção de seus destinatários”.²²² Neste modelo, a atenção é o produto a ser vendido, e quem melhor explorar vieses dos usuários/consumidores terá maior sucesso no empreendimento. Sob este enfoque, portanto, o microdirecionamento é uma estratégia promotora do desenvolvimento econômico na era da sociedade da informação.

É, entretanto, uma ferramenta que também produz efeitos deletérios consideráveis, traduzindo dilemas éticos. O principal dilema decorre do fato de que a exploração destes vieses para atender fins lucrativos das redes sociais e de terceiros anunciantes se dá de forma pouco transparente, sem que o usuário tenha perfeita ciência dos riscos a que está exposto, o que não

²²⁰ A captura maciça de dados é fundamental para a construção de um perfil do usuário (*profiling*), que possibilita o uso de análises psicométricas capazes de melhor identificar os gostos dos usuários. A técnica, usualmente referenciada como um “paradigma preditivo de avaliação das emoções” foi, durante muito tempo, a principal base para análises de comportamento realizadas pelas principais plataformas. Detalhes sobre o funcionamento da técnica podem ser conferidos em KOSINSKI, Michal, STIWELL, David; GRAEPEL, Thore. Private traits and attributes are predictable from digital records of human behavior. *PNAS*, [s.l.], v. 110, n. 15, mai. 2013.

²²¹ O psicólogo e autodenominado *designer* comportamental (*behavior designer*) B. J. Fogg afirma que, mais importante que a capacidade preditiva, é a construção de algoritmos capazes de criar hábitos e influenciar atitudes e comportamentos de forma persuasiva (é o que ele denomina matriz captológica algorítmica ou *computer as persuasive technology — captology*). Os algoritmos seriam assim verdadeiras armadilhas cognitivas orientadas à captura de atenção. A transição de uma matriz exclusivamente preditiva para uma captológica se verifica pela mudança de métricas utilizadas pelas principais redes sociais para avaliar a eficiência de seus algoritmos de recomendação. Mais do que curtidas e *likes*, plataformas como o Youtube, por exemplo, têm usado o tempo de permanência da experiência oferecida como métrica principal, conferindo menor peso ao sentimento que a experiência gera ao usuário. Cf. FOGG, B. J. *Persuasive technology: using computers to change what we think and do*. Burlington: Morgan Kaufmann, 2003. Também neste sentido, EYAL, Nir. *Hooked: como construir produtos e serviços formadores de hábitos*. Trad. Edson Furmankiewicz. Cascavel: Alfacon, 2020.

²²² Trecho de palestra de Herbert Simon (Designing Organizations for in Information Rich-World. In: GREENBERGER, M. (Ed.). *Computers, Communication, and the public interest*. Baltimore, MD: The Johns Hopkins Press, 1971. Disponível em: <https://digitalcollections.library.cmu.edu>. Acesso em: 13 abr. 2022).

só agrava a ansiedade e produz novos padrões de vício em seus usuários²²³, como também contribui para a amplificação do alcance de conteúdos que reforçam, não raro, mensagens negativas ou sensacionalistas por serem aquelas capazes de atrair mais atenção de um número maior de pessoas. O foco na capacidade de engajamento elimina métricas de qualidade, tais como aquelas relacionadas à confiabilidade da fonte²²⁴. Importa mais aquilo que é capaz de viralizar e produzir polêmicas, mantendo usuários ativos na rede. Estas opções produzem maior polarização, acirram conflitos sociais e contribuem para a formação do que se convencionou denominar de “filtros bolhas” (*bubble filters*) ou “câmaras de eco” (*echo chambers*).

A expressão “filtro bolha” foi usada, pela primeira vez, por Eli Parisier para fazer referência a algoritmos que recomendam conteúdos próprios às preferências dos usuários, de modo a mantê-los com a atenção capturada²²⁵.

O algoritmo das redes sociais funciona para apresentar ao usuário conteúdos relacionados a escolhas que foram feitas anteriormente, partindo da premissa de que o contato com o que lhe despertou mais atenção irá mantê-lo “conectado” ou “engajado” àquela rede social. Grosso modo, o algoritmo se baseia no interesse do usuário por determinado assunto e na proximidade ou alcance da pessoa que compartilha o artigo, conteúdo ou vídeo para definir o grau de visibilidade/invisibilidade²²⁶ que aquele material terá no seu *feed* de notícias.

“Câmara de eco”, por sua vez, é um termo mais antigo que remete à época em que a camada de conteúdo da rede era dominada por *blogs*, e consiste em um conjunto de pessoas que pensam da mesma forma e compartilham da mesma opinião, havendo pouco espaço para o dissenso.²²⁷ A diferença principal entre filtros-bolhas e câmaras de eco é que a última pode surgir de forma orgânica.

²²³ A respeito, confira-se artigo que destrincha as inúmeras estratégias destas redes sociais para explorar a ansiedade e vícios de seus usuários (BHARGAVA, Vikram; VELASQUEZ, Manuel. Ethics of Attention Economy: The Problem of Social Media Addiction. *Business Ethics Quarterly*, [s.l.], v. 31, n. 3, p. 321-359, jul. 2021).

²²⁴ O psicólogo Jonathan Haidt sugere que a curadoria de conteúdos pelas redes sociais deveria usar outras métricas para além da popularidade, adicionando pontos de fricção sempre que uma *postagem* for tóxica e, ainda, criando grupos de auditoria para avaliar danos e vieses produzidos pelos algoritmos de recomendação. Informação disponível em: HAIDT, Jonathan; ROSE-STOCKWELL, Tobias. The dark psychology of social networks. *The Atlantic*, [s.l.], dez. 2019. Disponível em: <https://www.theatlantic.com/magazine/archive/2019/12/social-media-democracy/600763/>. Acesso em: 11 abr. 2022.

²²⁵ Cf. PARISIER, Eli. *The Filter Bubble. What the internet is Hiding from you*. Nova Iorque: The Penguin Press, 2011.

²²⁶ É neste sentido que Nitrini afirma que a filtragem algorítmica é “central para a governança de discursos mais ampla que essas redes sociais exercem, possibilitando uma nova forma de edição de conteúdos baseada no binômio visibilidade/invisibilidade e customizando essa edição, no limite, a um nível individual”. Cf. NITRINI, Rodrigo Vidal. *Liberdade de Expressão nas Redes Sociais*. São Paulo: Dialética, 2021, p. 75.

²²⁷ Cf. SUNSTEIN, Cass. *Republic.com 2.0*. Princeton, Oxford: Princeton University Press, 2009, p. 145.

No âmbito das redes sociais, não raro os fenômenos se misturam, razão pela qual a abordagem entre os dois termos ocorre de forma conjunta. No caso de pessoas com preferências preestabelecidas, é usual que já integrem alguma câmara de eco (*online* ou *offline*), mas o posicionamento tende a se radicalizar quando incentivado pelos filtros algorítmicos das redes. Por outro lado, mesmo pessoas com tendências “imparciais” — ainda que de forma mais sutil — são suscetíveis ao efeito filtro bolha dos algoritmos. Não raro, à medida que os algoritmos de recomendações se aperfeiçoam e fornecem respostas mais precisas ao interesse do usuário, é comum que estas pessoas se aproximem de outras, também sujeitas ao efeito das ações dos algoritmos de recomendação, que compartilhem de visões semelhantes, formando as câmaras de eco.

A abordagem recente das câmaras e filtros das redes geralmente confere ênfase ao seu impacto recente em eleições ao redor do mundo, com maior destaque para a eleição norte-americana em 2016.²²⁸ Embora o tema seja pertinente, é discutível se o uso do algoritmo de recomendação do Facebook e demais redes sociais foram suficientes para alterar o resultado destas eleições, não havendo estudos conclusivos neste sentido²²⁹. Este não é o ponto que se pretende enfrentar aqui. O enfoque será dado ao impacto que filtros bolhas e câmaras de eco podem ter sob o tecido social à medida em que acentuam o fenômeno da “polarização de grupo” e criam um cenário propício para a produção de “desordens informacionais”.

A polarização de grupo refere-se, segundo Sunstein, à maior probabilidade de as pessoas radicalizarem seu posicionamento originário após deliberarem com membros do grupo que compartilham da mesma orientação.²³⁰ A existência do fenômeno é fundada em inúmeras pesquisas empíricas. A título de exemplo, um experimento em Colorado, em 2005, trouxe três questionamentos polêmicos sobre casamento entre pessoas do mesmo sexo, aquecimento global e ações afirmativas nas relações de trabalho. Após a resposta individual anônima de cada um dos participantes da pesquisa, estes foram divididos em dois grupos (liberais e conservadores) para deliberar por cerca de quinze minutos sobre o tema. O que se verificou é que, após o debate, as opiniões tanto de liberais quanto de conservadores se tornaram mais intensas e radicais que as opiniões afirmadas individualmente. Do mesmo modo, as poucas divergências entre os

²²⁸ ALCOTT, H.; GENTZKOW, M. Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, [s.l.], v. 31, n. 2, p. 211-236, 2017.

²²⁹ Essa tese ganhou força após o escândalo Cambridge Analytica. Sem prejuízo do reconhecimento de que o uso massivo de dados sem consentimento dos usuários provocou danos à intimidade, o argumento de que os algoritmos das redes sociais teriam impactado o resultado da eleição é tido por incerto. Neste sentido, SUMPTER, David. *Dominado pelos números. Do Facebook e Google às fake news. Os algoritmos que controlam nossas vidas*. São Paulo: Bertrand, 2019; NELSON, J. L. Is ‘fake news’ a fake problem? *Columbia Journalism Review*, [s.l.], v. 31, 2017.

²³⁰ SUNSTEIN, Cass. Op. cit., p. 60

grupos eram rapidamente solucionadas.²³¹ O estudo foi ainda reproduzido em outros países apresentando resultados similares.²³² O que o resultado empírico evidencia, além da inclinação à polarização pelo grupo, é que a ausência de diversidade dentro do conjunto considerado obsta, inclusive, a aceitação de argumentos e pensamentos diversos.²³³

Dentre as causas para a polarização, destaca-se a assimetria das informações apresentadas (há uma tendência muito maior de receber elementos informativos que confirmem vieses cognitivos ou discriminatórios); o sentimento de pertencimento a um grupo, que leva à adoção de posicionamentos mais contundentes com o propósito de obter maior conhecimento interno e, ainda, a confiança criada pela corroboração do posicionamento por terceiros.

Assim, quanto mais “likes” e “curtidas”, maior a credibilidade que a pessoa alimenta de si própria e mais enviesado tende a ser seu posicionamento, tornando-a mais suscetível inclusive à crença em informações falsas²³⁴ e transformando as redes sociais em um espaço de produção do que vulgarmente se rotula como *fake news*, mas que, em um esforço de se conferir maior precisão ao termo, tem sido conceituado no âmbito acadêmico como “desordens informacionais”, um gênero que abarcaria ao menos três espécies de conteúdos problemáticos que tendem a se reproduzir em uma esfera fragmentada e altamente polarizada como a das redes sociais²³⁵, podendo produzir danos sociais. São elas a *disinformation*/desinformação (conteúdo falso produzido com aparência jornalística, usualmente para fins políticos ou econômicos, para causar dano), *misinformation* (informação falsa difundida sem o intuito de prejudicar) e a *mal-*

²³¹ SUNSTEIN, Cass. *Republic.com 2.0*. Princeton, Oxford: Princeton University Press, 2009, p. 61.

²³² Resultado similar foi obtido em pesquisas empíricas realizadas no Canadá, Índia, Bangladesh, Nova Zelândia, Alemanha e França. Cf. *Ibid.*, p. 62.

²³³ Alguns estudos empíricos mais recentes têm relativizado o peso das “câmaras de eco” e “filtros bolha” na produção de um grau extremado de polarização. Embora haja evidências sólidas de que o isolamento informacional produz polarização, há evidências de que o acesso a notícia de matizes ideológicos distintos não tem, propriamente, reduzido o grau de animosidade social. Ao revés, o contato com informações e pessoas com visões de mundo diferente nas redes sociais, especialmente entre usuários que consomem muito conteúdo político, tem produzido uma espécie de *backlash*, uma repulsa instantânea daquele que vê sua percepção confrontada por uma informação ou manifestação que o desagrade. Tais achados indicam que a submissão a um grau mais amplo de informações e manifestações deva vir, talvez, acompanhado de algum elemento que certifique a credibilidade daquilo que se divulga. Recomenda-se também a supressão de métricas de quantificações, que estimulam a publicações “mais inflamadas”. Neste sentido, cf. BARBERÁ, Pablo. Social Media, Echo Chambers and Political Polarization. In: _____. *Social media and democracy*. [S.l.]: Cambridge University Press, 2020. Disponível online em: <https://www.cambridge.org/core/books/social-media-and-democracy/social-media-echo-chambers-and-political-polarization/333A5B4DE1B67EFF787626118CCFE19> Acesso em: 06 mai. 2022.

²³⁴ David Sumpter destaca que os mais vulneráveis às informações falsas são justamente os mais convictos de sua opinião. *op. cit.* p.62. A assertiva é corroborada por pesquisas recentes do Instituto Reuters, da Universidade de Oxford. Para mais detalhes: MAZZA, Luigi. O rosto da desinformação. *Piauí*, [s.l.], 10 set. 2021. Disponível em: <https://piaui.folha.uol.com.br/o-rosto-da-desinformacao/>. Acesso em 20 abr. 2022.

²³⁵ A referência ao uso do gênero e suas espécies constam de diversos trabalhos acadêmicos sobre o tema. A título de exemplo, menciona-se relatório sobre desordem informacional produzido pelo Conselho da Europa (*Council of Europe*). Cf. WARDLE, Claire; DERAKHSHAN, Hossein. *Information Disorder: Toward an interdisciplinary framework for research and policy making*. [S.l.]: Council of Europe, 2017.

information (informação, que embora não seja propriamente falsa, é divulgada com o intuito de causar prejuízo, tais como vazamentos de fatos privados, discursos de ódio e assédio).²³⁶

As redes sociais são produtoras de bolhas e câmaras de eco e um potencial fomentador do fenômeno da polarização de grupo e de enclaves informacionais. Estas são as principais sequelas do uso de estruturas algorítmicas para exploração de vieses e captura de atenção. Neste contexto comunicacional, a preocupação maior não é gerar um diálogo entre interlocutores de modo que se chegue a uma interpretação plausível da realidade — como preconizado por Mill, Holmes, Rawls ou Habermas —, mas antes reforçar pré-compreensões e provar, a qualquer custo, que o outro está errado. O risco, portanto, é a transformação do ideal do pluralismo em uma rede de usuários incapazes de se respeitarem mutuamente.²³⁷

O acirramento da polarização social do qual as redes sociais dependem para sobreviver economicamente encontra principalmente na política um campo fértil para se desenvolver. A dinâmica da viralização²³⁸, que aposta mais em conteúdos políticos polêmicos capazes de serem impulsionados por algoritmos e câmaras de eco, em detrimento de mensagens de consenso que busquem a agregação dos múltiplos pontos de vistas que compõem uma sociedade plural, potencializam as teorias conspiratórias, o escândalo e o conteúdo destrutivo²³⁹, transformando-as em tendência na comunicação estabelecida entre a esfera pública — cada vez mais digital — e as esferas formais do exercício do poder político.

Os efeitos deste contexto comunicacional sobre a política são descritos por Giuliano da Empoli²⁴⁰ nos seguintes termos:

²³⁶ Os dois últimos termos não possuem uma tradução precisa para o português.

²³⁷ Um sintoma da polarização exacerbada nas redes é o surgimento de novas expressões, que buscam descrever estratégias discursivas de ataque coletivo de usuários a reputação de terceiros cada vez mais comuns na atualidade, tais como *trolling*, *ciberbullying*, *shaming* e a cultura do cancelamento.

²³⁸ As *cibercascatas* (*cibercascades*) — uma espécie de efeito manada — são uma consequência da dinâmica de viralização nas redes. Informações impactantes e negativas se expandem com maior facilidade para além das bolhas epistêmicas atingindo inúmeras pessoas, que tendem a acreditar na notícia, especialmente quando compartilhadas por pessoas próximas (SUNSTEIN, Cass. *Republic.com 2.0*. Princeton, Oxford: Princeton University Press, 2009, p. 88).

²³⁹ Segundo Kahneman, nosso cérebro é moldado para reagir mais rapidamente a ameaças mesmo que puramente simbólicas: “(...) palavras emocionalmente carregadas chamam rapidamente a atenção e palavras ruins (guerra, crime) chamam a atenção mais do que palavras felizes (paz, amor). Não há ameaça real, mas o mero lembrete de um exemplo ruim é tratado pelo Sistema 1 como ameaçador”. E, continua: “A sensibilidade a ameaças se estende ao processamento de manifestações de opiniões das quais discordamos fortemente. Por exemplo, dependendo de sua atitude em relação à eutanásia, levaria menos de um quarto de segundo para seu cérebro registrar a ameaça numa frase que começa com “Acho que a eutanásia é um procedimento aceitável/inaceitável...” (KAHNEMAN, Daniel. *Rápido e Devagar: Duas Formas de Pensar*. São Paulo: Objetiva, 2018, p. 376).

²⁴⁰ EMPOLI, Giuliano da. *Os Engenheiros do Caos: como as Fake News, as Teorias da Conspiração e os Algoritmos estão sendo utilizados para disseminar o ódio, medo e influenciar as eleições*. Belo Horizonte: Vestígio, 2019, p. 91-92.

Se antes, cada líder político tinha que se dirigir ao eleitor médio com mensagens moderadas, em torno das quais devem convergir o maior número de pessoas, o foco agora é inflamar o maior número possível de grupos, para em seguida adicioná-los — mesmo à revelia deles. As contradições do conteúdo, mensagens e informações direcionadas a um e outro usuário permanecerão, de forma geral, invisível aos olhos da mídia e do público em geral. Como consequência, candidatos com tendências extremistas são trazidos para o centro do jogo político.

A retórica política naturalmente recorre a uma linguagem de afeto, que demanda a construção de um ideal de “nós” para se contrapor àquele que se busca derrotar nas eleições (“eles”)²⁴¹, mas é preciso que haja um acordo mínimo sobre as regras necessárias à manutenção da própria democracia, o que tende a se inviabilizar com a contínua escalada de acirramento da polarização.

Decerto que a instabilidade institucional que assola o Ocidente nos últimos anos não decorre, tão somente, das redes sociais. Crises morais e econômicas gravíssimas são elementos nucleares que contribuirão significativamente para a instabilidade institucional²⁴², que atinge parte das democracias ocidentais. É inegável, no entanto, que a forma como as informações circulam nas principais redes sociais globais contribui para o acirramento da polarização social e política, pois os algoritmos que a orientam reforçam vieses e preconceitos dos usuários de forma sub-reptícia, contaminando a formação de suas opiniões.

Em síntese, a sobrevivência econômica das redes sociais depende de filtros de recomendação, publicidades direcionadas e impulsionamento de conteúdos capazes de produzir engajamento, uma forma mais eficiente de exploração do mercado de captura da atenção. Por outro lado, tais expedientes implicam, em alguma medida, a manipulação, pouco transparente, de vieses dos seus usuários e fomento à produção constante de conflitos na sociedade, criando

²⁴¹ Para Mouffe, todo discurso político tem pretensões hegemônicas e se constrói com base no antagonismo (nós x eles). Discursos de “afeto” são fundamentais, portanto, para o exercício de qualquer pretensão política e não podem ser abandonados em prol de um ideal puro de racionalidade jurídica, inexistente na prática, o que não significa dizer que o “antagonismo” não deva ser contido em uma forma de “agonismo” que permita às partes se respeitarem enquanto adversários. Cf. MOUFFE, Chantal. *Por um populismo de esquerda*. São Paulo: Autonomia Literária, 2020.

²⁴² Os principais estudiosos do processo de erosão das democracias ocidentais costumam apontar quatro grandes razões para este fenômeno: (I) crises econômicas devastadoras ocorridas nos últimos anos; (II) inúmeros escândalos morais que afetam a confiança na classe política; (III) a percepção social da ausência de representatividade das instituições políticas; e (IV) o uso de novas esferas comunicacionais — geralmente, redes sociais — como instrumento fomentador de desinformação e conflitos. A título de exemplo, cf. CASTELLS, Manuel. *A ruptura: a crise da democracia liberal*. Rio de Janeiro: Zahar, 2018; MOUNK, Yascha. *O povo contra a democracia: porque nossa liberdade corre perigo e como salvá-la*. Trad. Cássio de Arantes Lete e Débora Landsberg. São Paulo: Companhia das Letras, 2019; ABRANCHES, Sérgio. *O tempo dos Governantes Incidentais*. São Paulo: Companhia das Letras, 2020; SOUZA NETO, Cláudio Pereira. *A Democracia em Crise no Brasil: valores constitucionais, antagonismo político e dinâmica institucional*. São Paulo: Contracorrente, 2020.

um ambiente propício para a produção de desinformação enquanto externalidade negativa²⁴³, o que gera um custo significativo à sociedade e a qualquer projeto de autogoverno coletivo. O desafio de regulação da liberdade de expressão nas redes requer, portanto, que tais empresas façam mais para internalizar estes custos.

2.6.2 Moderação e censura privada

O aprimoramento da atividade de moderação pelas plataformas, com a ampliação de seus termos de uso e normas comunitárias é a resposta padrão que as principais plataformas sociais têm dado para enfrentar as críticas de que tais empresas estariam sendo coniventes com discursos de ódios, ataques e falsidades que circulam por suas páginas, buscando aumentar a confiabilidade dos seus usuários na qualidade de seus serviços.

Para efeitos de comparação, até 2009, a política de moderação do Facebook continha apenas uma página, tendo apenas doze responsáveis pela aplicação destas regras de moderação, sem que houvesse uma metodologia clara para sua aplicação.²⁴⁴ Atualmente, o Facebook não só possui um robusto conjunto de normas comunitárias e termos de uso, tendo criado inclusive *guidelines* para aplicação de normas especiais de moderação durante a pandemia,²⁴⁵ como também contribuiu para a criação de um Conselho de Supervisão supostamente independente para avaliar as decisões de moderação tomadas no âmbito da rede social (o denominado *Oversight Board*). O movimento de incremento das políticas de moderação — ainda que em menor escala — também é observado em outras redes sociais de grande porte.²⁴⁶

Balkin, contudo, alerta para o risco da moderação realizada sem transparência e sem observância do devido processo legal., que poderia redundar no que ele denomina de *digital prior restraint*, algo como “censura prévia digital” pelas plataformas privadas.²⁴⁷ Opta-se neste

²⁴³ Esta parece também ser a opinião de Gustavo Binbenojm em: BINENBOJM, Gustavo. Fake news como externalidades negativas. *Jota*, [s.l.], 23 jun. 2020. Disponível em: <https://www.jota.info/opiniao-e-analise/colunas/publicistas/fake-news-como-externalidades-negativas-23062020>. Acesso: 20 abr. 2022.

²⁴⁴ Informação disponível em: ARHEGAS, João Victor; ESTARQUE, Marina. *Redes Sociais e Moderação de Conteúdo*: criando regras para o debate público a partir da esfera privada. Rio de Janeiro: ITS, 2020.

²⁴⁵ Informação disponível em: <https://pt-br.facebook.com/business/help/1123969894625935?id=434838534925385>. Acesso em: 20 dez. 2022.

²⁴⁶ A título de exemplo: ATUALIZAÇÕES no Centro de Transparência do Twitter. *Blog Twitter*, [s.l.], 14 jul. 2021. Disponível em: https://blog.twitter.com/pt_br/topics/company/2021/atualizacoes-no-centro-de-transparencia-do-twitter. Acesso em: 19 abr. 2022.

²⁴⁷ BALKIN, Jack. Free Speech in the Algorithmic Society: Big Data, Private Governance and New School Speech Regulation. *UC Davis Law Review*, Davis, v. 51, n. 3, p. 1149-1210, fev. 2018. Mimeografado, p. 1177-1179.

trabalho, no entanto, pela referência apenas à expressão “censura privada”. A ênfase no adjetivo “privado” se deve ao fato de que sanções de exclusão impostas pelas principais plataformas (entes privados) produzem efeitos sensíveis próximos ao silenciamento promovido por órgãos de censura estatais. Ademais, as políticas de moderação ocorrem tanto *ex ante*, quanto *ex post*, razão pela qual não parece fazer sentido a menção apenas ao risco de uma censura “prévia”.

Moderação é um termo guarda-chuva usualmente utilizado para se referir à maneira como plataformas privadas de redes sociais formulam e aplicam suas regras sobre o que os usuários de seus serviços podem fazer ou dizer.²⁴⁸ O termo, portanto, engloba tanto o procedimento de elaboração dos termos de uso e normas comunitárias, quanto sua aplicação.

No campo da elaboração das políticas de uso, as plataformas sociais como todo agente privado gozam de ampla autonomia, desde que, por óbvio, estejam em conformidade com as leis dos países onde prestam serviços. O que é ilícito no mundo *offline* também o é no universo *online*. Assim, por exemplo, ainda que seja legítima a criação de plataformas de redes sociais com normas de uso própria, inclusive ideologicamente enviesadas (para uso exclusivo de conservadores ou de liberais, por exemplo), estas não podem permitir comportamentos considerados ilícitos à luz da legislação nacional (no caso do Brasil, por exemplo, não será jamais permitido plataformas coniventes com comportamentos que impliquem crime de racismo em todas as suas acepções).

No campo da aplicação, a execução das normas de moderação de conteúdo pode se dar de diversas formas. Ela será *ex ante* ou *ex post* a depender se o controle é feito antes (entre o *upload* e a publicação) ou posteriormente à publicação do conteúdo. Ela pode se *reativa* quando decorre da provocação de outros usuários incomodados com alguma publicação (*flagging*) ou *proativa*, quando uma equipe da empresa busca ativamente conteúdos violadores das suas normas comunitárias e termos de uso. Tais decisões podem se dar, ainda, de forma manual, quando tomada por humanos, ou automaticamente, quando orientadas por um *software*.²⁴⁹

Na moderação *ex ante*, a atuação da plataforma privada é sempre espontânea, e neste ponto reside o receio da censura prévia digital (*digital prior restraint*) a que alude Balkin. O

²⁴⁸ DOUEK, Evelyn; FRANÇOIS, Camille. The Accidental Origins, Underappreciated Limits and Enduring Promises of Platform Transparency Reporting About Information Operations. *Journal of Online Trust and Safety*, [s.l.], out. 2021, p. 1.

²⁴⁹ Cf. KLONICK, Kate. The New Governors: The People, Rules, and Processes Governing Online Speech. *Harv. L. Rev.*, [s.l.], v. 131, n. 1598, mar. 2017, p. 1636.

principal risco desta espécie de moderação é a ausência de transparência dos mecanismos de inteligência artificial que orientam sua utilização²⁵⁰.

A moderação *ex post* também ocorre de forma majoritariamente proativa (usualmente conteúdos ligados a *spams*, contas falsas, nudez, violação sexual e conteúdos gráficos violentos)²⁵¹, mas os casos mais controversos (discursos de ódio, assédio ou informação falsa) demandam, por vezes, indicação de outros usuários. As principais plataformas sociais oferecem aos usuários a oportunidade de denunciar conteúdo ou perfis, indicando algumas opções padrões como “discurso de ódio”, “spam”, “nudez”, “informação falsa”, “golpe ou fraude” “*bullying*”, “assédio”, dentre outros.

Mesmo que a moderação *ex post* sofra a influência do trabalho manual de um time de moderadores contratados por estas plataformas privadas²⁵², a maior parte desta moderação também requer o uso maciço de mecanismos automatizados, mais ou menos sofisticados²⁵³, para dar conta da enorme quantidade de pedidos de remoção que tomam conta destas plataformas.²⁵⁴

²⁵⁰ Exemplos mais comuns de *software* são: o *PhotoDna* da Microsoft, que cria uma impressão digital identificada para cada imagem *postada* na *internet*, produzindo um número que a identifica. O cruzamento desta imagem com um banco de imagens de abuso sexual infantil permite ao algoritmo identificar a compatibilidade entre os números que formam a impressão digital e as imagens. Trata-se de uma técnica denominada *hashing*. Programas de *geoblocking*, por vezes, são utilizados por plataformas para proibir a circulação de conteúdos considerados ilícitos em uma determinada região. O Youtube também tem o *software* do *IDCloud*, que se vale da técnica de *hashing* para criar uma impressão digital a conteúdos protegidos por direitos autorais. O objetivo é que o *software* impeça a publicação de conteúdos que possam violar direitos autorais de terceiro (Ibid., p. 1637).

²⁵¹ Segundo Kate Klonick “(...) In the third quarter of 2019, nearly 100% of spam and fake accounts, 98.4% of adult nudity and sexual violations, and 98.6% of graphic and violent content on Facebook were found and flagged by the site before users notified the platform” (KLONICK, Kate. The Facebook Oversight Board: Creating an Independent Institution to Adjudicate Online Free Expression. *Yale Law Journal*, [s.l.], v. 129, n. 2418, jul. 2020, p. 2432).

²⁵² Esta atividade de moderação, no Facebook, ocorre em um esquema de pirâmide. Em um primeiro nível, mais amplo normalmente, formado por revisores terceirizados em empresas espalhadas pelo mundo, é feita a moderação do dia a dia, aplicando as regras mais claramente estabelecidas para *flags*, referentes a nudez, pornografia, insultos baseados em religião, etnia ou orientação sexual, incitação à violência contra pessoas ou animais etc. Um segundo nível, formado por funcionários mais experientes e normalmente sediados nos Estados Unidos, é responsável pela equipe mais básica (incluindo decisões que são repassadas do primeiro nível ao segundo), além das decisões sobre temas previamente definidos como prioritários, tais como ameaças iminentes de violência, autolesões, terrorismo e suicídio. No terceiro e último nível atuam os empregados mais graduados do próprio Facebook, sediados na Califórnia. Cf. NITRINI, Rodrigo Vidal. *Liberdade de Expressão nas Redes Sociais*. São Paulo: Dialética, 2021, p. 68.

²⁵³ A maioria dos mecanismos utilizados na moderação de conteúdo, embora automatizados, não são propriamente instrumentos orientados por inteligência artificial. Trata-se antes de meros mecanismos automatizados que comparam novos conteúdos com uma lista negativa de exemplos, não sendo capaz de identificar novos casos de assédio, discursos de ódio ou pornografia. A moderação humana é ainda substancialmente relevante para identificar os casos mais controversos de violação aos termos de uso onde ainda não há uma quantidade razoável de padrões antigos que possam ser evocados como referência. Cf. GILLEPSIE, Tarleton. Content moderation, AI, and the question of scale. *Big Data & Society*, [s.l.], jul./dez. 2020.

²⁵⁴ Só no terceiro trimestre de 2019, 4,28 bilhões de peças de conteúdo foram removidas do Facebook. Desses, apenas 40,9 milhões foram resultados de pedidos de usuários (menos de 1% do total) (KLONICK, Kate. Op. cit., p. 2434).

O uso de mecanismos automatizados como ferramenta preferencial na moderação de conteúdo suscita algumas dificuldades. A licitude de uma expressão ou opinião engloba a compreensão de contextos. Treinar algoritmos a perceberem corretamente contextos políticos, sociais e culturais é algo bastante complicado de se conseguir, mesmo com o uso de técnicas avançadas de *machine learning*. É possível ainda que a base de dados utilizada pelo algoritmo que orienta a moderação esteja enviesado ou sub-representado, tendo poucos elementos capazes de rotular adequadamente certas espécies de discursos culturalmente minoritários, que poderão ser desproporcionalmente afetados pelo impacto do uso de instrumentos automatizados.²⁵⁵ Por sua vez, há uma necessidade de constante adaptação dos algoritmos ante a evolução dinâmica da linguagem, o que demanda a construção de mecanismos de moderação cada vez mais complexos, elevando os custos e energia que estas plataformas precisam despender para a manutenção de um modelo de moderação eficiente. O risco de falibilidade, portanto, é tão presente no uso de técnicas automatizadas, quanto no uso de moderação humana. A preferência por instrumentos automatizados de moderação decorre apenas do fato de se tratar de instrumentos que viabilizam uma tomada de decisão mais rápida em um contexto em que os pedidos de remoção se avolumam aos milhões.

O temor da falibilidade de mecanismo de controle à liberdade de expressão comunica-se ao risco de *chilling effect* usualmente associado a regulações governamentais excessivamente genéricas, incapazes de indicar de forma objetiva qual tipo de discurso deve ser limitado à luz do interesse público. Se, no âmbito público, a resposta a este risco requer a criação de leis e regulamentos mais objetivos, que reduzam tanto quanto possível a margem de discricionariedade daqueles que as interpretarão, é preciso que haja parâmetros que confirmem maior segurança aos usuários sobre a forma como estes mecanismos automatizados serão utilizados como instrumento de controle do discurso no âmbito das redes sociais, na medida em que se reconhece como inevitável o uso destas ferramentas para se promover uma moderação eficiente.

A doutrina aponta dois caminhos que podem conferir maior transparência e legitimidade à atuação das plataformas na moderação. O primeiro é uma maior clareza sobre as taxas de

²⁵⁵ A título de exemplo, redes sociais já excluíram indevidamente — e em mais de uma oportunidade — conteúdos *postados* por grupos indígenas sob o argumento de violação às normas de nudez (FACEBOOK censura fotos de Sebastião Salgado de tribo indígena da Amazônia. *Folha de S. Paulo*, [s.l.], 4 set. 2018. Disponível em: <https://www1.folha.uol.com.br/poder/2018/09/facebook-censura-fotos-de-sebastiao-salgado-de-tribo-indigena-da-amazonia.shtml>. Acesso em 20/04/2022. Acesso em: 20 abr. 2022).

erros e *inputs* associados ao uso de mecanismos automatizados.²⁵⁶ No atual estado da arte, não há mecanismos automatizados infalíveis. Ao revés, o sistema de moderação predominantemente automatizado vem produzindo alguns equívocos emblemáticos.²⁵⁷ Neste cenário, o mais importante seria um sistema transparente de informação quanto às taxas de erro destes mecanismos de moderação, viabilizando o que Douek chama de uma *avaliação pragmática probabilística* do direito à liberdade de expressão.²⁵⁸ Uma vez que os erros são inevitáveis, estes só serão admitidos se houver transparência na informação do maior ou menor risco assumido para que possamos calibrá-las a níveis aceitáveis em cada caso. Trata-se, em síntese, de “uma escolha entre erros”.²⁵⁹

Assim, por exemplo, poder-se-ia admitir ferramentas de moderação automatizadas com limites mais estritos (e *ex ante*) que levassem a maiores riscos de “falsos positivos” quando relacionados a assuntos considerados pela sociedade como mais sensíveis, tais como pornografia infantil, divulgação de crimes violentos ou apologia ao terrorismo. Por sua vez, mecanismos de moderação com limites mais frouxos, com maiores taxas de imprecisão, onde os riscos de “falso negativo” sejam maiores, poderiam ser considerados legítimos em casos cuja identificação seja mais complexa, podendo variar a depender do contexto cultural, como ocorre nos casos de discursos de ódio, *bullying* ou de desinformação.

Isto não obsta o dever de as plataformas aprimorarem o uso destas ferramentas, mas evidencia que o mais importante não é a cobrança imediata por resultados perfeitos, mas uma transparência sobre os *inputs* que informam estes mecanismos de automatização e a mensuração

²⁵⁶ A falta de transparência da lógica algorítmica e a necessidade premente de se jogar luzes sobre os critérios que a orientam é aprofundado na obra seminal de Frank Pasquale (PASQUALE, Frank. *The Black Box Society: The Secret Algorithms that Control Money and Information*. Cambridge, MA: Harvard University Press, 2015).

²⁵⁷ Cite-se dois exemplos recentes. O emblemático caso da supressão da icônica foto da Guerra do Vietnã (A menina do Napalm), identificada como apologia à nudez infantil em: HOLTER, Mikael. Facebook remove foto da 'menina do napalm' e gera protestos. *Tilt Uol*, [s.l.], 9 set. 2016. Disponível em: <https://www.uol.com.br/tilt/noticias/redacao/2016/09/09/facebook-remove-foto-da-menina-do-napalm-e-gera-protestos.htm>. Acesso em: 20 dez. 2022. Mais recentemente, no auge da pandemia de COVID-19, o Facebook derrubou conteúdo *postado* por voluntários que anunciavam iniciativas de doação de máscaras costuradas à mão para profissionais de saúde no EUA, que sofriam com a falta do equipamento. Houve ameaças, inclusive, de proibição de novas *postagens*. O argumento pela proibição era a observância das diretrizes da Organização Mundial de Saúde, que não recomendava o uso de máscaras que não fossem as cirúrgicas. Em contrapartida, golpistas que vendiam máscaras banidas não eram identificados pelo mesmo sistema de moderação automatizada. Informação disponível em: ISAAC, Mike. Facebook hampers do-it-yourself mask efforts. *The New York Times*, [s.l.], 5 abr. 2020. Disponível em: <https://www.nytimes.com/2020/04/05/technology/coronavirus-facebook-masks.html>. Acesso em: 20 abr. 2022; e SILVERMAN, Craig. Facebook Banned Mask Ads. They're Still Running. *BuzzFeedNews*, [s.l.], 13 mai. 2020. Disponível em: <https://www.buzzfeednews.com/article/craigsilverman/facebook-mask-ads-ban-zestads-coronavirus>. Acesso em: 20 abr. 2022.

²⁵⁸ DOUEK, Evelyn. Governing Online Speech: From 'Posts-As-Trumps' to Proportionality and Probability. *Colum. L. Rev.*, [s.l.], v. 121, n. 759, ago. 2020. Disponível em: <https://ssrn.com/abstract=3679607>. Acesso em: 20 dez. 2022, p. 811.

²⁵⁹ *Ibid.*, p. 812.

dos riscos de falha em cada caso, de modo a possibilitar que o Estado e sociedade civil tenham como compreender quais os critérios e riscos associados ao uso destes mecanismos automatizados para exigir os aprimoramentos que se façam necessários, requerer o monitoramento/revisão humana, ou até mesmo para solicitar o abandono efetivo destes mecanismos, na hipótese de as taxas de erros se mostrarem insuportáveis.

O segundo caminho decorre da necessidade de observância de uma garantia análoga ao devido processo legal na aplicação destas normas de moderação²⁶⁰. Por devido processo legal, leia-se o direito dos usuários de saberem por qual razão determinado conteúdo fora rotulado, contingenciado, suprimido ou proibido no âmbito da rede social, conferindo-lhe a oportunidade de contestar tal classificação promovida, seja pela moderação automatizada ou pela moderação humana. Parte substancial da legitimidade de qualquer instituição, pública ou privada, decorre da concretização de um procedimento que confira razoabilidade à tomada de decisão do agente.²⁶¹

O impacto social e político dos discursos e materiais publicizados nas principais plataformas acaba por requerer a observância de algumas garantias procedimentais que assegurem por parte destes agentes privados o respeito aos direitos fundamentais. Uma decisão de moderação que implique a exclusão de uma notícia, vídeo e, pior, perfil de usuário, pode ter um impacto substancial não só para o indivíduo, que muitas vezes depende da rede social para exercício de atividade econômica ou profissional, mas também para a coletividade, quando possa implicar o silenciamento de autoridades públicas relevantes.

Os riscos de censura privada e moderação traduz, portanto, uma questão de legitimidade. As atividades de moderação só serão adequadas se razoáveis, cognoscíveis e precedidas de um procedimento minimamente justificatório. Em que pese as plataformas privadas tenham a autonomia gerencial reconhecida pelo art. 19 do Marco Civil da *Internet*, seu impacto substancial sobre a vida contemporânea exige um comportamento mais transparente e coerente na condução de suas políticas de uso. Somente assim será assegurado o respeito que deve ser dispensado ao direito fundamental à liberdade de expressão.

²⁶⁰ BALKIN, Jack. Free Speech in the Algorithmic Society: Big Data, Private Governance and New School Speech Regulation. *UC Davis Law Review*, Davis, v. 51, n. 3, p. 1149-1210, fev. 2018. Mimeografado, p. 1177. O receio é descrito na seguinte passagem: “(...) *the key problem of administrative prior restraint is that it involves informal or bureaucratic censorship. Decisions about who gets to speak are made by bureaucrats — or in the digital age, by programmers or algorithms — often without notice or an opportunity to be heard and without the civil liberties and other procedural protections that apply in judicial proceedings (...) New school speech regulation is analogous to prior restraint; when speech is filtered, blocked, or taken down by a filter, an algorithm, or a corporate employee, speakers get none of these procedural benefits*”.

²⁶¹ Neste sentido, HAGGART, Blayne; Clara Iglesias. Democratic Legitimacy in global platform governance. *Telecommunications Policy*, [s.l.], v. 45, n. 6, jul. 2021. Disponível em: <https://doi.org/10.1016/j.telpol.2021.102152>. Acesso em: 20 dez. 2022.

2.7 A insuficiência do art. 19 do Marco Civil da *Internet*: a necessária busca por um modelo de regulação das plataformas de redes sociais

A norma da imunidade dos intermediários, consagrada na cláusula 230 do CDA, e importada para o ordenamento jurídico nacional com sutis mitigações, por meio do art. 19 do Marco Civil da *Internet* revela-se, portanto, incapaz de dar conta das mudanças que a inclusão de um novo ator — as redes sociais — trazem para a estrutura da esfera pública.

A principal utilidade da teoria triangular concebida por Balkin é conseguir explicar os impactos à liberdade de expressão produzido por uma esfera pública cada vez mais virtual. Ele esclarece os novos riscos decorrentes desta relação triangular entre Estados, intermediários da *internet* (com ênfase, neste trabalho, nas redes sociais) e sociedade civil.

Se, antes, a ameaça ao livre mercado de ideias vinha apenas da censura oriunda de órgãos burocráticos estatais, uma vez que as formas de comunicação de massa eram concentradas nas mãos de poucos agentes; na esfera pública virtual a ameaça advém de outras frentes distintas.

Há o risco captura do aparato tecnológico privado — essencial ao contínuo fluxo de informações — pelo Estado, seja por meio de regulações excessivamente abrangentes, seja pelo uso abusivo de sua arquitetura algorítmica para ataque coordenado a opositores. Por outro lado, a manipulação algorítmica para fins lucrativos ou políticos tem promovido distúrbios informacionais de toda ordem, problema que se acentua pelo pouco conhecimento sobre os critérios que norteiam o funcionamento da tecnologia utilizada pelas plataformas, o que dificulta a busca por soluções regulatórias adequadas.

Há um evidente *techlash* em curso. Em um primeiro momento, a *ciberesfera* é percebida por sua relevância para o desenvolvimento econômico e a construção de uma cultura efetivamente democrática²⁶², o que impulsionou a defesa de modelos autorregulatórios na esfera digital. Atualmente, há um profundo ceticismo sobre a capacidade destes agentes privados exercerem — sem quaisquer tipos de limites externos — a governança do discurso nas plataformas de redes sociais.

²⁶² Nesta época, ganhou força a defesa de um direito fundamental de acesso à *internet* como meio essencial à viabilização de um projeto de concretização de um direito amplo à informação. Cf. CARVALHO, Ana Cristina Azevedo P. *Marco Civil da Internet no Brasil: análise da Lei nº 12.965/14 e do Direito de Informação*. Editora Rio de Janeiro: Alta Books, 2014, p. 79-88. Tem um livro de uma orientanda minha sobre esse assunto, a Bárbara Coutinho (título é direito fundamental de acesso à internet), depois você dá uma olhada.

Antes de finalizar o capítulo, passa-se brevemente sobre os episódios-paradigma que levaram à essa mudança de visão sobre o papel das redes sociais como promotoras de espaços democráticos. O ponto é relevante para materializar os danos oriundos destas redes. São eles: o escândalo *Cambridge Analytica* e o papel central do Facebook no massacre da minoria étnica Rohingya, no Myanmar.

O caso *Cambridge Analytica* revelou como o Facebook permitiu que um programa criado por uma empresa inglesa (a *Cambridge Analytica*) coletasse informações de milhões de usuários sem autorização. Alguns usuários do Facebook se submeteram a um teste rápido de personalidade disponível na plataforma social, consentindo com o uso dos dados obtidos para fins acadêmicos²⁶³. O Facebook, contudo, permitiu não só a obtenção das informações pessoais daqueles que participaram da pesquisa como também de todas as pessoas que fossem “amigos” dos usuários que se voluntariaram para o teste *online*. A *Cambridge Analytica* utilizou, então, essa enorme gama de dados para fins políticos, coordenando campanhas políticas²⁶⁴ por meio de estratégias de *microtargeting* e desinformação, que tinham como objetivo principal explorar, de forma insidiosa, os vieses dos usuários cujos dados foram ilegalmente coletados.²⁶⁵

O massacre da minoria Rohingya, no Myanmar, foi ainda mais dramático. A perseguição à minoria étnica islâmica do Myanmar é um fenômeno histórico, que se intensifica desde 2012 naquele país, por força de uma política constante de perseguição e de negativas de direito àquela etnia, minoria em um país majoritariamente budista. No dia 25 de agosto de 2017, o confronto alcançou seu ápice quando uma operação militar liderada pelo Exército de Myanmar — em represália a um ataque da minoria Rohingya a alguns postos fronteiriços oficiais do Exército — destruiu inúmeras cidades e vilas onde o grupo étnico tinha maior presença, causando dezenas de milhares de mortes e, ainda, provocando a fuga de mais de 700.000 (setecentos mil) membros do grupo étnico para Bangladesh.

Relatório publicado pela ONU em 2018 destacou o papel central do Facebook para tragédia, visto que a rede social fora utilizada para amplificar o alcance dos discursos de ódio

²⁶³ The Great British Brexit Robbery: how our democracy was hijacked. Disponível em CADWALLADR, Carole. The great British Brexit robbery: how our democracy was hijacked. *The Guardian*, [s.l.], 7 mai. 2017. Disponível em: <https://www.theguardian.com/technology/2017/may/07/the-great-british-brex-it-robbery-hijacked-democracy>. Acesso em: 21 abr. 2022.

²⁶⁴ Destaca-se, em especial, a campanha presidencial norte-americana em 2016 e o plebiscito que definiu a saída do Reino Unido da União Europeia.

²⁶⁵ Para mais detalhes, cf. KAISER, Brittany. *Manipulados: Como a Cambridge Analytica e o Facebook invadiram a privacidade de milhões e botaram a democracia em xeque*. Trad. Roberta Clapp e Bruno Fiuza. Rio de Janeiro: Harper Collins, 2020.

em face da minoria islâmica, e foi pouco responsivo na moderação a centenas de posts que desumanizavam e estigmatizavam os Rohingya e muçulmanos em geral.²⁶⁶

Estes dois eventos, possivelmente os mais impactantes para a reputação das mídias sociais, somam-se a tantos outros²⁶⁷ para marcar o início de um movimento que se traduz por uma profunda desconfiança da sociedade civil não só com a forma como as redes sociais angariam a maior parte das suas receitas (e.g., a partir de coleta de dados), mas também com os critérios pouco transparentes que definem quais discursos terão o alcance amplificado em suas plataformas.

A despeito dos eventos narrados permitirem múltiplas leituras sobre os fatores que contribuíram para esta percepção de desconfiança, é possível extrair três causas principais para tal mudança: a queda do mito da isenção das plataformas; o impacto concreto dos debates e discursos *online* sobre o mundo *offline*, e a ausência de transparência na aplicação das regras de moderação.

O mito da isenção das plataformas ancora-se em um discurso, predominante ao menos até 2016, de que as mídias sociais eram apenas um canal intermediário entre seus usuários e não um *publisher* e, portanto, não lhes caberia propriamente orientar ou promover a governança dos discursos em rede.²⁶⁸ Esta tese caiu por terra. Como bem explica Evelyn Douek, o discurso nas plataformas é fruto de uma complexa interação entre interesses dos usuários, recursos da plataforma e escolhas algorítmicas²⁶⁹, cujo objetivo final é aumentar o tempo de permanência do usuário nas redes sociais, capturando sua atenção.

As plataformas de redes sociais não são, portanto, isentas. Há uma curadoria de conteúdo produzida pelos próprios algoritmos das redes que, em interação com o comportamento de seus usuários, define quais tipos de postagens serão amplificadas e quais

²⁶⁶ HUMAN RIGHTS COUNCIL. *Report of the independent international fact-finding mission on Myanmar*. [S.l.]: HRC, 2018. Disponível em: https://www.ohchr.org/Documents/HRBodies/HRCouncil/FFM-Myanmar/A_HRC_39_64.pdf. Acesso em: 20 dez. 2022.

²⁶⁷ A título de exemplo, podem ser citados: (i) acusações de interferência russa nas eleições norte-americanas e francesas por meio de produção de desinformação em plataformas de redes sociais; (ii) o uso do whatsapp para divulgação de desinformação em massa durante as eleições da Nigéria em 2015; (iii) a anulação da eleição presidencial no Quênia, em 2017, pela Corte Constitucional daquele país, sob o argumento de que a empresa Cambridge Analytica, manipulou os resultados eleitorais por meio de produção de desinformação em grande escala e promoções de ações de segregação étnica; (iv) na Índia, notícias falsas e discursos de ódio tem produzido uma onda de linchamentos e mortes. No caso brasileiro, o uso de mecanismos de disparos em massa — especialmente no whatsapp — para produção de desinformação nas eleições de 2018 foi a tônica tal como ocorreu na Nigéria. Sobre este último caso: cf. MELLO, Patrícia Campos. *A Máquina do Ódio: notas de uma repórter sobre Fake News e Violência Digital*. São Paulo: Companhia das Letras, 2020.

²⁶⁸ KLONICK, Kate. The New Governors: The People, Rules, and Processes Governing Online Speech. *Harv. L. Rev.*, [s.l.], v. 131, n. 1598, mar. 2017.

²⁶⁹ DOUEK, Evelyn. Governing Online Speech: From 'Posts-As-Trumps' to Proportionality and Probability. *Colum. L. Rev.*, [s.l.], v. 121, n. 759, ago. 2020. Disponível em: <https://ssrn.com/abstract=3679607>. Acesso em: 20 dez. 2022.

serão restringidas, tendo como medida a capacidade de engajamento de um determinado conteúdo. Como já se expôs, tal critério privilegia conteúdos sensacionalistas e falsos, que tendem a circularem de forma mais rápida que informações verdadeiras. Partindo desta premissa, seria no mínimo irrazoável não imputar às redes sociais o dever de mitigar as externalidades negativas provocadas por um modelo de exploração econômica que depende desta lógica para sobreviver.

A segunda causa é decorrência dos efeitos concretos produzidos pela polarização dos discursos nas redes. O acirramento da polarização e o aumento da desinformação nas redes sociais já produziu alguns episódios no mundo *offline* com consequências desastrosas. Além dos exemplos de tentativa de manipulação da opinião pública pela empresa *Cambridge Analytica* e o massacre étnico em Myanmar, é possível indicar exemplos mais recentes, tais como os prejuízos que notícias falsas sobre vacinas, medicamentos e tratamentos precoces representaram para a condução das políticas de saúde pública de combate à COVID-19²⁷⁰, sem esquecer a radicalização dos discursos *online* durante as últimas eleições norte-americanas, que foram o combustível para a invasão ao Capitólio no início de 2021²⁷¹.

A terceira causa do *techlash* é a falta de transparência das redes sociais na elaboração e, especialmente, aplicação das suas políticas de moderação de conteúdo. O relevante papel que as plataformas passam a exercer a partir de uma demanda mais robusta de moderação requer maior clareza e coerência sobre critérios de julgamento das plataformas.²⁷²

De fato, uma política autorregulatória de moderação sempre conviverá com a sombra do conflito de interesses, o que justifica a desconfiança da sociedade civil. Como conferir credibilidade à moderação de um agente privado se o lucro destas empresas depende de perfis

²⁷⁰ Neste sentido, matéria da Piauí divulgou pesquisa da IPEC (Inteligência em Pesquisa e Consultoria) que retrata que ao menos 46% dos brasileiros concordam com uma das *fake news* sobre efeitos adversos da vacina contra a COVID-19. Informação disponível em: GORZIZA, Amanda. Uma Mentira Sobre Vacina Enganou 46% dos Brasileiros. *Piauí*, [s.l.], 22 mar. 2021. Disponível em: <https://piaui.folha.uol.com.br/quase-metade-dos-brasileiros-acredita-em-boatos-sobre-vacinas/>. Acesso em: 20 abr. 2022.

²⁷¹ O acirramento da polarização foi capitaneado por *posts* e comentários do então Presidente Donald Trump, o que resultou na sua exclusão de inúmeras redes sociais (CRUZ, Bruna Souza. Veja motivos do Facebook, Twitter e outras plataformas para bloquear Trump. *Tilt Uol*, [s.l.], 7 jan. 2021. Disponível em: <https://www.uol.com.br/tilt/noticias/redacao/2021/01/07/trump-bloqueado-saiba-os-motivos-da-acao-do-facebook-instagram-e-twitter.htm>. Acesso em: 14 dez. 2021). As redes sociais foram também de suma importância para a radicalização do discurso político no Brasil, que levaram ao acirramento do discurso de contestação ao resultado das últimas eleições presidenciais, culminando no lamentável episódio de 08 de janeiro de 2023. A relevância das redes sociais para disseminação de discursos golpistas foi reconhecida pela própria “Meta”. Informação disponível em: BBC. META vai bloquear conteúdo pró-invasão em Brasília das Redes Sociais. 09 de janeiro de 2023. Disponível em: <https://g1.globo.com/tecnologia/noticia/2023/01/09/meta-vai-bloquear-conteudo-pro-invasao-em-brasilia-das-redes-sociais.ghtml>. Acesso em 13/01/2023.

²⁷² Cf. DOUEK, Evelyn. Governing Online Speech: From 'Posts-As-Trumps' to Proportionality and Probability. *Colum. L. Rev.*, [s.l.], v. 121, n. 759, ago. 2020. Disponível em: <https://ssrn.com/abstract=3679607>. Acesso em: 20 dez. 2022.

e conteúdos capazes de engajar e capturar maior atenção dos usuários das redes independentemente da credibilidade ou veracidade daquilo que se veicula? Escândalos recentes, trazidos por um conjunto de séries investigativas do *Wall Street Journals (Facebook Files)* reforçam a falta de compromisso das redes sociais com a proteção de direitos fundamentais. Dentre as revelações, destaca-se: a negligência do Facebook na adoção de medidas para reduzir efeitos adversos das redes sociais para autoestima de crianças e adolescentes, ainda que tais efeitos tenham sido reportados em relatórios internos da empresa; a ausência de um padrão isonômico na aplicação das regras de moderação (usuários com maior capacidade de engajamento contavam com a leniência da empresa para publicar desinformação, promover assédio ou incitar violência sem qualquer espécie de sanção) e, ainda, a construção de algoritmos que — em detrimento do objetivo anunciado de promoção da comunicação — acabavam por tornar as pessoas mais raivosas, incentivando o conflito.²⁷³

Tais elementos indicam a necessidade de se estabelecer algumas obrigações complementares às redes sociais, de modo que elas passem a se portar, de fato, como um fórum público de debates conforme reiterados em suas normas comunitárias e termos de uso, agindo com maior transparência e lealdade com os usuários, garantindo a devida observância dos direitos fundamentais no ambiente *online*.²⁷⁴

O modelo preconizado pelo art. 19 do MCI, que aposta exclusivamente na judicialização, como já se antecipou, é insuficiente para tal finalidade, pois é incapaz de orientar de forma ampla a atuação destes agentes privados, dada especialmente as limitações impostas pela seletividade das questões que alcançam o Poder Judiciário.

A insuficiência do modelo atual de responsabilização para lidar com os compromissos exigidos pela proteção dos direitos fundamentais requer a experimentação de novos arranjos regulatórios. Existem uma série de respostas possíveis das mais às menos intrusivas. Desde regulações do tipo “comando e controle”, passando pela correção e, até mesmo, uma autorregulação orientada por compromissos públicos por maior transparência e assertividade no combate aos danos colaterais produzidos pela arquitetura algorítmica das redes sociais.²⁷⁵

²⁷³ Informação disponível em: THE facebook files. *The Wall Street Journal*, [s.l.], [2021]. Disponível em: <https://www.wsj.com/articles/the-facebook-files-11631713039>. Acesso em: 21 abre. 2022.

²⁷⁴ Neste mesmo sentido, BALKIN, Jack. How to regulate social media. *Journal of Free Speech Law*, [s.l.], v. 1, n. 71, nov. 2019. Mimeografado.

²⁷⁵ BELLI, Luca; FRANCISCO, Pedro Augusto; ZINGALES, Nicolo. Lei do Estado ou lei da Plataforma? Cuidado com a privatização da regulação e da polícia. In: BELLI, Luca; CAVALLI, Olga (Orgs.). *Governança e Regulações da Internet na América Latina: análise sobre infraestrutura, privacidade, cibersegurança e evoluções tecnológicas em homenagem aos dez anos da South School on Internet Governance*. Rio de Janeiro: FGV, 2019, p. 447-470.

Fecha-se este capítulo, portanto, com destaque para um ponto de partida fundamental para qualquer proposta que se faça em matéria de regulação da liberdade de expressão em plataformas de redes sociais.

Os debates regulatórios não devem perder de vista os riscos promovidos por novas tentativas de captura da liberdade discursiva pelo Estado (por meio de novas táticas de censura como a reversa ou colateral), mas também devem ter em mente o principal objetivo destas medidas: internalizar os custos sociais decorrentes da polarização exacerbada e desordens informacionais produzidas pela estrutura algorítmica das redes sociais e, ainda, exigir maior transparência quanto aos *inputs*, resultados e procedimentos, que orientam o exercício da atividade de moderação por parte destas plataformas.

3 CONTEÚDO, ARQUITETURAS E (COR)REGULAÇÃO: UMA ANÁLISE DAS ESTRATÉGIAS E ARRANJOS REGULATÓRIOS

3.1 Regulação e combate à desinformação: conteúdo x arquitetura

No capítulo anterior, foram apresentados os novos danos que justificam eventual intervenção estatal sobre o exercício da atividade econômica de empresas que administram e gerenciam plataformas de redes sociais, bem como os principais riscos que decorrem da atividade regulatória estatal com potencial efeito limitador sobre a liberdade de expressão.

Como resposta a estes desafios, Estados têm desenvolvido múltiplas estratégias regulatórias. Tais estratégias podem, ao nosso ver, ser estruturadas em dois grandes grupos: um primeiro, que tem foco no conteúdo, isto é, legislações que visam a proibição de circulação de determinadas espécies de discurso; e uma segunda, voltada à arquitetura dos intermediários, com ênfase em leis que buscam conferir maior transparência às práticas e critérios que as redes sociais operam para realizar tanto a curadoria quanto a moderação de conteúdo em suas plataformas, seja de forma humana ou algorítmica.²⁷⁶

O caminho que se seguiu até aqui aponta uma nítida preferência por estratégias com maior enfoque nas arquiteturas. Conforme já abordado no capítulo 1.4, a opção pela ênfase na “arquitetura de escolhas” é mais inteligente inclusive para reduzir a chance de êxito no caso de eventuais impugnações judiciais, já que conta com precedentes favoráveis do Supremo Tribunal Federal.

Nos dizeres da Suprema Corte, intervenções legislativas que limitem a autonomia das empresas que atuam na área de comunicação para determinar a priorização de certos conteúdos (1070522/PE e ADI 4293/DF), ou, ainda, para criar um dever de emissão de alertas informativos sobre a características dos programas que transmitem (ADI 2404/DF), configuram uma espécie de restrição razoável à liberdade de expressão, não configurando censura.

²⁷⁶ Em sentido similar, confira-se: ARHEGAS, João Victor; BOTTINO, Celina; PERRONE, Christian. Moderação de Conteúdo em 2021: quem regula o moderador? *Jota*, [s.l.], 17 jan. 2022. Disponível em: <https://www.jota.info/opiniao-e-analise/artigos/moderacao-de-conteudo-2021-quem-regula-moderador-17012022>. Acesso em: 17 mai. 2022. No mesmo caminho, embora acrescente ainda uma terceira espécie de estratégia (a estratégia focada na limitação ao uso de dados), KELLER, Clara Iglesias. Don't Shoot the Message: Regulate Disinformation Beyond Content. *RDP*, Brasília, v. 18, n. 99, p. 486-515, jul./set. 2021, p. 486-514.

De todo modo, não se afasta em absoluto a possibilidade de restrições legislativas diretamente sobre o conteúdo — como, aliás, já existem no ordenamento jurídico brasileiro²⁷⁷ —, embora seja certo que tais hipóteses devam ser bastante excepcionais. Como já se antecipou, restrições sobre conteúdo são possíveis, desde que amparadas em razões sujeitas a amplo escrutínio público, o que recomenda a construção destas normas por meio de processos legislativos que assegurem ampla participação social, resguardado sempre o papel do Poder Judiciário de atuar como revisor deste procedimento, ante seu caráter contramajoritário. É natural, portanto, que diante de fenômenos parcialmente novos²⁷⁸ como a *desinformação*, *ciberbullying*, *stalking* e similares, o legislador veja-se tentado a responder com a criação de novos tipos penais ou infrações administrativas.

Regulações pelo conteúdo buscam identificar certas espécies de discurso como danosos (em geral fazendo referências a termos como “desinformação”, “falsidades” “terrorismo”, “assédio” e “*bullying*”), e visam bani-los de circulação por meio da ameaça de sanções criminais aos usuários ou imputação de sanções administrativas e até mesmo criminais aos intermediários (e seus representantes), que se omitem no dever de retirada destes conteúdos, sendo coniventes com a sua livre distribuição por suas plataformas virtuais.

A título ilustrativo, elenca-se alguns exemplos. Na Malásia, em abril de 2018, foi promulgada a primeira lei de anti-*fake news* do país. A referida lei definia *fake news* como “*notícias, informações, dados e relatórios que são parciais ou totalmente falsos*”, e punia com até seis anos de prisão e multa de 500 mil *ringgits* quem disseminasse tais informações dentro e fora da Malásia.²⁷⁹ Assim que entrou em vigor, a lei foi taxada como um atentado à liberdade de expressão pelos críticos²⁸⁰ e já em 2019, após a vitória da oposição ao então atual governo, a norma fora revogada. Em 2021, no entanto, a previsão retornou por meio de uma lei

²⁷⁷ Aqui é o caso não só do supracitado exemplo da Lei nº 7716/89, como também de tipos penais como calúnia, difamação e injúria constantes no Código Penal (arts. 138, 139 e 140 do CP). Embora a existência destes dispositivos sejam objeto de críticas por parte da doutrina, por se tratar de tipos penais cuja vigência se iniciou anteriormente à incidência da Constituição de 1988 representando duvidosa limitação à liberdade de se expressar, é certo que o tipo penal da “injúria” teve seu alcance ampliado — pós-1988 — a partir de uma nova qualificadora denominada “injúria racial”. Para maiores detalhes sobre a crítica a estes vetustos tipos penais, confira-se SCHREIBER, Anderson. De São Conrado para o STF: uma nota sobre a liberdade de expressão. *Jota*, [s.l.], 3 ago. 2021. Disponível em: <https://www.jota.info/opiniao-e-analise/colunas/coluna-do-anderson-schreiber/de-sao-conrado-para-o-stf-uma-nota-sobre-a-liberdade-de-expressao-03082021>. Acesso em 23 mai. 2022.

²⁷⁸ Diz-se parcialmente, já que a veiculação da mentira e o assédio moral sempre estiveram presente em nossa sociedade. Muda-se, contudo, a dimensão do a seu alcance. A *internet* e as plataformas de rede sociais reduzem os custos de tempo e materiais de expressões prejudiciais, que podem viralizar em poucos segundos.

²⁷⁹ Informação disponível em: MALÁSIA é o primeiro país do mundo a revogar lei de combate às fake news. *Consultor Jurídico*, [s.l.], 19 ago. 2028. Disponível em: <https://www.conjur.com.br/2018-ago-19/malasia-primeiro-pais-revogar-lei-combate-fake-news>. Acesso em: 20 mai. 2022.

²⁸⁰ Informação disponível em: NGUI, Yantoultra. Malaysia Passes ‘Fake News’ Law That Critics Call an Assault on Speech. *The Wall Street Journal*, [s.l.], 3 abr. 2018. Disponível em: <https://www.wsj.com/articles/malaysia-passes-fake-news-law-that-critics-call-an-assault-on-speech-1522757225>. Acesso em 20 mai. 2022.

emergencial que criminalizou a disseminação de notícias falsas relacionadas à COVID-19, trazendo sanções criminais idênticas à lei revogada em 2019.²⁸¹

Em 2016, a Indonésia emendou a Lei nº 11/2008 do país (Lei nº 19/2016), que regulava a tramitação de informações eletrônicas do país. Dentre as inovações inseridas pela emenda, destaca-se a previsão de sanção de até seis anos de prisão e um valor máximo de multa na cifra de um bilhão de rúpias a qualquer pessoa que distribua intencionalmente ou crie informação eletrônica que seja violenta, insultante, difamatória, ameaçadora, falsa, enganosa ou que infrinja ódio ou dissenso entre indivíduos ou grupos, por causa de raça, religião ou etnia.²⁸² Leis similares foram aprovadas no Camboja, Quênia, Etiópia²⁸³ e Venezuela²⁸⁴.

Trata-se, assim, de leis que buscam responsabilizar o usuário, não importando ao menos em um primeiro momento qualquer tipo de restrição à atuação das empresas, embora não raro tais previsões normativas sejam acompanhadas de dispositivos que confirmam aos intermediários o dever de auxiliar nas investigações criminais.

Outros países, como Singapura, criaram leis (*Protections from online Falsehoods and Manipulation Act – POFMA*) que mesclam a possibilidade de responsabilização individual dos usuários por crimes de divulgação de notícias falsas (prevendo penas de até 05 anos) com a possibilidade de responsabilização da figura dos intermediários que deixem de cumprir as ordens do Poder Executivo de bloqueio de acesso a tais declarações, quando prejudiciais ao interesse público. O valor das multas pode chegar a \$ 500.000 por ofensa.²⁸⁵ A vagueza das expressões utilizadas pela lei — em especial a referência ao “interesse público” como fundamento para imposição de restrições de acesso e aplicação de sanções — tem sido amplamente criticada por permitir a “consolidação do poder político, aumentar a supervisão

²⁸¹ A lei restringe um pouco mais o conceito de *fake news* ao defini-la como “qualquer notícia, informação e relatórios que sejam no todo ou em parte falsos relacionados à COVID-19 ou proclamação do estado de emergência, seja na forma de recursos visuais ou de gravações de áudio ou em qualquer outra forma capaz de sugerir palavras ou ideias”.

²⁸² A íntegra da lei pode ser conferida em: https://www.zaico.nl/files/RUU-ITE_english.pdf. As condutas tipificadas constam do art. 27 a 34 da lei. As sanções são cominadas pelos arts. 46 e ss. da lei. Acesso em 20 mai. 2022.

²⁸³ KELLER, Clara Iglesias. Don't Shoot the Message: Regulate Disinformation Beyond Content. *RDP*, Brasília, v. 18, n. 99, p. 486-515, jul./set. 2021, p. 272.

²⁸⁴ A lei ameaça com sanção de dez a vinte anos quem publica, promove ou incita o ódio, discriminação ou violência baseada em identidade gênero, política, sexual, religiosa, étnica ou social. Cf. FISS, Joele; MCHANGAMA, Jacob. *The Digital Berlin Wall: How Germany (Accidentally) Created a Prototype for Global Online Censorship*. Copenhagen: Justitia, 2020. Mimeografado.

²⁸⁵ A íntegra da lei pode ser conferida em: <https://sso.agc.gov.sg/Act/POFMA2019?TransactionDate=20191001235959&ProvIds=P19-#pr62->. Mais especificamente sobre o ponto, dispõe a lei: “An internet access service provider that does not comply with an access blocking order under subsection (2) or (3) shall be guilty of an offence and shall be liable on conviction to a fine not exceeding \$20,000 for each day during any part of which that order is not fully complied with, up to a total of \$500,000 for each offence”.

regulatória sobre as redes sociais e censurar críticos do governo antes da realização das eleições nacionais”.²⁸⁶ A Rússia²⁸⁷ e a China²⁸⁸ também possuem legislação que segue este caminho.

Por fim, há legislações que buscam primordialmente a responsabilização dos intermediários por eventuais excessos ou negligência no exercício da moderação, tendo por referência não só o procedimento adotado, mas principalmente o conteúdo que tem o alcance limitado ou não.

No direito comparado, a Austrália tem exemplo de lei (*The Criminal Code Amendment Act 2019 — Sharing of Abhorrent Violent Material*), que reputa os intermediários da rede (dentre os quais se incluem as redes sociais) como responsáveis por falha na comunicação dentro de um prazo razoável à Polícia Federal Australiana de conteúdos violentos abomináveis (*abhorrent violent material*) do qual tenha ciência, bem como a falha em remover, de forma expedita (*expeditiously*) tais conteúdos de seus serviços. A sanção consiste em multas administrativas vultuosas. Há, portanto, um maior rigor na coerção de materiais ligados à divulgação de ataques terroristas ou crimes violentos, definidos no Código Penal Australiano, em detrimento de outros conteúdos, o que exigiria a adoção de procedimentos mais “expeditos”.

289-290

Por fim, a Alemanha tem a mais famosa das legislações sobre o tema — o *Network Enforcement Act (NetzDg)* — que, embora priorize normas orientadas ao aprimoramento do

²⁸⁶ NEO, Ric. The Securitisation of Fake News in Singapore. *International Politics*, [s.l.], v. 57, p. 724-740, 2020. Disponível em: <https://link.springer.com/article/10.1057/s41311-019-00198-4>. Acesso em: 20 dez. 2022.

²⁸⁷ É o caso da *Federal Law of Information, Information Technologies and Information Protection (The Information Law)*, que prevê a possibilidade de usuários serem sancionados com multas e prisões administrativas (se as condutas não configurarem crimes) ou prisões judiciais, no caso de reconhecimento de responsabilidade criminal por conteúdos divulgados. As intermediárias podem ainda ser responsabilizadas administrativamente (e sofrerem multas) se falharem no seu dever de impedir a disseminação de informações para propósitos ilegais. O escopo da norma é bem amplo, abarcando não só *fake news*, mas também casos de violência, assédio, discursos de ódio, pornografia, terrorismo e demais conteúdos considerados ilegais no país. Um impacto negativo desta norma foi exemplificado quando da apresentação do conceito de censura colateral. Para rememorar, o governo russo definiu que o uso do termo “guerra” para nominar o conflito que se desenrola entre Rússia e Ucrânia configuraria uma espécie de informação falsa. Para mais informações, confira-se: HEBERT SMITH FREEHILLS. *Race to Regulate: Online Harms*. [S.l.]: Herbert Smith Freehills, 2019. Disponível em: <https://www.herbertsmithfreehills.com/latest-thinking/online-harmful-content>. Acesso em: 23 mai. 2022.

²⁸⁸ É o caso da *Cybersecurity Law*, art. 47 (a referida lei, vale dizer, embora trate incidentalmente de falsidades e controle de informações, é amplamente focada na questão da proteção de dados pessoais).

²⁸⁹ A norma é fruto de uma emenda ao código penal daquele país e teve como motor a transmissão ao vivo no Facebook do ataque terrorista contra mesquitas muçulmanas por um supremacista branco na cidade de Christchurch, na Nova Zelândia. O evento resultou em 51 mortos e 40 feridos, e a lentidão da plataforma para derrubar o referido conteúdo (que viralizou em poucos minutos) e interromper a transmissão ao vivo foi alvo de duras críticas.

²⁹⁰ Recentemente, o Governo Australiano aprovou o *Online Safety Act (2021)*, que prevê a possibilidade de responsabilização de intermediários por negligência na derrubada de conteúdos pornográficos (não consensuais), que configurem *cyberbullying*, e conteúdo *online* seriamente prejudicial ou, ainda, negligência no bloqueio de páginas ou endereços IP com materiais que promovam conteúdos violentos abomináveis. A norma entrou em vigor no início de 2022. Para mais informações, *vide* *ibid*.

funcionamento das redes sociais (voltada à arquitetura das redes), traz também regras que buscam inviabilizar a circulação de certos tipos de discurso (com foco no conteúdo), sob pena de responsabilização administrativa dos intermediários. Diferentemente dos outros países, no entanto, a Alemanha não criou tipos penais ou conceitos jurídicos novos para abordar temas como desinformação, divulgação de material violento, terrorismo ou de assédio. Em verdade, a lei cria para as plataformas uma obrigação de remoção de conteúdo “manifestamente ilícito” no prazo de 24 horas, e de conteúdo “ilícito”²⁹¹ no prazo de sete ou mais dias a contar da reclamação, sob pena de multa ao intermediário. A referência para avaliação da licitude/ilicitude da conduta é a própria legislação vigente e os tipos penais já previstos naquele país (daí a lei ser chamada — numa tradução livre — de “lei para a melhoria da aplicação da lei na *internet*”).

Ao focar em conteúdo específicos, que mereçam especial restrição, normas que trabalhem com o conceito de “informações falsas” ou *fake news*²⁹² traduzem um risco evidente de *chilling effect* (quando o foco é o usuário) e censura colateral (quando o foco é o intermediário) ante a vagueza das suas expressões e a ausência de um parâmetro sólido para avaliar o que é verdade ou não, uma discussão que traz dificuldades similares ao reconhecimento de um “direito ao esquecimento” (abordado no capítulo 2.4.1), qual seja, o fato de não se tratar ainda de expressão com um significado consolidado pela experiência compartilhada, que permita uma compreensão adequada sobre seu conteúdo.

No caso de *fake news*, o fenômeno é ainda mais grave, porquanto o uso da expressão está em disputa política²⁹³. De fato, embora o termo tenha caído no uso da imprensa norte-americana crítica ao ex-Presidente Donald Trump para fazer referência a informações falsas por ele disseminadas, o próprio presidente e seus aliados se valiam da expressão para

²⁹¹ A diferença entre o conteúdo manifestamente ilícito e o ilícito é que, neste último caso, a decisão sobre ilicitude depende da “inverdade de uma alegação de fato” ou “de outras circunstâncias de natureza factual” ou, ainda, quando a rede social remeter a ilicitude da decisão para a “instituição de regulação autorregulada”. Nestes casos, o prazo é maior, pois há a exigência de se facultar ao prejudicado a oportunidade de defesa e posicionamento no âmbito antes da tomada de decisão (§ 3º, 3). Cf. EIFERT, Martin. Lei Alemã para a Melhoria da Aplicação nas Redes Sociais (Lei de aplicação da Lei nas redes – NetzDG). Trad. Pedro Henrique Ribeiro e Revisão de Júlio Carvalho. In: ABOUD, Georges; CAMPOS, Ricardo; NERY JR., Nelson (Orgs.). *Fake News e Regulação*. 2. ed. São Paulo: Revista dos Tribunais, 2020.

²⁹² Embora o enfoque maior da dissertação seja a desinformação, o mesmo questionamento poderia ser estendido para expressões tais como “conteúdo violento abominável”, “*ciberbullying*” e afins.

²⁹³ Neste sentido, é possível afirmar — com bases nas lições de Ernesto Laclau — que a expressão *fake news* é um significado flutuante, isto é, um termo ainda sob disputa política. Cf. FARKAS, Johan; SCHOU, Jannick. *Post-Truth, Fake News and Democracy: Mapping the Politics of Falsehood*. Londres: Routledge, 2020, p. 73-79.

descreditar notícias veiculadas nos canais tradicionais da mídia, como estratégia retórica para neutralizar eventuais informações que fossem negativas ao governo.²⁹⁴

No entanto, a sociedade civil não pode ignorar o fato de que a disseminação de informações falsas é um fato presente no nosso cotidiano e, portanto, tema merecedor de estudos e de um conceito que permita compreender adequadamente não só seu funcionamento, mas também que traga elementos que permitam definir quando sua produção ou disseminação é passível de punição.²⁹⁵ Daí a academia, em especial, abandonar o uso da expressão *fake news* e passar, cada vez mais, a trabalhar com a expressão de “desordens informacionais”, que abarcaria tanto a desinformação, como a *misinformation* e a *mal-information*.

Conforme antecipado em passagens anteriores, a desinformação se distinguiria das demais não só pela falsidade do conteúdo, mas especialmente pela intenção de se produzir o dano, evidenciada pela aparência jornalística do conteúdo, seja para angariar lucro ou para obter vantagens políticas. A intenção de produzir o dano é também presente na *mal-information*, que, no entanto, não traria um conteúdo propriamente falso, mas sim a veiculação de informação discriminatória ou vazamento de fatos descontextualizados ou afeitos à vida privada com o intuito de assediar ou prejudicar terceiro. A *misinformation*, pois embora faça referência a um fato falso, não há intenção de causar dano. A pessoa genuinamente crê na veracidade daquilo que se alega, mas acaba contribuindo para o ciclo de desordem informacional, em razão de um excessivo enviesamento ou por ignorância sobre aquilo que se expressa.

Há uma inclinação, portanto, em aproximar o conceito de desinformação nas redes sociais a uma premissa reconhecida pela doutrina nacional²⁹⁶ e por precedentes de tribunais estrangeiros²⁹⁷, quando aborda os limites da liberdade de imprensa e informação. Trata-se da

²⁹⁴ Daí o paradoxo da expressão *fake news* ter sido considerada a palavra do ano em 2017, pelo dicionário Oxford, em razão do seu uso exacerbado pelo Presidente Trump, embora o próprio fosse corriqueiramente acusado de produzir informações falsas. Informação disponível em: ‘FAKE News’ é eleita palavra do ano e vai ganhar menção em dicionário britânico. *GI*, [s.l.], 2 nov. 2017. Disponível em: <https://g1.globo.com/educacao/noticia/fake-news-e-eleita-palavra-do-ano-e-vai-ganhar-mencao-em-dicionario-britanico.ghtml>. Acesso em: 24 mai. 2022.

²⁹⁵ A própria “democracia” é em si um conceito sob constante disputa política. Há uma nítida disputa entre aqueles que enfatizam seu aspecto formal (vontade da maioria) com aqueles que associam a democracia a um componente liberal de proteção das minorias, por meio do necessário respeito aos direitos fundamentais. Esta disputa nunca impediu que estudiosos se debruçassem sobre os elementos que integrariam este conceito.

²⁹⁶ Por todos, BARROSO, Luís Roberto. Interpretação constitucionalmente adequada do Código Civil e da Lei de Imprensa. *Revista de Direito Administrativo*, Rio de Janeiro, v. 235, p. 1-36, jan./mar. 2014, p. 18. A verificação da culpa ou dolo, segundo o autor, decorre da observância de seis parâmetros. A saber: a) a veracidade subjetiva dos fatos (ao qual já referenciamos como a exigência de comprovação do dolo ou culpa grave na divulgação de informação falsa); b) licitude no meio empregado para obtenção de informação; c) personalidade pública ou privada da pessoa objeto de notícia; d) local do fato; e) natureza do fato; f) existência de interesse público na divulgação.

²⁹⁷ O referido parâmetro encontra nítida inspiração na doutrina da *actual malice* do direito norte-americano. A lembrar, segundo o parâmetro construído no caso *New York Times vs Sullivan* (ação ajuizada por um funcionário

noção de “veracidade subjetiva”, compreendida aqui por um dolo ou culpa grave na divulgação de um fato noticioso, posteriormente descoberto como falso, e que visa responsabilizar o veículo de imprensa ou jornalista, tão somente, se houver comprovável negligência na pesquisa de fontes e obtenção do material noticiado.

Há, no entanto, algumas peculiaridades ao tratar da desinformação na *internet*. Se, ao responsabilizar jornalistas ou canais de imprensa por veiculação de notícias falsas, é possível considerar a conduta jornalística como adequada por meio da avaliação de parâmetros próprios de atuação daqueles que exercem a atividade profissional, exigir tal rigor no âmbito das redes sociais pode soar excessivo e provocar um efeito resfriador desmedido nestas plataformas. Uma outra dificuldade decorre do fato de que as desinformações nas redes nem sempre são direcionadas ao ataque da reputação de uma ou mais pessoas determinadas, o que torna a judicialização um mecanismo muitas vezes inútil. A desinformação é, muitas vezes, a publicização de um fato falso que — se disseminado — pode causar dano a toda a coletividade.

A imaginar, por exemplo, que um grupo de antivacinas nas redes sociais divulguem *sites*, informações e *links* de duvidosa credibilidade que vinculem as vacinas contra a COVID ao surgimento de novas doenças. Por mais que seja evidente a necessidade de que os principais canais de imprensa, instituições de checagem de fatos, autoridades públicas, especialistas e pesquisadores orientem a coletividade quanto à falsidade das afirmações²⁹⁸, não é possível argumentar pela responsabilização daqueles que contribuíram para a disseminação com base em uma mera negligência. Em verdade, é preciso haver sólida evidência da má-fé, isto é, que o conteúdo foi compartilhado para obtenção de alguma vantagem material ou política. Por isso, é absolutamente censurável, em um contexto democrático, a construção de tipos penais como os adotados na Malásia, Indonésia ou Singapura, que considera a divulgação de fato noticioso falso (atrelado ou não a um prejuízo de interesse público genericamente considerado) como elemento suficiente para a responsabilização de qualquer usuário ou como ferramenta de intimidação dos intermediários da *internet*. Considerando que a definição daquilo que é

público visando a responsabilização do jornal New York Times por veiculação de notícias difamatórias a seu respeito). Embora a doutrina tenha se originado nos EUA, parâmetros similares — e com nítida inspiração no referido julgado — foram utilizados na Espanha, Inglaterra, Itália, Austrália e, inclusive, em Cortes Internacionais de Direitos Humanos. No âmbito do STF, embora se encontrem diversos precedentes que façam referência à primazia *prima facie* da liberdade de expressão (o que aproxima o Brasil da noção norte-americana ao menos quando se fala em “liberdade de informação e imprensa”), tal parâmetro é muitas vezes esquecido nas instâncias ordinárias. Para mais informações, cf. CHEQUER, Cláudio. *A Liberdade de Expressão como Direito Preferencial Prima Facie*: análise crítica de revisão ao padrão jurisprudencial brasileiro. 2. ed. Rio de Janeiro: Lumen Juris, 2017, p. 124-139.

²⁹⁸ A nosso ver, seria inclusive legítimo exigir de redes sociais que etiquetem ou que adotem medidas para reduzir a amplificação desse tipo de conteúdo. O ponto será aprofundado no último capítulo.

verdadeiro ou falso é fruto de um contínuo debate racional e reflexivo²⁹⁹, não é difícil imaginar que criminalizar a desinformação com base em uma suposta “falsidade” a ser apontada por órgãos estatais pode vir a se tornar pretexto para perseguição e censura de críticos, seja por meio de ações legislativas diretas (mediante criminalização da conduta de usuários), seja por meio da censura colateral (mediante pressão de multas exacerbadas sobre as plataformas de redes sociais).

Isso não implica, contudo, abraçar aqui uma concepção radicalmente relativista de que todas as verdades sejam parciais e dependem da perspectiva de uma pessoa. Há coisas que simplesmente não podem ser equiparadas. Valendo-se das palavras de Michiko Kakutani, “não há como igualar pontos de vista quando um lado representa um consenso esmagador e outro é praticamente uma aberração na comunidade científica”³⁰⁰, tampouco é correto dizer que, necessariamente, “todos os lados têm o mesmo valor”³⁰¹. Democracia não é sinônimo de neutralidade e é preciso que afirmações se sustentem em uma base fática mais ou menos detectável.³⁰²

Em verdade, o relativismo moral e a disseminação desenfreada de mentiras, especialmente quando patrocinada por agentes políticos, para semear dúvidas e ataques a qualquer base comum de concordância é tão ruim à democracia quanto o peso da censura estatal que cala a liberdade daqueles que ousam desafiar quem esteja no poder. Como bem percebeu Orwell, “quando a verdade é tão fragmentada e relativa abre-se caminho para que um líder ou grupo dominante diga em que se deva acreditar: ‘Se o líder diz que determinado evento ‘nunca aconteceu’ — bem, aquilo nunca aconteceu’”.³⁰³

Por essa razão, defende-se aqui que, se há algum espaço para responsabilização por desinformação, é para fins de repreensão à conduta de agentes políticos que se valem

²⁹⁹ Aqui, cumpre referenciar a noção de “livre mercado de ideias” ou de “esfera pública” a que aludimos no capítulo 1 (itens 1.2.1 e 1.2.2).

³⁰⁰ Cf. KAKUTANI, Michiko. *A morte da Verdade: notas sobre a mentira na Era Trump*. Trad. André Czarnobai e Marcela Duarte. Rio de Janeiro: Intrínseca, 2018, p. 89.

³⁰¹ A crítica também é de Michiko Kakutani ao comentar declaração de Trump que tentou equiparar as pessoas que se manifestavam contra supremacistas brancos aos neonazistas que haviam se reunido em Charlottesville, na Virgínia, para protestar contra a remoção das estátuas dos soldados confederados. Havia “gente boa nos dois lados” declarou Trump. Afirma a jornalista que igualar vítima ao agressor é “criar uma falsa equivalência factual que o torna cúmplice dos crimes e consequências mais abomináveis” (Ibid., p. 87 e 90).

³⁰² Segundo Hannah Arendt, o “súdito ideal do governo totalitário não é o nazista convicto nem o comunista convicto, mas aquele para quem já não existe a diferença entre fato e ficção (isto é, a realidade da experiência) e entre o verdadeiro e o falso (isto é, os critérios do pensamento)”. Cf. ARENDT, Hannah. *As Origens do Totalitarismo*. Trad. Roberto Raposo. São Paulo: Companhia das Letras, 2012, p. 632.

³⁰³ ORWELL, George. *Looking Back on the Spanish War*. [S.l.]: Orwell Foundation, 1943. Disponível em: <https://www.orwellfoundation.com/the-orwell-foundation/orwell/essays-and-other-works/looking-back-on-the-spanish-war/>. Acesso em: 25 mai. 2022.

sistematicamente de práticas de “censura reversa” para produzir desinformações que mirem instituições democráticas.

Há alguns argumentos fáticos e jurídicos para amparar este posicionamento. Dentre os argumentos fáticos, reforça-se a evidência de que mesmo Estados com práticas censoras estabelecidas (tais como China e Rússia) têm se valido cada vez mais do expediente da produção de desinformação para incitar a divisão e o conflito (especialmente, para influenciar processos eleitorais em outras nações) e silenciar opositores. Remete-se aqui às considerações já apresentadas ao capítulo 2.4.1.

É possível, ainda, adicionar um argumento trazido pela Psicologia. Segundo Jonathan Haidt, o jeito mais fácil de se desfazer intuições e construções emocionais elaboradas com base no nosso sistema cognitivo (o sistema nº 1) é por meio do que ele denomina de “persuasão social” e “persuasão racionalizada”.³⁰⁴

A “persuasão social” depende da respeitabilidade e apreço que o falante gera ao seu interlocutor. Assim, somos naturalmente mais atentos à opinião de um familiar ou amigo, um professor ou um político pelo qual se nutre apreço. A partir desta aproximação e uma ligação empática entre falante e interlocutor, é possível que os argumentos racionais (“persuasão racionalizada”) sejam efetivamente ouvidos e possibilitem um diálogo plural pautado por tolerância, civilidade e razoabilidade.

A constatação destes estudos evidencia por qual razão a liberdade de expressão de agentes políticos merece especial atenção do Direito. A especial legitimidade que o processo eleitoral confere a estes agentes cria um vínculo especial de afetividade entre eleitos e o restante da coletividade, o que, por certo, demanda maior responsabilidade e deveres destes agentes, tanto na promoção da tolerância e civismo, quanto na valorização da verdade.

Para que se escape do cenário caótico pintado por Giuliano da Empoli,³⁰⁵ as redes sociais não podem ser um local por meio do qual agentes políticos manipulam as estruturas algorítmicas da esfera digital para prestigiar conteúdos sensacionalistas, teorias conspiratórias e discursos de intolerância, que esgarcem o tecido social, confirmem preconceitos e produzam divisão, pondo em risco as bases da democracia. Antes, deve ser um espaço no qual agentes políticos, cientes das falhas propiciadas pelo modelo da economia de atenção explorado pelas redes sociais, assumam um compromisso de não promover o diversionismo e a intolerância,

³⁰⁴ HAITD, Jonathan. *A Mente Moralista: por que as pessoas boas se separam por causa da política e da religião?* [S.l.]: Alta Cult, 2019, p. 59-61.

³⁰⁵ Vide nota de rodapé nº 241.

respeitando as noções de pluralidade, tolerância, civismo e o progresso social, fins últimos da proteção especial conferida à liberdade de expressão.

Em síntese, se é correto afirmar que a liberdade de expressão do agente político deva ser qualificada por força do interesse público subjacente ao exercício da sua função, é certo também que são maiores suas responsabilidades na manifestação de opiniões quando se valem das redes sociais para se tornarem um canal direto de comunicação com os cidadãos. Neste contexto, tolerância, zelo e razoabilidade deixam de ser mera virtudes a serem perseguidas e tornam-se deveres jurídicos essenciais à preservação da democracia.

Sob o ponto de vista jurídico, eventual legislação que impute responsabilização a agentes políticos encontra guarida especialmente no princípio da “democracia militante”, que teria sido abarcado como cláusula implícita pelo art. 17 da Constituição da República. Com efeito, se a própria participação no sufrágio eleitoral requer, necessariamente, o ingresso em partido político que respeite o regime democrático e os direitos fundamentais, parece razoável conceber leis que responsabilizem agentes políticos que se valham da especial posição que lhe foi conferida pelo próprio processo democrático para divulgar informações falsas a respeito de opositores e demais instituições essenciais à preservação do regime democrático, como uma ferramenta de “autodefesa” da própria democracia.³⁰⁶

Não se ignora aqui que a ideia de uma autodefesa democrática a exigir um modelo militante de democracia enseja riscos, pois o próprio conceito de “democracia” é relativamente cambiante e pode vir a ser ocasionalmente capturado³⁰⁷. Por essa razão, não se vai tão longe a ponto de sustentar que o princípio da democracia militante — implícito ao art. 17 da Constituição Federal — possa justificar arroubos judiciais que viabilizem cassação de mandatos ou sancionamento de manifestações sem o devido amparo legal, sob pena de instituir-se um *Macarthismo* às avessas³⁰⁸.

O princípio, contudo, é útil para sustentar a constitucionalidade de leis que busquem responsabilizar administrativa, política ou criminalmente agentes políticos que atentem contra o Estado Democrático do Direito, por meio de manifestações odiosas ou falsas. Há que se

³⁰⁶ Daniel Sarmiento e João Gabriel Madeira Pontes partem de argumentação jurídica similar para sustentar, inclusive, um limite à garantia das imunidades parlamentares. SARMENTO, Daniel; PONTES, João Gabriel Madeira. Democracia Militante e Imunidade Material dos Parlamentares: Limites Constitucionais aos Discursos de Deputados e Senadores. *Revista da AJURIS*, Porto Alegre, v. 47, n. 149, dez. 2020, p. 68-93.

³⁰⁷ Frise-se que a única vez que o princípio fora utilizado para cancelamento de registro de candidatura se deu por decisão do Tribunal Superior Eleitoral, em 1947 (no contexto da Guerra Fria) em prejuízo do PCB (Partido Comunista Brasileiro), sob o argumento de que a Constituição de 1946 consagrava uma democracia militante.

³⁰⁸ O termo se refere a um período que perdurou durante a década de 50 nos Estados Unidos, de acentuada perseguição política a opositores do governo sob a acusação de serem comunistas ou simpatizantes do regime soviético.

considerar, por óbvio, uma série de cautelas na elaboração deste tipo de infração, como na escolha dos verbos que irão compor o tipo do ilícito de modo a reduzir, tanto quanto possível, a margem para juízos discricionários na sua interpretação, bem como é recomendável que seja especificado o dano almejado pela desinformação³⁰⁹ como cautela para redução de sua generalidade.³¹⁰

Com efeito, a criação de sanções penais ou administrativas de veiculação de “fake news” ou “desinformação”, com caráter amplo e geral, não são recomendadas por se tratar de conceitos em construção e pelo risco de que a noção de “verdade” seja transformada em um pretexto para perseguição de usuários críticos ao governo. O único espaço em que nos parece possível sustentar a possibilidade de regulação de conteúdo — com base no princípio constitucional da democracia militante — é na definição de condições para punição de agentes políticos, que se valham sistemática e dolosamente de medidas de censura reversa para produzir danos específicos ao Estado Democrático de Direito.

Sem embargo, reforça-se: a intervenção regulatória estatal deve ser prioritariamente orientada à alteração da arquitetura das redes sociais. Na esteira das lições de Lessig, “todo controle social e regulamentação governamental do mundo digital passa pela necessidade de assimilação do código”³¹¹ e, continua, “é o código que incorpora certos valores e impossibilita a realização de outros valores atuando como um modelador de expectativas cotidianas”.³¹² Não há regulação eficiente no mundo digital que não perpassasse pela avaliação dos fins que inspiram a construção de códigos e algoritmos na *internet*.

No caso das redes sociais, o fundamento para esta intervenção decorre da constatação trazida no capítulo anterior de que as ameaças às liberdades fundamentais não decorrem

³⁰⁹ Até porque nem toda mentira é capaz, por si só, de causar dano significativo que justifique a criminalização.

³¹⁰ A nova lei de defesa do Estado Democrático de Direito (Lei nº 14.197/2021) — uma materialização legislativa do princípio da democracia militante — previu, de forma mais específica, em sua redação originária o crime de “Comunicação Odiosa em Massa”, que tinha a seguinte redação: “Promover ou financiar, pessoalmente ou por interposta pessoa, mediante uso de expediente não fornecido diretamente pelo provedor de aplicação de mensagem privada, campanha ou iniciativa para disseminar fatos que sabe inverídicos, e que sejam capazes de comprometer a higidez do processo eleitoral”. O dispositivo foi, contudo, vetado pelo atual presidente e um dos argumentos foi justamente a amplitude do verbo “promover”, pois não deixava claro se a conduta objeto de criminalização seria daquele que gerou a notícia (responsável pela desinformação) ou daquele que a compartilhou (*misinformation*). Fonte: PRESIDÊNCIA DA REPÚBLICA. *Mensagem nº 427, de 1º de setembro de 2021*. Brasília, DF: Presidência da República, 2021. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2019-2022/2021/Msg/VEP/VEP-427.htm#:~:text=%E2%80%9CA%20proposi%C3%A7%C3%A3o%20legislativa%20estabelece%20como,para%20disseminar%20fatos%20que%20sabe. Acesso em: 26 mai. 2022. Uma forma interessante de delimitar o alcance do tipo — que se dirigia a todo e qualquer usuário — seria restringi-lo aos agentes políticos que se valem das ações elencadas no referido tipo penal, isentando de responsabilização demais usuários, que poderia ser a “interposta pessoa” integrante do tipo, mas não o sujeito passivo da sanção criminal.

³¹¹ LESSIG, Lawrence. The Law of the Horse: What Cyber Law Might Teach. *Harvard Law Review*, [s.l.], v. 113, p. 501-546, mar. 1999, p. 506.

³¹² *Ibid.*, p. 510.

exclusivamente do Estado, mas da própria relação entre particulares em razão do modelo econômico desenvolvido pela maior parte das plataformas de redes sociais. O desafio, contudo, é descobrir como obter uma regulação de plataformas digitais apta a mitigar os riscos elencados nesta dissertação se o Estado carece do completo conhecimento sobre o funcionamento dos códigos, algoritmos e ferramentas de inteligência artificial, e quase nunca será capaz de dar respostas rápidas à realidade extremamente dinâmica do mundo digital.³¹³

A busca por tal resposta parte da necessidade de construção de parâmetros (*standards*) de direitos fundamentais para a ação destas plataformas privadas e a priorização de instrumentos de ação legislativa indireta³¹⁴ que reconheçam, em alguma medida, a responsabilidade de tais plataformas por conteúdos divulgados por seus usuários, impondo-lhe alguns deveres de cuidado e transparência, de modo a reorientar a noção de “governança das plataformas” (e a construção dos próprios códigos) a parâmetros constitucionalmente aceitáveis.

O objetivo dos próximos tópicos é enriquecer a análise com a avaliação de arranjos regulatórios, que buscam justamente inculcar parâmetros de atuação de respeito aos direitos fundamentais ao exercício do poder de governança das redes sociais. O enfoque será na legislação alemã (*NetzDG*) outrora mencionada, e compará-la com proposta de autorregulação apresentada pela principal empresa do setor (a Meta, que administra o Facebook e o Instagram).

A pretensão é contrapor o modelo concebido pela lei alemã, que se vale de um arranjo corre regulatório, com a mais sofisticada proposta já apresentada por uma plataforma privada de regulação da moderação, qual seja, o *Oversight Board*, um órgão supostamente autônomo que teria o condão de revisar as decisões do Facebook e Instagram, sugerindo a adoção de soluções técnicas que mitiguem o problema da desinformação e disseminação do discurso de ódio nestas plataformas. A partir deste contraponto, é possível se avaliar as virtudes e lacunas de cada arranjo, tendo por fim obter indícios de uma resposta regulatória que enderece de forma adequada o problema da ausência de transparência e das externalidades negativas produzidas pelas redes sociais, sem que isto produza um cenário de censura estatal.

A escolha da lei alemã para o estudo de caso se deu por algumas razões. Em especial, por ser uma lei pioneira no trato da matéria: (I) é a primeira lei de um país democrático que vincula expressamente as plataformas à observação das normas que limitam a liberdade de expressão no âmbito do país; (II) é a primeira lei a conceber a corre regulação como estratégia

³¹³ Sobre os limites de eventual ação regulatória estatal, confira-se o disposto no capítulo 4.4.

³¹⁴ LESSIG, Lawrence. The limits in Open Code: Regulatory Standards and Future of Net. *Berkeley Law Technology Journal*, [s.l.], n. 759,1999, p. 763.

para lidar com os desafios da governança de discursos nas redes sociais; (III) é a primeira lei a orientar a arquitetura das redes no exercício da moderação de conteúdo. O segundo motivo é o fato de a lei ser corriqueiramente evocada como inspiração, tanto por países com tradição democrática (vide casos do Reino Unido e França), como por países que não possuem um regime democrático consolidado (Indonésia, Singapura, Rússia, Venezuela e Honduras), o que gera alguns mal-entendidos sob seu escopo.³¹⁵ A terceira razão se conecta à segunda, qual seja, o fato de o *NetzDG* ser a principal inspiração do Projeto de Lei nº 2630/2020, o mais avançado projeto em trâmite no Congresso Nacional, que visa criar um modelo regulatório para o exercício da liberdade de expressão no âmbito das redes sociais no Brasil.³¹⁶

No entanto, antes de avançar sobre este ponto, é preciso fazer um breve desvio de curso para aprofundamento das distinções conceituais entre autorregulação, autorregulação regulada e correção, termos-chave para compreensão das conclusões que serão apresentadas ao fim deste capítulo.

3.2 Aclaramentos conceituais: autorregulação, correção e autorregulação regulada

A perda pelo Estado do monopólio do poder regulatório é uma constatação decorrente do avanço do poder de influência de inúmeros atores sociais sobre aspectos da vida privada dos cidadãos em geral.

O Estado pós-moderno, na visão de Chevallier, é um novo ator político, cujo poder soberano é *enquadrado* por uma base de valores jurídicos (o Estado de Direito), políticos (predomínio das Democracias Liberais) e econômicos (superioridade dos mecanismos de mercado) a serem observados; *rivalizado* por grandes conglomerados econômicos e agentes privados com quem mantém relações ambivalentes de poder (as grandes empresas e organizações privadas precisam dos Estados para se consolidar em novos mercados e os Estados precisam destes operadores econômicos para proteger empregos e assegurar o desenvolvimento

³¹⁵ A influência da legislação alemã pode ser percebida ao menos em 17 (dezesete) países. Cf. FISS, Joele; MCHANGAMA, Jacob. *The Digital Berlin Wall: How Germany (Accidentally) Created a Prototype for Global Online Censorship*. Copenhagen: Justitia, 2020. Mimeografado.

³¹⁶ Confira-se: SCHREIBER, Mariana. A controversa lei alemã que inspira o projeto de lei das Fake News. *O Globo*, [s.l.], 26 ago. 2020. Disponível em: <https://oglobo.globo.com/epoca/brasil/a-controversa-lei-alema-que-inspira-projeto-de-lei-das-fake-news-1-24606576>. Acesso em: 26 mai. 2022.

econômico), e *englobada* pela necessidade de constituição de espaços ampliados de regulação, capazes de lidar com a dimensão transnacional do mercado.³¹⁷

O fortalecimento das relações de interdependência entre setor público e privado eleva outros atores sociais, outrora percebidos como mero destinatários de normas nacionais e internacionais, ao papel de coautores, haja vista sua capacidade ampliada de exercer pressão junto ao Poder Público.³¹⁸ Soma-se, a este aspecto, a vetusta crítica sobre a incapacidade de o Estado lidar com a informação dispersa na sociedade, como argumento justificador de uma postura não-intervencionista do Poder Público. Segundo essa linha, há uma presunção relativa de que espaços ampliados de regulação privada seriam mais eficientes, já que usualmente mais informados acerca daquilo que se regula.³¹⁹

Cresce, assim, o fenômeno da autorregulação privada ou, nos dizeres de Julia Black, um modelo descentralizado de regulação³²⁰, onde o poder de influência estatal sobre a vida privada — fruto de normas públicas em geral — passa a concorrer cada vez mais com espaços privados de influência sobre terceiros. É neste sentido que a referida autora passa a definir a regulação como “um processo, que envolve uma tentativa focada e sustentada de tentar alterar ou influir no comportamento de terceiros, de acordo com parâmetros ou *standards* estabelecidos, para alcançar resultados bem definidos”.³²¹

Há uma desvinculação do conceito de regulação da figura do Estado. A heterorregulação estatal é apenas mais um arranjo regulatório, que coexiste com outros modelos como o da autorregulação privada.

Há alguma celeuma, contudo, na definição do alcance da expressão “autorregulação”. Uma concepção mais estrita vê a autorregulação como um espaço privado de orientação de condutas adotado de forma *voluntária* por particulares, no qual a ingerência do Poder Público sobre a esfera de atuação do particular é nula ou reduzidíssima (limitado a decisões judiciais casuísticas ou acordos de cooperação eventuais)³²². Há concepções mais amplas, no entanto, que não consideram a autorregulação como um modelo necessariamente voluntário. Segundo Saddy, o termo englobaria quatro elementos: (I) o estabelecimento de padrões de

³¹⁷ Cf. CHEVALLIER, Jacques. *O Estado Pós-Moderno*. Trad. Marçal Justen Filho. Belo Horizonte: Fórum, 2009, p. 38-54.

³¹⁸ *Ibid.*, p. 48.

³¹⁹ HAYEK, Friedrich von August. The Use of Knowledge in Society. *The American Economic Review*, [s.l.], v. 35, n. 4, set. 1945.

³²⁰ BLACK, Julia. Decentring Regulation: Understanding the Role of Regulation and Self-Regulation in a “Post-Regulatory” World. *Current Legal Problems*, Oxford: Oxford University Press, v. 54, n. 1., p. 103-146, jan. 2001.

³²¹ *Ibid.*

³²² Sentido que parece ser o adotado por Clara Iglesias Keller em sua tese de doutorado. Cf. KELLER, Clara Iglesias. *Regulação Nacional de Serviços na Internet: Exceção, Legitimidade e o Papel do Estado*. Rio de Janeiro: Lumen Juris, 2019, p. 183.

comportamento, (II) criado por um ente extraestatal ou não, (III) por meio de um documento escrito (IV) que autolimite a vontade de quem os elabora, adere, aprova ou subscreve.³²³ Sob essa ótica, a correção também seria uma espécie de modelo autorregulatório, no qual seria atribuído maior poder de ingerência ao Poder Público, que poderia limitar a autonomia de agentes privados a parâmetros gerais, fiscalizá-los ou condicionar os *standards* privados a atos de aprovação pelo Poder Público.

A concepção mais estrita (que agrega o requisito da voluntariedade) é a adotada neste trabalho por parecer a que melhor explica as limitações da autorregulação para lidar com questões de direitos fundamentais no âmbito da *internet*. Nos parece mais útil tratar a correção como uma espécie autônoma do arranjo autorregulatório, principalmente porque o fato determinante para avaliação das vantagens e desvantagens de um ou outro modelo é, justamente, a menor ou maior vinculação do agente privado às pautas públicas na construção da sua regulação, elemento que distingue as duas espécies regulatórias.

De fato, o modelo autorregulatório fora muito festejado por permitir soluções mais eficientes ante o maior conhecimento técnico dos agentes econômicos para construir regras que melhor direcionem os riscos e falhas de mercados decorrentes de uma determinada atividade. Outras vantagens podem ser enumeradas: como se trata de parâmetro construído pelos próprios agentes privados, contaria com maior adesão e receptibilidade, sendo também uma ferramenta flexível, que permite rápidos rearranjos a um custo reduzido para o Estado.³²⁴

O modelo, contudo, possui algumas falhas. A principal é o conflito de interesses que pode acometer os agentes privados que se autorregulam. Como confiar a um particular o dever de se autolimitar se isso puder vir a lhe gerar prejuízos financeiros? Se a atividade econômica estiver intrinsecamente ligada à vulneração de direitos basilares, como viabilizar o arranjo autorregulatório?

Este é um cenário a ser considerado quando se fala de provedores de redes sociais e intermediários da *internet*. Se, no início, não era sabido ao certo os riscos e externalidades negativas decorrentes deste novo ecossistema informacional e, portanto, era compreensível a aposta em um modelo autorregulatório para impulsionar o surgimento de uma nova tecnologia, hoje já se sabe que a atividade-fim destas plataformas de redes sociais viabilizam-se economicamente por um modelo de atividade que depende do uso (muitas vezes, abusivo) de

³²³ SADDY, André. *Regulação Estatal, Autorregulação Privada e Códigos de Condutas e Boas Práticas*. Rio de Janeiro: Lumen Juris, 2019, p. 88.

³²⁴ Para maior detalhe sobre as vantagens da autorregulação, confira-se: BALDWIN, Robert; CAVE, Martin; LODGE, Martin. *Understanding Regulation*. 2. ed. Nova Iorque: Oxford University Press, 2012, p. 137-145.

dados de particulares e fomento à construção de enclaves informacionais, que contribuem para a amplificação de fenômenos como a “desinformação”, cujos efeitos sociais e políticos já foram exemplificados.

Soma-se isso ao fato de que o mercado de plataformas de redes sociais é bastante concentrado³²⁵. Neste cenário, mesmo soluções regulatórias aparentemente mais sofisticadas como o *Oversight Board* podem ser encaradas mais como mero mecanismos de relações públicas ou instrumento para a imposição de barreiras de entrada a *players* novos no mercado, do que uma séria tentativa de orientar racionalmente o debate sobre os limites do exercício da liberdade de expressão nas redes sociais.

Também seguindo esta esteira, Anthony Ogus destaca que modelos autorregulatórios puros costumam funcionar bem em cenários específicos onde é possível estabelecer uma concorrência legítima entre arranjos autorregulatórios ou, ainda, quando as externalidades produzidas por uma atividade econômica forem absorvidas por um grupo determinado e relativamente homogêneos de pessoas. Em casos em que o dano produzido à sociedade é difuso, o melhor talvez seja apostar em modelos regulatórios que permitam “um limitado grau de controle centralizado”.³²⁶

A correção é uma resposta a estas indagações e surge como um modelo híbrido entre a autorregulação privada e a tradicional regulação heteronormativa estatal, que, em tese, permitiria conciliar o melhor de cada um dos modelos. Ela se caracteriza pela “colaboração entre a regulação estatal e a autorregulação privada”,³²⁷ onde “há uma fundada divisão de tarefas e responsabilidade entre o Estado, agentes regulados e partes interessadas”³²⁸, no qual é possível que “os particulares sejam exortados a desenvolver ou dar concretude a um corpo de princípios básicos editados pelo Estado, por meio da lei ou de um regulamento, sob a supervisão de um regulador estatal”³²⁹.

³²⁵ Mesmo a resolução do problema de concentração de mercado no âmbito das plataformas de redes sociais não é vista como suficiente para mitigar as externalidades produzidas por suas atividades. Segundo Zuboff, ainda que o mercado tenha condições razoavelmente competitivas, estaríamos diante de diversas empresas que apoiaram seu lucro em um modelo de capitalismo que depende essencialmente da captura de dados de particulares para a construção de algoritmos que, de forma pouco transparente, moldam o comportamento de terceiros para atender suas próprias necessidades lucrativas. Cf. ZUBOFF, Shoshana. *A Era do Capitalismo de Vigilância: a luta por um futuro humano na nova fronteira do poder*. Rio de Janeiro: Intrínseca, 2020.

³²⁶ Neste sentido, cf. OGUS, Anthony. Rethinking Self-Regulation. *Oxford Journal of Legal Studies*, [s.l.], v. 15, n. 1, 1995, p. 97-108.

³²⁷ SADDY, André. *Regulação Estatal, Autorregulação Privada e Códigos de Condutas e Boas Práticas*. Rio de Janeiro: Lumen Juris, 2019, p. 112

³²⁸ BINENBOJM, Gustavo. *Poder de Polícia, Ordenação e Regulação: Transformações Político-Jurídicas, Econômicas e Institucionais do Direito Administrativo Ordenador*. 3. ed. Belo Horizonte: Fórum, 2020, p. 325.

³²⁹ *Ibid.*, loc. cit.

A vantagem deste modelo está no fato de que tal arranjo permite ao Estado “endereçar questões de governança e direitos humanos de forma mais transparente”³³⁰, pois aumenta o escrutínio público sobre o tratamento dos direitos fundamentais (*accountability*), sendo um modelo também mais suscetível a soluções experimentais, que viabilizam um meio termo entre a opção dual entre regular e não regular.

A opção corregulatória é especialmente interessante no âmbito das redes sociais, pois evita ainda a intervenção direta do Estado sobre a autonomia de quem discursa, algo que seria em tese vedado — para alguns — por força de uma *eficácia bloqueadora* dos direitos fundamentais³³¹, pois a opção por um modelo corregulador mantém a autonomia dos usuários que utilizam as redes sociais para se comunicar, sem afastar a possibilidade de a regulação pública orientar minimamente o poder de governança dos intermediários das redes a fins constitucionalmente legítimos.

Isto não implica esquecer que mesmo o modelo corregulatório pode apresentar falhas, especialmente em cenários de profunda assimetria informacional, onde é possível que empresas manipulem seus interesses sobre os da regulação, tornando tal sistema opaco e ineficiente. Tão importante quanto a corregulação é definir os pontos substanciais onde a intervenção estatal fiscalizatória e/ou sancionatória se faz necessária para reduzir tais assimetrias.

Por fim, cumpre abordar o conceito de “autorregulação regulada”. Segundo Binenbojm, o termo “autorregulação regulada” designa um “conjunto de arranjos em que a ordenação é exercida predominantemente por entidades privadas, com variações entre si quanto à forma, momento, e intensidade em que se perfaz o controle ou supervisão estatal”³³². Seria, portanto, “um gênero do qual a corregulação seria espécie”³³³, onde a ação estatal seria complementar ou residual.

Parte considerável da doutrina trata os termos quase como sinônimos.³³⁴ Em ambos, há um protagonismo compartilhado entre agentes públicos e privados, no qual caberia a este último

³³⁰ KELLER, Clara Iglesias. *Regulação Nacional de Serviços na Internet: Exceção, Legitimidade e o Papel do Estado*. Rio de Janeiro: Lumen Juris, 2019, p. 194.

³³¹ O ponto é defendido por Bruno Boquimpani Silva com base em precedente do STF, que afastou a exigência de qualificação profissional — como o diploma de jornalista — para o desempenho de atividade jornalística (STF, RE nº511.961. Min.Rel. Gilmar Mendes. Pleno, julgado em:17/06/2009, Dje: 13/11/2009.). Cf. SILVA, Bruno Boquimpani. *A autorregulação: delineamento de um modelo policêntrico de regulação jurídica*. 2010. Dissertação (Mestrado em Direito Público) – Faculdade de Direito, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2010, p. 68-69.

³³² BINENBOJM, Gustavo. *Poder de Polícia, Ordenação e Regulação: Transformações Político-Jurídicas, Econômicas e Institucionais do Direito Administrativo Ordenador*. 3. ed. Belo Horizonte: Fórum, 2020, p. 325.

³³³ *Ibid.*, p. 326.

³³⁴ Neste sentido parece indicar SADDY, André. *Regulação Estatal, Autorregulação Privada e Códigos de Condutas e Boas Práticas*. Rio de Janeiro: Lumen Juris, 2019, p. 116.

criar mecanismos e instrumentos que deem concretude a pautas gerais elaboradas pelo Estado, que poderão estar, ainda, sob a supervisão de regulador estatal ou de um ente privado reconhecido como legítimo para tanto.

No âmbito das redes sociais, é possível cogitar que a lei venha a criar uma distinção mais objetiva entre as duas expressões³³⁵, uma vez que o Projeto de Lei nº 2630/2020 — inspirado pela “legislação alemã de aplicação da lei nas redes sociais” que guarda disposição quase idêntica — utiliza o termo “autorregulação regulada” para referenciar a possibilidade de os provedores de redes sociais criarem uma instituição privada cujo objetivo precípua seja assegurar as diretrizes públicas fixadas no referido projeto de lei.³³⁶⁻³³⁷

Parte-se, portanto, da premissa de que em toda regulação um ou mais atores buscam fixar padrões de comportamento, de forma intencional, para alcançar finalidades pré-estabelecidas, influenciando a ação de terceiros ou de si próprios. O que diferirá um modelo do outro é o grau de intervenção estatal. Na autorregulação privada, o Estado não estabelece parâmetros, tampouco participa do monitoramento de forma sistêmica, havendo ampla autonomia ao ator ou atores privados afetados pelo documento autorregulatório. Na correção, há uma divisão de protagonismo entre Estado e agentes privados. Ao primeiro, caberá definir pautas e finalidades gerais que limitam a autonomia privada. Ao segundo, cabe minuciar a forma para cumprimento daquelas premissas. O monitoramento dos fins gerais da lei — uma espécie de *compliance* externo — será feito por um órgão ou ente estatal (correção em sentido estrito) ou uma pessoa jurídica de direito privado, que será reconhecida pelo Poder Público como competente para tanto (autorregulação regulada).

³³⁵ Distinção essa que — se adotada — aproxima-se do conceito trazido por Binbenojm, que vê na autorregulação regulada um arranjo com menor grau de intervenção do Poder Estatal.

³³⁶ Confira-se a íntegra do projeto de lei em: <https://www.camara.leg.br/propostas-legislativas/2256735>.

³³⁷ Figura jurídica similar é a do Operador Nacional do Sistema Elétrico (art. 13 da Lei nº 9648/98), pessoa jurídica privada, sem fins lucrativos, a quem cabe a coordenação e controle da operação da geração e da transmissão de energia elétrica integrantes do Sistema Interligado Nacional (SIN). Frise-se, contudo, que o poder regulatório desta instituição privada — cuja autoridade é expressamente reconhecida pela lei — sempre fora entendida pela doutrina como espécie de “correção”, não havendo menção expressa ao termo “autorregulação regulada”, que pareceu ganhar mais força no âmbito do direito digital. Cf. ARAGÃO, Alexandre Santos de. *Direito dos Serviços Públicos*. 4. ed. Belo Horizonte: Fórum, 2017.

3.3 Um novo modelo autorregulatório: uma avaliação do *Oversight Board*

O Conselho de Supervisão (*Oversight Board*) é estruturado com a finalidade de superar a percepção geral de que modelos puramente autorregulatórios têm se mostrado inaptos para endereçar adequadamente matérias de direitos fundamentais no âmbito da *internet*. Trata-se de uma ferramenta de revisão humana colegiada, supostamente independente, de decisões tomadas de forma automatizada ou pelo *staff* do Facebook ou Instagram, representando tentativa da empresa de manter a credibilidade do Facebook (Meta) no mercado junto aos seus acionistas e usuários diante dos inúmeros episódios polêmicos em que a empresa se viu recentemente envolvida. Embora não seja possível afirmar, por ora, se o arranjo preconizado representa de fato um avanço na regulação da liberdade de expressão no âmbito das redes sociais³³⁸ ou uma peça publicitária empresarial (*ethical washing*), é possível projetar, pelos seus atos constitutivos e decisões mais relevantes, as principais limitações e virtudes da proposta.

A legitimidade do conselho se assenta, basicamente, sob duas premissas: independência e transparência. De um lado, uma pressuposta pretensão de crescente autonomia do órgão em relação à Meta. Com efeito, ainda que as decisões de moderação tenham por base os padrões de comunidade do Facebook e do Instagram, espera-se que sejam tomadas por julgadores não submissos à esfera gerencial da empresa. Ademais, a moderação deve ser construída mediante um prévio procedimento de análise transparente e pautado em uma razão pública.

Kate Klonick evidencia que a independência requer necessariamente a aferição de três tipos de autonomia distintas: a autonomia financeira, jurisdicional e intelectual do conselho e de seus representantes.³³⁹ Assim, os esforços do Facebook pela independência dependem da análise específica de cada um destes elementos à luz da sua Carta de Criação (o estatuto do conselho) e do regulamento interno que disciplina seu funcionamento.

³³⁸ Kate Klonick chega a mencionar que o êxito do *Oversight Board* pode implicar o reconhecimento da sua autoridade pelo Estado e até por outras plataformas, o que a aproximaria do modelo de instituição privada de autorregulação regulada a que alude o Projeto de Lei nacional nº 2630/2020 (KLONICK, Kate. *The Facebook Oversight Board: Creating an Independent Institution to Adjudicate Online Free Expression*. *Yale Law Journal*, [s.l.], v. 129, n. 2418, jul. 2020, p. 2422-2499). Em sentido similar, há quem defenda a criação de uma estrutura administrativa análoga a uma Corte Constitucional como uma decorrência do constitucionalismo digital. Cf. MILOŠ, Matija; PELIĆ, Toni. *Constitutional Reasoning There and Back Again: The Facebook Oversight Board as a Source of Transnational Constitutional Advice*. In: DE POORTER, J. et al. (Eds). *European Yearbook of Constitutional Law 2021: Constitutional Advice*, vol. 3. [s.l.]: Springer & T.M.C. Asser Press, The Hague, 2022, p. 197-223

³³⁹ KLONICK, Kate. *The Facebook Oversight Board: Creating an Independent Institution to Adjudicate Online Free Expression*. *Yale Law Journal*, [s.l.], v. 129, n. 2418, jul. 2020.

A autonomia financeira do Conselho é assegurada por um *trust* cujo aporte inicial fora uma doação irrevogável de 130 milhões de dólares do Facebook.³⁴⁰ O objetivo deste *trust* é garantir a criação, funcionamento e supervisão das operações do Conselho de Supervisão. Atualmente, esta estrutura é gerenciada por um conjunto de seis administradores (*trustees*)³⁴¹, não tendo o Facebook, a princípio, ingerência direta sobre as atividades de gestão. A suposta independência do *trust* pode ser, contudo, limitada em circunstâncias excepcionais ou para decisão de questões-chave.³⁴² O acordo entre o *trust* e o conselho, infelizmente, não especifica quais são estas questões-chave.

A autonomia jurisdicional, por outro lado, corresponde à capacidade de o *board* fazer com que suas decisões sejam observadas pela Meta.³⁴³ O estatuto e os regulamentos do conselho são os instrumentos que fornecem os principais elementos para avaliação da extensão dessa independência.

O escopo do órgão é definido pelo estatuto. Segundo art. 2º, seu objetivo primordial abrange solicitações de análise da pessoa responsável pela publicação original do conteúdo ou por uma pessoa que tenha enviado anteriormente o conteúdo.³⁴⁴ A Meta poderá, também, enviar solicitações para análise e perguntas adicionais relacionadas ao tratamento do conteúdo, que vão além do questionamento se o conteúdo deve ser permitido ou removido completamente.

A despeito desta amplitude e do compromisso da Meta de respeitar as decisões do órgão, os regulamentos internos do comitê designam somente peças individuais para análise (*posts*, fotos, vídeos e comentários), que tenham violado seus termos de uso e padrões de comunidade (não abarca conteúdos que violem a lei do país onde opera), e o usuário deve ter uma conta ativa no Facebook ou Instagram. No mais, é permitido ao órgão — além de decidir pela exclusão ou manutenção de conteúdo — instruir, de forma *não vinculante*, o Facebook e Instagram à alteração nos seus padrões de comunidade e outras políticas relevantes. Pedidos de informações sobre algoritmos e *design* das redes, entretanto, só serão fornecidos se a Meta julgar razoavelmente necessário para as deliberações do *Oversight Board*.³⁴⁵

³⁴⁰ Este valor seria suficiente para financiar dois mandatos ou seis anos de funcionamento do comitê. Cf. *ibid.*

³⁴¹ Informação disponível em: GOVERNANCE. *Oversight Board*, [s.l.], [2018]. Disponível em: <https://oversightboard.com/governance/>. Acesso em 08 mar. 2022.

³⁴² Oversight Board Trust Agreement, Section 2.2. Disponível em: <https://about.fb.com/wp-content/uploads/2019/12/Trust-Agreement.pdf>.

³⁴³ KLONICK, Kate. The Facebook Oversight Board: Creating an Independent Institution to Adjudicate Online Free Expression. *Yale Law Journal*, [s.l.], v. 129, n. 2418, jul. 2020.

³⁴⁴ Artigo 2º, seção 1.

³⁴⁵ Artigo 1º, seção 4 do Estatuto do Conselho. O ponto é reforçado pelo Estatuto Interno do Conselho, que assim dispõe no item 2.2.2: “Facebook may decline such requests where Facebook determines that the information is not reasonably required for decision-making in accordance with the intent of the charter, is not technically feasible

Como muito dos termos do estatuto são bem amplos, a independência jurisdicional vai depender, em parte, da deferência do Meta na interpretação de seus dispositivos e no atendimento das recomendações de mudança das regras de suas plataformas. É preciso destacar, ainda, que alterações no regulamento interno e no Estatuto do próprio *Oversight Board* dependem de aprovação do Facebook (Meta), em conjunto com a maioria do Conselho e os *trustees*.³⁴⁶ Não há possibilidade, portanto, de ampliar o escopo da atuação do conselho por meio de alterações no estatuto e no regulamento sem a concordância do Facebook (Meta).

A independência intelectual, por sua vez, é garantida pela ausência de qualquer pressão externa dos membros do conselho, quando da tomada de decisões, bem como a capacidade do órgão decisório de evitar a influência de vieses, por meio de um procedimento de seleção e remoção idôneo de seus membros.³⁴⁷

Por necessidades pragmáticas, o processo inicial de seleção de membros foi feito pelo próprio Facebook, mas, segundo informações da própria empresa, o procedimento excluiu do processo seletivo todos aqueles que são ou já foram empregados da empresa ou, ainda, casos de potencial conflito de interesse³⁴⁸. Os próximos processos seletivos serão executados pelo próprio Conselho, ainda que assegurada a indicação de candidatos pela Meta e pelo público.³⁴⁹ A indicação formal do membro é feita pelos *Trustees*, cabendo a este ainda instruir os processos para assegurar o balanceamento geográfico no processo de seleção.³⁵⁰ O mandato de cada membro será de três anos, por um máximo de três mandatos.³⁵¹

O processo de exclusão é disciplinado em grande parte pelo regulamento interno e é aplicável a quem descumprir o Código de Conduta do Conselho³⁵². A exclusão poderá advir de indicação do Diretor do *Board*, do próprio Conselho de Supervisão ou do público. O ponto mais importante do estatuto é a definição de que os membros do Conselho jamais poderão ser afastados em razão de suas decisões.³⁵³

A avaliação da legitimidade do órgão requer ainda transparência sobre o procedimento a ser adotado para “apelar” de análise das remoções de conteúdo — aceitas ou negadas —

to provide, is covered by attorney/ client privilege, and/or cannot or should not be provided because of legal, privacy, safety, or data protection restrictions or concerns”.

³⁴⁶ Art. 6º, seção 1 do Estatuto do Conselho.

³⁴⁷ KLONICK, Kate. The Facebook Oversight Board: Creating an Independent Institution to Adjudicate Online Free Expression. *Yale Law Journal*, [s.l.], v. 129, n. 2418, jul. 2020.

³⁴⁸ *Ibid.*

³⁴⁹ O regulamento interno traz previsão neste sentido (art. 1º, seção 1.2.2).

³⁵⁰ Art. 4º, seção 2.1.1 do Regulamento Interno.

³⁵¹ Art. 1º, seção 3 do Estatuto do Comitê.

³⁵² É o documento que rege os princípios que devem pautar as condutas profissionais e pessoais dos membros do conselho.

³⁵³ Art. 1º, seção 8 do Estatuto do Conselho.

levadas a cabo pelo Facebook ou Instagram. É necessário que seja também transparente, de modo a possibilitar que as pessoas tenham confiança no procedimento desenvolvido. Neste sentido, o estatuto prevê linhas gerais do processo decisório, que serão tomados por painéis (com até 05 membros). Após a decisão do Painel, a decisão é submetida à análise pelo restante do Conselho, que, por maioria, poderá aprová-la. A decisão será publicizada no sítio eletrônico do *Oversight Board*, em linguagem acessível em dezoito idiomas distintos³⁵⁴. O estatuto se preocupa em construir decisões por meio de consensos, embora estes não sejam um imperativo. Na eventualidade de o consenso não ser alcançado, os motivos de divergência poderão ser incluídos na decisão a ser publicada, a fim de abranger todos os pontos de vista e vozes dissonantes.³⁵⁵ Como ponto negativo, pode-se citar que parte substancial do procedimento de apelação ao comitê é disciplinada em regulamento interno, ainda não disponível em português³⁵⁶ — o que, de certo, limita o acesso do usuário ao órgão. Também não há previsão de identificação dos membros dos painéis (com até 5 membros) responsáveis pela decisão, visto que a opção do estatuto foi mantê-los anônimos.³⁵⁷ A forma de seleção dos recursos para julgamento — o estatuto fala no parâmetro da relevância do caso — não são também evidenciados de forma pública, embora haja a promessa de que serão elaborados relatórios anuais com a indicação dos critérios utilizados para publicizar o que é “relevante” para o comitê, como forma de prestação de contas.³⁵⁸

Em resumo, tanto o estatuto quanto o regulamento possuem significativas limitações. A independência financeira não é plena e está sujeita a uma cláusula geral que não evidencia as “questões-chave” que podem justificar a interferência da Meta sobre o conselho. E mesmo a independência jurisdicional depende da deferência da Meta com interpretações mais ampliadas do escopo de atuação do Conselho de Supervisão e da boa-vontade da empresa regulada para fornecer resposta quanto às solicitações de caráter técnico e propostas de alteração dos termos de uso e demais normas comunitárias. No mais, não há assunção de compromisso de observância da legislação, visto que só estarão sujeitos à análise do *Oversight*, manifestações potencialmente ofensivas aos termos de uso e padrões da comunidade do Facebook ou Instagram.

³⁵⁴ Art. 1º, seção 4 do Regulamento Interno.

³⁵⁵ Art. 3º, seção 4 do Estatuto do Conselho.

³⁵⁶ Em 2021, foi incluído no site um tutorial em português sobre o procedimento de apelação, o que mitiga este impacto negativo. Link: <https://oversightboard.com/appeals-process/>.

³⁵⁷ Art. 3º, seção 2 do Estatuto do Comitê. A opção pelo anonimato se ampara justamente numa necessidade de blindar os membros de pressões externas, garantindo isenção.

³⁵⁸ A observação é de Kate Klonick (The Facebook Oversight Board: Creating an Independent Institution to Adjudicate Online Free Expression. *Yale Law Journal*, [s.l.], v. 129, n. 2418, jul. 2020).

No campo da independência intelectual, a tendência é sua consolidação à medida em que o órgão avance em seu funcionamento e a influência inicial do Facebook na formação dos primeiros membros do comitê vá gradualmente se reduzindo. Há ainda sérias dúvidas se estes julgamentos serão, contudo, suficientes para orientar os milhões de pedidos de remoção que se acumulam no Facebook/Instagram todos os dias, o que pode implicar substancial redução da capacidade dos julgamentos do comitê influenciarem de forma ampla a prática de moderação da empresa.

A transparência procedimental parece, contudo, ser um ponto alto, ressalvada a recomendação de publicização dos membros do painel julgador e disponibilização dos regulamentos definidores do procedimento em outras línguas que não o inglês.

Um estudo da atuação do órgão pressupõe, no entanto, aferir como o *Oversight Board* opera na prática. Considerando o pequeno número de casos apreciado pelo Comitê no presente caso (23, ao todo), opta-se aqui pela avaliação daquele que é o mais relevante dos casos apresentados pelo órgão em sua curta vida.

No caso nº 2021-001-FB-UA, apreciou-se a proporcionalidade das sanções aplicadas pelo Facebook e Instagram de supressão do perfil do ex-Presidente Donald Trump de suas plataformas, por força de sucessivas declarações realizadas quanto à fraude nas eleições e, especialmente, duas declarações do dia 06 de janeiro que exaltavam as pessoas envolvidas no episódio de invasão ao Capitólio,³⁵⁹ o que violaria as regras dos padrões de comunidades do Facebook, que proíbem o uso da plataforma para apoiar pessoas envolvidas em violência. O conselho manifestou-se pela proporcionalidade da sanção de suspensão dos perfis, porém considerou inapropriado mantê-la por tempo indefinido. O conselho expediu, ainda, a pedido do Facebook, recomendações a serem adotadas quando postagens de usuários influentes representarem uma grande probabilidade de dano iminente.

Mais do que uma análise acurada do acerto ou desacerto substancial da decisão, cumpre aqui avaliar se o procedimento seguiu de maneira transparente como preconizado no estatuto e no regulamento.

É positivo que a decisão do comitê descreva de forma tão detalhada não só as razões que levaram à tomada de decisão, mas também todo o histórico que permeou o processo. Há relato do contexto fático subjacente à análise, qual seja, o processo eleitoral norte-americano

³⁵⁹ Foram consideradas duas declarações do dia 06 de janeiro e, em especial, os usos das expressões “Amo Vocês. Vocês são especiais”, na primeira publicação e “grandes patriotas” e “lembre-se desse dia para sempre”, na segunda publicação. Fonte: OVERSIGHT BOARD. Former President Trump’s suspension. *Oversight Board*, [s.l.], 7 jan. 2021. Disponível em: <https://oversightboard.com/decision/FB-691QAMHJ/>. Acesso em: 06 jun. 2022.

permeado por diversas alegações não comprovadas de fraude do ex-Presidente Donald Trump; indicação dos parâmetros normativos que orientaram a tomada de decisão (os padrões de comunidades e termos de serviços do Facebook e Instagram e normas internacionais de direitos humanos incorporadas pela política corporativa de direitos humanos do Facebook); o conteúdo da declaração de defesa apresentada pelo ex-Presidente Donald Trump; explicações do Facebook sobre a decisão objeto de apelação; comentários públicos encaminhados por partes interessadas e indicações de eventuais dissensos que surgiram durante a votação.

Ao fim, estes elementos foram considerados para certificar a necessidade de se adequar as sanções a um prazo estabelecido pelo Facebook, não podendo a empresa se esquivar de suas responsabilidades ao pedir ao conselho a fixação de um prazo proporcional para sanção.

A clareza na exposição dos elementos e transparência argumentativa na enunciação das razões de decisão são deveres que, uma vez cumpridos, conferem uma disponibilidade daqueles que se submetem às suas decisões a acolher suas perspectivas, dando-lhe substancial legitimidade procedimental.

Há, no entanto, um ponto negativo substancial na decisão, que indica um limite de difícil superação pelo comitê. O Facebook se recusou a responder às principais perguntas relacionadas ao funcionamento dos algoritmos que orientam a arquitetura de escolhas em suas plataformas. Em síntese, recusou-se a responder como seus recursos afetaram a visibilidade do conteúdo do ex-Presidente Donald Trump antes da aplicação da sanção de remoção do perfil; se pretende revisar suas políticas de *design* a fim de evitar outros episódios violentos como o do dia 06 de janeiro de 2021 (especialmente contribuindo para a redução de discursos de ódios na rede) e, ainda, se a suspensão e exclusão da conta afeta a capacidade de os anunciantes fazerem direcionamentos às contas dos seguidores.

Vale lembrar que o Estatuto do Conselho não exige que o Facebook responda todas as perguntas feitas, mas somente aquelas que julgue indispensável para as deliberações do comitê. Em outras palavras, o próprio Facebook irá decidir o que irá expor ao público. Frise-se que a recusa não se deu unicamente no caso envolvendo o ex-presidente. Em ao menos outros três precedentes, o Facebook se recusou expressamente a prestar informações técnicas exigidas pelo conselho tanto acerca dos critérios que orientam tanto a curadoria quanto a moderação de conteúdos, inclusive indicando se a supressão de conteúdo se deu por pressões de órgãos estatais. O fundamento é sempre uma referência genérica ao artigo 2º, seção 2.2.2 do Estatuto do Conselho de Supervisão, que permite à empresa Meta recusar o fornecimento de informações

técnicas que não julgue razoável para a análise dos casos, devido a preocupações legais relacionadas à privacidade, segurança e proteção de dados.³⁶⁰

Outro ponto frágil é que o número reduzido de julgamentos do comitê ao longo destes dois anos é substancialmente incapaz de dar conta dos milhões de casos analisados pelas plataformas do Facebook e Instagram. Como já se expôs, a Meta somente é obrigada a atender os pedidos de manutenção ou exclusão de uma determina peça de conteúdo ou perfil.

A forma mais fácil de conferir alcance geral a estes precedentes seria vincular as recomendações de reformulação dos padrões de comunidade da empresa ou, ainda, de aprimoramento das técnicas humanas e automatizadas de moderação. Em 21 dos 23 casos submetidos ao *Oversight Board*³⁶¹, há recomendações que caminham no sentido de reformulação das diretrizes adotadas pelos padrões de comunidades das plataformas³⁶² ou,

³⁶⁰ Além do caso Trump, há outros três precedentes em que informações foram solicitadas e a empresa Meta se recusou em fornecê-las: 1) caso 2021-009-FB-UA: o Conselho de Supervisão concordou com uma decisão do Facebook que reverteu sua decisão original de remover conteúdo postado por influente usuário egípcio, que compartilhava uma postagem de notícias — da Al Jazeera — sobre uma ameaça de violência da ala militar do grupo palestino Hamas. A manutenção da derrubada se deu mesmo após revisão humana. Depois que o caso foi selecionado pelo conselho, o Facebook prontamente anulou sua decisão anterior e retomou com o conteúdo. Durante o julgamento, o conselho solicitou informações quanto às alegações de que o conteúdo palestino teria sido “censurado” por pressão do Governo Israelense. Embora o Facebook tenha alegado genericamente não haver pressão, recusou-se a entrar em maiores detalhes sobre eventual requerimento do Governo de Israel; 2) decisão do caso 2021-006-IG-UA: o Conselho de Supervisão anulou decisão original da Meta de remover uma postagem do Instagram incentivando as pessoas a discutirem o confinamento solitário de Abdullah Ocallan, membro fundador do Partido dos Trabalhadores do Curdistão (PKK). Durante o julgamento se percebeu que a supressão do conteúdo se deu por conta de uma falha de orientação interna sobre a política de “Indivíduos e Organizações Perigosas”, que foi “inadvertidamente não transferida” para um novo sistema de revisão em 2018. O Conselho indagou a Meta acerca de quantos conteúdos foram removidos indevidamente em razão desta falha, no que foi respondido que a resposta de tal requerimento era “tecnicamente inviável”; 3) decisão do caso nº 2021-003-FB-UA: o conselho anulou decisão de remover uma postagem e limitar o acesso ao perfil de usuário indiano, que alegava que o primeiro ministro indiano estaria formulando uma ameaça de “genocídio dos sikhs” (minorias religiosas daquele país), sob conselho do Presidente da RSS (*Rashtriya Swayamsevak Sangh*) uma organização ultranacionalista indiana. A Meta se recusou a responder se houve pedido de autoridades indianas para derrubada ou contingenciamento na distribuição do conteúdo, por não as considerar razoavelmente necessária para julgamento do caso.

³⁶¹ Os dados são de junho de 2022.

³⁶² Cf. decisão nº 2020-006-FB-FBR — trata-se de caso de postagem de usuário francês que compartilhava notícia que recomendava o uso de hidroxicloroquina e azitromicina para tratamento da COVID, tecendo duras críticas à agência de saúde francesa, por proibir a prescrição de tais medicamentos. O conteúdo foi derrubado sob o rótulo de desinformação produtora de “danos iminentes”. O conselho determinou a retomada do conteúdo, mas teceu diversas considerações sobre a necessidade de reformulação dos padrões de comunidade da empresa, especialmente quanto à definição mais concreta do que seriam “danos iminentes”, bem como de aprimoramento das técnicas de moderação. A título de destaque se sugeriu: a) criação de medidas menos intrusivas para políticas de desinformação em saúde, tais como etiquetagem, citação de fontes da OMS, criação de medidas de fricção (redução de compartilhamentos e rebaixamento de visualizações no *feed* de notícias, por exemplo); b) aprimoramento dos relatórios de transparência sobre as políticas de desinformação em saúde. Recomendações similares foram feitas no precedente (2021-008-FB-UA), em que página de associação estadual de médicos teceu críticas a medidas de isolamento e segregação compulsória impostas por Governos Estaduais e Municipais no Brasil, com base em uma notícia que descontextualizava manifestação do porta-voz da Organização Mundial de Saúde (OMS) sobre o tema. Vale destacar que, em todos os vinte e um precedentes com recomendações, há menção também da necessidade de informar ao usuário sobre as razões que embasam a supressão do conteúdo, conferindo-lhe prazo para pedidos de revisão da decisão.

ainda, que recomendem maior transparência aos usuários sobre os critérios que pautam a aplicação das regras de moderação³⁶³. No entanto, como já exposto, tais alertas são meramente instrutivos. Não há qualquer mecanismo que confira caráter vinculatório a tais determinações, nem instrumento de prestação de contas da empresa ao *Oversight* quanto ao atendimento destas determinações.

Estes são os pontos que traduzem a maior vulnerabilidade do *Oversight Board*. Primeiro, há uma autonomia bastante limitada do conselho, especialmente no aspecto jurisdicional, pois suas decisões só vinculam quanto ao caso concreto, tendo mero efeitos de recomendação na adoção de políticas que busquem prevenir a má aplicação das regras de moderação ou instar a empresa a adotar políticas informativas para reduzir a disseminação de discursos danosos, o que torna o modelo tal como apresentado incapaz de construir uma política geral de efetivo respeito aos direitos fundamentais nestas plataformas. Segundo, e mais importante, a Meta é pouco transparente, visto que eventualmente se recusa a fornecer, de forma completa, informações quanto à sua política de construção algorítmica.

No limite, tal modelo — desde que aperfeiçoado³⁶⁴ — pode ser interessante para fins de redução dos riscos de “censura privada” e até ser evocado como modelo para uma instituição de “autorregulação regulada”, mas será incapaz de lidar com os danos colaterais provocados pelos algoritmos.

Como exposto, os poderes restritos atribuídos ao conselho permitem à empresa Meta ocultar toda e qualquer informação sobre os fins que orientam sua programação algorítmica, sempre que isso lhe puder gerar algum prejuízo reputacional ou econômico.

³⁶³ Por exemplo, caso 2020-004-IG-BR — decisão do Conselho de Supervisão que anulou decisão da Meta de supressão do Instagram de conteúdo com nudez, que fora publicizado com o intuito de promover a campanha do “outubro rosa”, no Brasil. Além da anulação da decisão, foi recomendada a necessidade de se incluir uma exceção a ser apreendida pelos mecanismos de moderação automatizada que vedam a promoção de conteúdo com nudez, para as hipóteses de “promoção de campanha de conscientização contra o câncer” e de que revise as Diretrizes de Comunidade do Instagram a fim de esclarecer que a proibição de nudez não é absoluta. Essa recomendação foi de fato seguida pela rede social, que hoje traz exceções à vedação à nudez em suas diretrizes da comunidade (veja-se: <https://pt-br.facebook.com/help/instagram/477434105621119>. Acesso em: 07 jun. 2022).

³⁶⁴ Para isso, é fundamental que a Meta atenda, de fato, as recomendações do *Oversight* a despeito de seu caráter não coercitivo e seja mais transparente na sua relação com Estados, até mesmo para evitar pressões pouco republicanas, que ocultem instrumentos de censura colateral. A respeito, confira-se: ARUN, Chinmaayi. Facebook’s Face. *Harvard Law Review*, [s.l.], v. 135, abr. 2021. Mimeografado.

3.4 O exemplo corregulatório alemão: a *NetzDg*³⁶⁵

A NetzDG, lei alemã para aplicação da lei na *internet*, é, até hoje, principal exemplo de construção de um regime de responsabilidade corregulatório no âmbito das redes sociais. A norma apresenta um escopo principal: assegurar a aplicação da legislação alemã ao ambiente *online*, a fim de evitar que as redes sociais se tornem um livre espaço para prática de atos ilícitos.³⁶⁶

Para tanto, a lei define redes sociais como “provedores de telecomunicação que, com cunho lucrativo, operam plataformas na *internet* que funcionam de modo que usuários possam compartilhar conteúdos indeterminados com outros usuários ou torná-los acessíveis ao público”³⁶⁷, dispensando do cumprimento das obrigações, a rede social que tiver, domesticamente, menos de dois milhões de usuários registrados.

Consideram-se conteúdos ilícitos os indicados no § 1º (3) da Lei, que faz referência a uma série de tipos do Código Penal Alemão, tais como incitação ao ódio (*hatred*), calúnia e difamação.

A consecução do objetivo da norma decorre do reconhecimento de dois deveres centrais à plataforma. Um primeiro relacionado à obrigação de criação de um sistema de um “procedimento contínuo, efetivo e transparente (...) para tratamento de reclamações de conteúdos ilícitos” (§ 4). A eficiência deste procedimento requer que o provedor da rede social

³⁶⁵ A análise se baseia em versão traduzida da lei disponibilizada na coletânea “Fake News e Regulação”. (ABBOUD, Georges; CAMPOS, Ricardo; NERY JR., Nelson (Orgs.). *Fake News e Regulação*. 2. ed. São Paulo: Revista dos Tribunais, 2020).

³⁶⁶ A elaboração da lei teve como mote um ambiente de hostilidade que se instalou nas redes sociais, principalmente em relação à política de imigração alemã, que não raro descambava para divulgação de atos difamatórios e produção de discursos de ódio em face de imigrantes islâmicos. O fato considerado símbolo do movimento foi a viralização de uma foto de um refugiado sírio (Modanami) junto à chanceler Ângela Merkel, no final de 2015, que normalmente vinha acompanhada da divulgação de histórias de que o refugiado teria participado de atentados ocorridos na Europa. As falsas notícias buscavam vincular o governo alemão ao terrorismo. Quando Modanami instou o Facebook à retirada dos conteúdos difamatórios, a rede social concordou em retirar e usar ferramentas de *geoblocking* em relação a algumas postagens difamatórias. Houve uma recusa, contudo, em monitorar ativamente conteúdos novos que viessem a ser postados, de modo a viabilizar sua rápida retirada. Modanami chegou a acionar o judiciário requerendo que o Facebook fosse responsabilizado por novas publicações, pedido que lhe foi negado. Informação disponível em: UN REFUGIADO sírio pierde su batalla judicial contra Facebook por un ‘selfie’ con Merkel. *20 minutos*, [s.l.], 7 mar. 2017. Disponível em: <https://www.20minutos.es/noticia/2978420/0/refugiado-sirio-batalla-facebook-alemania/foto-montaje-terrorista/>. Acesso em: 15 jun. 2022.

³⁶⁷ A lei expressamente afasta do seu escopo “plataformas com ofertas de cunho jornalístico e redacional” e “plataformas direcionadas para comunicação individual ou difusão de conteúdos específicos”.

crie mecanismos aos usuários que permitam a reclamação de conteúdo ilícito e cujo bloqueio deverá ocorrer em até 24 (vinte e quatro) horas, quando a ilicitude for evidente. Este prazo poderá ser de sete dias nos demais casos, podendo ultrapassá-lo, quando a decisão da ilicitude depender da “inverdade de uma alegação de fato ou depender evidentemente de outras circunstâncias de natureza fatural”, hipótese na qual será assegurada ao usuário a oportunidade de defesa antes da tomada de decisão ou, ainda, quando a rede social “remeta a decisão da ilicitude para uma instituição de autorregulação regulada”.

Este procedimento deve informar “sem atraso” o reclamante e o usuário sobre toda decisão e os fundamentos que a justificam. O procedimento será supervisionado pela diretoria da rede social por controle mensal e poderá ser supervisionada, ainda, por uma agência especificamente autorizada, que a norma denomina de “instituição de autorregulação regulada”. Esta instituição deverá cumprir, minimamente, os seguintes requisitos: a) garantia de independência e conhecimento técnico de seus analistas; b) uma estrutura de equipamentos que sejam tecnicamente apropriados; c) um regramento procedimental que regule a abrangência e o procedimento de análise, bem como as obrigações de submissão da rede social associada, além de prever a possibilidade de revisão das decisões; d) uma ouvidoria para reclamações; e) seja composta por diversos provedores de redes sociais ou instituições que assegurem uma estrutura e equipamentos tecnicamente adequados, além de ter a abertura assegurada para ingresso de outros provedores. Caberá às autoridades administrativas indicadas na lei reconhecerem uma instituição como de autorregulação regulada à luz dos parâmetros mínimos acima elencados.

A aferição da construção deste procedimento interno de governança de conteúdos potencialmente ilícitos depende do cumprimento de um outro dever: o dever de prestar informações às autoridades alemãs. Esta obrigação consta do § 2 da lei e prevê que redes sociais que tiverem, no ano civil, mais de 100 reclamações que tratem de conteúdos ilícitos são obrigadas a apresentar um relatório escrito em alemão sobre os procedimentos e medidas tomados em relação às reclamações relacionadas a conteúdos ilícitos nas plataformas. Este relatório será publicado no Diário Oficial da República Federal da Alemanha e no próprio *website* da rede social. O prazo para apresentação destes relatórios é semestral. O relatório deve conter, no mínimo, as seguintes informações: a) considerações gerais sobre os esforços que o provedor da rede social tomou para impedir ações ilícitas; b) apresentação de mecanismos de envio e submissões sobre conteúdos ilícitos, além de critérios utilizados para apagamento, retirada ou bloqueio de conteúdos ilícitos; c) números de reclamações relativas a conteúdos ilícitos recebidos no período abarcado pelo relatório, devendo categorizar as reclamações

provenientes ou não de usuários; d) organização, recurso pessoal, competência técnica e linguística das unidades de trabalho responsáveis para o processamento de reclamações, além do treinamento e supervisão do pessoal competente para o processamento das reclamações; e) número de reclamações nas quais uma agência ou entidade externa foi consultada para se chegar a uma decisão; f) número de reclamações recebidas no período abarcado pelo relatório, que levaram ao apagamento, retirada ou bloqueio de conteúdos contestados, devidamente categorizados por reclamações provenientes ou não de usuários; g) o tempo decorrido entre a entrada da reclamação da rede social e a retirada, apagamento ou bloqueio de conteúdo ilícito, devendo ser categorizado pelos períodos “dentro de 24 horas”/ “dentro de 48 horas”/ “dentro de uma semana”/ “mais que uma semana”; h) medidas tomadas para orientar e instruir tanto o usuário afetado quanto o usuário reclamante sobre o conteúdo da decisão oriunda da reclamação.

A norma traz ao final sanções administrativas (multas, em geral), no caso de atuação intencional ou negligente das redes sociais, dentre as quais cumpre destacar: a) falhar em produzir um relatório ou fazê-lo de forma incompleta, incorreta ou intempestiva ou ainda falhar na sua publicação; b) deixar de prover ou garantir, de maneira incorreta ou incompleta, o procedimento para tratamento de conteúdos ilícitos; c) deixar de nomear um mandatário autorizado a receber citações na República Federativa da Alemanha; d) deixar de prestar as informações requeridas, dentre outras. Há uma preocupação de as autoridades administrativas competentes emitirem “princípios administrativos” sobre o exercício do poder de aplicação de multas pecuniárias, a fim de evitar abuso no exercício do poder sancionatório. É ainda prevista uma hipótese de revisão judicial no caso de o governo impor multas por falhas sistemáticas na remoção ou bloqueio de conteúdo ilegal.

Duas foram as principais críticas feitas à lei. A primeira, ao uso da expressão “manifestamente ilícito”, que justificaria a derrubada do conteúdo ilícito em 24 horas e incentivaria um *chilling effect* (*overblocking*) na sua aplicação. A outra, de que a norma estaria privatizando o exercício do poder de polícia estatal ao atribuir, a agentes privados, o dever de criação de um procedimento célere e eficiente para derrubada de conteúdos ilegais (e não propriamente àqueles que violam seus termos de uso — ponto que não é abordado pela lei).³⁶⁸

³⁶⁸ WISCHMEYER, Thomas. “What is illegal offline is also illegal online”: The German Network Enforcement Act 2017. *Bilyana Petkova & Tuomas Ojanen: Fundamental Rights Protection Online: The Future Regulation of Intermediaries*, [s.l.], out. 2018. Mimeografado.

Quanto ao risco de *overblocking*, a doutrina alemã compreende, de forma amplamente majoritária,³⁶⁹ que o intuito da norma é punir tão somente aquelas redes sociais que falham, sistematicamente, no seu dever de construir um procedimento eficiente e justo para tratamento de reclamações sobre conteúdos ilícitos. Assim, a amplitude da expressão “manifestamente ilícita”, que poderia dar ensejo a uma onda de “falsos positivos” pela sua generalidade, é contrabalanceada pela exigência de que a falha seja evidenciada de forma reiterada, de modo que se constate — acima de qualquer dúvida — a ineficiência do sistema de *compliance* interno da rede social, mesmo em casos mais claros de violação à legislação daquele país.

Tão logo a lei entrou em vigor, houve apagamento de alguns conteúdos lícitos de grande projeção (principalmente satíricos) pelas redes sociais.³⁷⁰ A médio prazo, contudo, tais práticas cessaram, não havendo evidências sólidas de *overblocking* mesmo após quatro anos após vigência da lei³⁷¹. Isso se reforça pelo fato de que a ameaça de sanção administrativa nunca se concretizou quanto a este ponto. Nenhuma plataforma de rede social foi, até o momento, multada por excessos ou negligência no apagamento de conteúdo supostamente ilícito.³⁷²

Sobre o ponto de privatização de uma atribuição que competiria aos poderes públicos, por meio da transferência às plataformas do apagamento de conteúdo ilícito, a necessidade de se recorrer a tais agentes decorre, primeiro, de uma constatação fática. A atuação prioritária dos agentes que operam tais plataformas é uma consequência jurídica normal resultante de uma rede de relações oriundas de plataformas que têm caráter de direito privado, e da incapacidade de o Poder Público acompanhar diretamente tais relações³⁷³. Ademais, esse dever prioritário

³⁶⁹ EIFERT, Martin. Lei Alemã para a Melhoria da Aplicação nas Redes Sociais (Lei de aplicação da Lei nas redes – NetzDG). Trad. Pedro Henrique Ribeiro e Revisão de Júlio Carvalho. In: ABOUD, Georges; CAMPOS, Ricardo; NERY JR., Nelson (Orgs.). *Fake News e Regulação*. 2. ed. São Paulo: Revista dos Tribunais, 2020, p. 161-192.

³⁷⁰ EIFERT, Martin. Lei Alemã para a Melhoria da Aplicação nas Redes Sociais (Lei de aplicação da Lei nas redes – NetzDG). Trad. Pedro Henrique Ribeiro e Revisão de Júlio Carvalho. In: ABOUD, Georges; CAMPOS, Ricardo; NERY JR., Nelson (Orgs.). *Fake News e Regulação*. 2. ed. São Paulo: Revista dos Tribunais, 2020, p. 161-192. Na visão deste autor, estes “equivocos” foram frutos de articulação das próprias redes sociais, que buscavam alardear os efeitos censores da norma jogando a opinião pública contra a lei.

³⁷¹ Segundo relatórios de 2018, 27% das reclamações ao Youtube resultaram em bloqueio dos conteúdos, com fundamento na lei (pouco mais de 58.000 reclamações em um universo de quase 215.000 reclamações). No Twitter, essa taxa foi de 11% e, no Facebook, de 21%. Tais dados são substancialmente inferiores ao número de vídeos e conteúdos removidos por estas plataformas com base em suas normas comunitárias. A título de comparação, somente entre janeiro e março de 2018, o Youtube removeu mais de 9 milhões de vídeos de suas plataformas. Cf. WISCHMEYER, Thomas. “What is illegal offline is also illegal online”: The German Network Enforcement Act 2017. *Bilyana Petkova & Tuomas Ojanen*: Fundamental Rights Protection Online: The Future Regulation of Intermediaries, [s.l.], out. 2018. Mimeografado.

³⁷² SCHULZ, Wolfgang. Roles and Responsibilities of Information Intermediaries: fighting misinformation as a test case for a human rights-respecting governance of social media platforms. *A Hoover Institute Essay*, [s.l.], n. 1904, 2019.

³⁷³ Neste sentido, reforça-se mais uma vez as lições de Lessig sobre a essencialidade da ação legislativa indireta para regulação de espaços digitais.

seria normativamente justificável. Se tais plataformas contribuem para a facilitação da difusão e amplificação do alcance de conteúdos não só socialmente produtivos, mas também danosos, é juridicamente lógico que assumam a obrigação pelo apagamento daquele conteúdo que viole a lei. Frise-se, por fim, que o argumento da “privatização” só faria sentido se o usuário prejudicado pela ação destas empresas fosse impedido de recorrer à justiça em casos de abusos, o que não é vedado pela legislação alemã.³⁷⁴

Relativizada as críticas iniciais, o principal ponto de êxito da legislação, na visão da doutrina alemã, tem sido no aprofundamento do debate público sobre o tema, ampliando uma discussão política continuada da *internet* sobre suas práticas de decisão. Isso se reflete em duas circunstâncias. Primeiro, uma avaliação inicial de que a lei tornou as plataformas mais diligentes na aplicação de seus próprios termos de uso, por força do dever de transparência quanto à forma como lida com as reclamações percebidas.³⁷⁵ Segundo, pela mudança de impressão que se teve da lei na Europa. Em um primeiro momento, a norma foi vista como uma ameaça de violação a direitos humanos.³⁷⁶ Hoje, é fruto de apoio popular dentro de seu país e acabou sendo um precursor de arranjos corregulatórios idealizados pela própria União Europeia.³⁷⁷

A legislação alemã caminha, no entanto, timidamente na redução das assimetrias informacionais, tendo até o momento contribuído pouco para a resolução do mistério sobre o funcionamento dos algoritmos das redes sociais.

Explica-se. Como o foco da lei é o cumprimento da lei penal alemã — e não a aplicação dos termos de uso das próprias redes — há uma subnotificação no fornecimento de dados oferecidos ao governo daquele país.

Boa parte dos termos das principais plataformas já proíbe comportamentos que possam configurar crime de racismo, apologia ao nazismo, difamação e outras espécies de discursos odiosos vedados pela legislação alemã, havendo uma sobreposição entre os termos de uso e a norma nacional. Como a arquitetura das plataformas tende a priorizar reclamações fundadas em seus próprios termos de uso, sob o argumento de que o custo para apagamento de conteúdo

³⁷⁴ Cf. HELDT, Amélie. Reading Between the lines and the numbers: an analysis of the first NetzDG reports. Internet Policy Review. *Journal on Internet Regulation*, [s.l.], v. 8, n. 2, jun. 2019.

³⁷⁵ LEERSEN, Paddy; TWOREK, Heidi. *An analysis of Germany's NetzDG law*. Amsterdam: Transatlantic High Level Working Group on Content Moderation Online and Freedom of Expression, 2019.

³⁷⁶ UNITED NATIONS. *Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*. [S.l.]: UN, 2017. Disponível em: <https://digitallibrary.un.org/record/1304394>. Acesso em: 15 jun. 2022.

³⁷⁷ É o *Digital Services Act* (DAS), norma recém promulgada, que estabelece o dever de intermediários digitais — dentre os quais se inserem as plataformas de redes sociais — de criarem mecanismos de denúncia amigáveis aos usuários para notificar conteúdo ilegal sob seus serviços, assegurado o devido processo legal. A lei ainda engloba preocupação com a divulgação de publicidade nas redes, ponto que não é abarcado pelo *NetzDG*.

nestes casos, por se dar de forma universal (sem necessidade de uso de filtros *geoblocking*, por exemplo) seria menor, a amostra quantitativa de reclamações com base no *NetzDG* tende a ser reduzida, fornecendo ao Poder Público poucas evidências sobre o funcionamento da estrutura de curadoria e moderação destas redes³⁷⁸.

Ao limitar o dever de garantir de um tratamento adequado de conteúdos à legislação penal alemã, o *NetzDG* reduz significativamente o escrutínio público sobre procedimentos e arquiteturas que orientam a governança privada das plataformas.

O principal exemplo desta limitação é o Facebook, que tem se mostrado bastante resistente ao cumprimento da lei. Para efeitos de comparação, o Facebook relatou, no ano de 2018, 886 reclamações com fundamento no *NetzDG*, contra 214.827 do Youtube e 264.818 do Twitter. No segundo semestre daquele ano, números similares se repetiram. Ao se comparar a arquitetura das três redes sociais, fica claro porque o Facebook (atual Meta), apesar de ter uma rede de usuários registrados na Alemanha significativamente maior, apresenta um número substancialmente inferior de reclamações. Enquanto Youtube e Twitter criaram ferramentas de denúncia que incluem a violação ao *NetzDG* como uma categoria nova ao lado de tantas outras que estas redes sociais apresentavam, o Facebook mantém tal ferramenta de denúncia em um ponto apartado de sua página no qual o usuário precisa acessar um link que o direciona a outro site onde a denúncia pode ser feita. A criação destes elementos de fricção passa a impressão de operar tal como um *sludge*³⁷⁹, que busca atravancar o bom funcionamento deste novo canal de denúncias. Em resposta, o Governo Alemão aplicou uma multa de 2 (dois) milhões de euros ao Facebook, sob o argumento de que os dados apresentados em relatórios estavam subnotificados.³⁸⁰ A má vontade das plataformas com a lei também se constata pelo fato de que até hoje nunca foi criada qualquer entidade de autorregulação regulada.

³⁷⁸ Talvez o ponto mais interessante dos relatórios tenha sido as informações consolidadas de que mecanismos automatizados de moderação, de fato, não são hoje capazes de fazerem uma sólida diferenciação entre discursos de ódio e discursos humorísticos/satíricos, dada sua dificuldade de compreender contextos. O ponto tem sido relatado, em especial, por relatórios prestados pelo Google, o que tem levado a pedidos pela ampliação do número de moderadores humanos por parte destas empresas. Cf. HELDT, Amélie. Reading Between the lines and the numbers: an analysis of the first NetzDG reports. *Internet Policy Review. Journal on Internet Regulation*, [s.l.], v. 8, n. 2, jun. 2019.

³⁷⁹ *Sludges* são arquiteturas que criam fricções (entraves burocráticos, em geral) capazes de operar de duas formas. Desencorajando a pessoa de atuar em prol de seu próprio interesse ou encorajando-a comportamentos autodestrutivos. Cf. THALER, Richard. Nudge, not sludge. *Science*, [s.l.], v. 361, n. 6401, ago. 2018. Disponível em: <https://science.sciencemag.org/content/361/6401/431>. Acesso em: 14 jun. 2022.

³⁸⁰ Informação disponível em: ALEMANHA multa Facebook em 2 milhões de euros por violação à lei de transparência online. *GI*, [s.l.], 2 jul. 2019. Disponível em: <https://g1.globo.com/economia/tecnologia/noticia/2019/07/02/alemanha-multa-facebook-em-2-milhoes-de-euros-por-violacao-a-lei-de-transparencia-online.ghtml>. Acesso em: 14 jun. 2022.

Apesar da multa ao Facebook, os dados prestados ao Governo alemão ainda parecem aquém do objetivo pretendido de redução das assimetrias informacionais, trazendo poucos esclarecimentos sobre a “caixa preta” das redes sociais³⁸¹.

3.5 Fundamentos constitucionais para a priorização da correção das redes sociais

A comparação entre os dois modelos nos permite aferir os pontos positivos e negativos de cada proposta. O modelo autorregulatório, que aposta em um Conselho de Supervisão privado, mostra-se capaz de seguir procedimentos mais ágeis, que permitam a construção de decisões bem fundamentadas acerca do limite à liberdade de expressão. De outro, a ausência de um elemento de fiscalização efetivamente independente parece prejudicar a capacidade do *Oversight Board*, de fato, influenciar a política de governança das plataformas.

O modelo correção alemão³⁸², por sua vez, vem se mostrando positivo para o debate público, chamando maior atenção à opacidade da estrutura algorítmica destas redes e conferindo maior transparência às plataformas. Efetivamente, contudo, o modelo avançou pouco na redução das assimetrias informacionais em razão da opção legislativa de vincular a observância dos deveres de monitoramento e informação apenas às reclamações que tenham como fundamento a violação da lei penal alemã. Um modelo correção que ignore a centralidade dos termos de uso e normas comunitárias, que pauta a maior parte das ações das plataformas, tende a ser parcialmente ineficaz.

³⁸¹ É a crítica de Amélie Heldt: “*Rather than providing numbers without context, the reports are supposed to help understand the connection between the grounds on which social networks delete or block unlawful content and the provisions provided by law. Unfortunately, this expectation was not fulfilled*” (HELDT, Amélie. Reading Between the lines and the numbers: an analysis of the first NetzDG reports. Internet Policy Review. *Journal on Internet Regulation*, [s.l.], v. 8, n. 2, jun. 2019).

³⁸² Apesar da lei alemã exigir, em determinado momento, o estabelecimento de um procedimento de comunicação e envio de reclamação sobre conteúdos ilícitos de fácil reconhecimento, exigindo prazos específicos para atuação (o que a aproxima de uma regulação tradicional de comando e controle, neste ponto), renuncia, no geral, a uma configuração pormenorizada sobre a implementação de um sistema de devido processo legal para denúncia e retirada de conteúdos assegurando alto grau de liberdade para as empresas em sua implementação. Efetivamente, o que se demanda é um “procedimento contínuo, efetivo e transparente” de tratamento sobre as reclamações de conteúdos ilícitos. Procedimento que será monitorado diretamente pelas próprias plataformas por meio de controles mensais ou por uma instituição de autorregulação regulada. Nas palavras de Martin Eifert, tais aspectos evidenciam que o controle direto de qualidade deste *compliance-system* permanece, majoritariamente, a cargo dos provedores, cabendo ao Estado “um controle de qualidade indireto e não específico desse sistema”. Cf. EIFERT, Martin. Lei Alemã para a Melhoria da Aplicação nas Redes Sociais (Lei de aplicação da Lei nas redes – NetzDG). Trad. Pedro Henrique Ribeiro e Revisão de Júlio Carvalho. In: ABOUD, Georges; CAMPOS, Ricardo; NERY JR., Nelson (Orgs.). *Fake News e Regulação*. 2. ed. São Paulo: Revista dos Tribunais, 2020. pp.169-170.

A contraposição entre os dois modelos, contudo, evidencia a maior capacidade dos arranjos corregulatórios pela sua maior aptidão para (re)orientar o poder de governança das plataformas de redes sociais à observância de parâmetros constitucionais, em especial pelo receio provocado pelo poder fiscalizatório de um agente externo efetivamente imparcial. Modelos puramente autorregulatórios em um mercado altamente concentrado e dependente de tecnologias cujo funcionamento é pouco conhecido parecem ser insuficientes para lidar com os danos produzidos por abusos no exercício da liberdade no âmbito das redes sociais.

A proposta corregulatória mostra-se mais adequada ao desafio de assegurar que tais plataformas conduzam seu poder de governança de discursos com a seriedade que o direito fundamental à liberdade de expressão exige, se comparada ao modelo hoje existente que aposta na autorregulação destas empresas preconizada pelo art. 19 do MCI, havendo múltiplas razões para se sustentar sua constitucionalidade.

Em especial, é proposta que atende o postulado da proporcionalidade, em sua dupla dimensão³⁸³, especialmente se considerarmos sua capacidade de superar medidas legislativas notoriamente insuficientes (art. 5º, LIV da CRFB). Se mais comumente a proporcionalidade confunde-se com a proibição do excesso, é possível também observá-la como mandado de proibição à proteção insuficiente.³⁸⁴

O princípio da proibição da proteção insuficiente vem sendo reconhecido pelo STF tanto como argumento justificador da inconstitucionalidade de leis e atos normativos que impliquem redução significativa da proteção aos direitos fundamentais, quanto para justificar a constitucionalidade de leis e decisões que busquem impor ao Estado um dever mínimo de proteção a certos bens jurídicos, o que permite antever tal fundamento como argumento justificador de uma lei que institua mecanismos corregulatórios dos provedores de redes sociais.³⁸⁵

³⁸³ SARLET, Ingo Wolfgang. Constituição e proporcionalidade: o direito penal e os direitos fundamentais entre proibição de excesso e proibição de insuficiência. *Revista Brasileira de Ciências Criminais*, [s.l.], n. 47, mar./abr. 2004, p. 60-122.

³⁸⁴ Neste sentido, o Ministro Gilmar Mendes, citando Schlink, destaca que “se o Estado nada faz para atingir um dado objetivo (...) parece ferir um dever de atuação decorrente do dever de legislar (...). Se, se comparam, contudo, situações do âmbito das medidas protetivas, tendo em vista a análise de sua eventual insuficiência, tem-se uma operação diversa da verificada (...) que, nada mais é, do ponto de vista metodológico do que considerar tal conduta como desproporcional em sentido estrito”. Cf. MENDES, Gilmar Ferreira; BRANCO, Paulo Gustavo Gonet. *Curso de Direito Constitucional*. Editora Saraiva. 8. ed. Saraiva: São Paulo, 2013, p. 227.

³⁸⁵ A título de exemplo da primeira hipótese, cf. STF, ADI 5676/RJ, Tribunal Pleno, Min. Ricardo Lewandowski, Julgado em 18/12/2021 e ADI 3470/RJ, Tribunal Pleno, Min. Relatora Rosa Weber, Julgado em 29/11/2017. Como exemplo da segunda hipótese: HC 104.410/RS, Segunda Turma, Min. Rel. Gilmar Mendes, Julgado em 06/03/2012. AgRg no RE 763.667/CE, Min. Rel. Celso de Mello. Julgado em 22/10/2013.

O art.19 do MCI, isoladamente considerado, confere hoje uma proteção insuficiente contra os danos que a disseminação de certas espécies de discurso, no âmbito das redes sociais, causa para outros indivíduos, a sociedade e a democracia.³⁸⁶ A coexistência deste modelo de responsabilização civil com a construção de um modelo corregulatório para as plataformas é, portanto, uma necessidade de primeira ordem.

Por outro lado, a defesa de um arranjo corregulatório é também uma resposta necessária à luz dos riscos abordados neste trabalho. Se não elimina, por completo, o receio de captura do aparato tecnológico pelo Estado ou a produção de efeitos resfriadores (*chilling effect*) é, de certo, uma opção mais segura e eficiente³⁸⁷ do que a mera criação de novas ilicitudes discursivas (tais como desinformação, assédio virtual, *ciberbullying* e afins) como ferramenta de bloqueio à profusão de discursos danosos que proliferam em plataformas virtuais.

A defesa de um regime corregulatório, entretanto, diz pouco se desacompanhada de maiores reflexões sobre as pautas gerais de vinculação externa que serão impostas aos agentes privados e dos critérios para indicação da autoridade competente pelo exercício deste *compliance* externo.

De maneira geral, são apontados dois grandes objetivos às propostas de regulação externa da atuação de plataformas das redes sociais: a necessidade de maior compreensão tanto por parte do Estado quanto por parte da sociedade civil da forma como os códigos e algoritmos das plataformas influem o comportamento dos usuários/cidadãos, e de outro, da necessidade de assegurar que as decisões de moderação destas plataformas sejam ao mesmo tempo eficientes e legítimas.

A criação de deveres gerais que permitam assegurar que as redes sociais cumpram tais finalidades é tema do último capítulo, havendo inúmeros argumentos constitucionais aptos a legitimá-los, mas, antes de adentrar neste ponto, é importante frisar que a entidade que fará as vezes de autoridade externa fiscalizatória precisa ser projetada com os cuidados necessários para garantir sua necessária neutralidade política, de modo a reduzir os riscos de captura que podem vir a transformar o modelo corregulatório em um novo método para promoção da censura.

³⁸⁶ Neste sentido, faz-se referências aos casos concretos apresentados no capítulo segundo.

³⁸⁷ A eficiência decorre do fato de que as plataformas — pela sua ampla capacidade de governança e monitoramento — estão mais bem posicionadas para lidar de forma mais ágil e menos custosa com conteúdos danosos, ainda que não propriamente ilícitos, como o das “desordens informacionais”.

Esta discussão não é inédita e permeou inclusive a elaboração da Lei Geral de Proteção de Dados no Brasil (LGPD)³⁸⁸. Nesta lei se discutia, dentre outras coisas, os limites do uso dos dados pessoais por intermediários da *internet* em geral, e o cumprimento dos comandos preconizados pela lei ficaria por conta de um órgão denominado Autoridade Nacional de Proteção de Dados (ANPD), que estaria submetida hierarquicamente à Presidência da República.

No texto originariamente aprovado pelo Congresso Nacional, a LGPD previa a criação de uma autoridade nacional voltada para a proteção de dados no âmbito da administração indireta, e cujo desenho institucional era similar ao de uma agência reguladora, espécie de autarquia especial, que se caracteriza por sua independência funcional diferenciada se comparada às autarquias tradicionais. O modelo era concebido como ideal para regulamentação em concreto de recomendações, diretrizes e eventual aplicação de sanções em casos de dúvidas ou violações ao direito de proteção de dados pessoais. A versão final da lei, contudo, acabou por conferir à norma o caráter de “órgão da Administração Pública” (art. 5º, XIX, da Lei nº 13.709/2018), em que pese ter sido resguardada a possibilidade de conversão do órgão em autarquia (art.55-A, § 1º na redação anterior à estabelecida pela MP nº 1124/2022).

A referida decisão foi, acertadamente, objeto de diversas críticas por não afastar a ANPD de interesses políticos transitórios, dado o caráter técnico das discussões que deveriam ser travadas no órgão e a natureza fundamental do direito em análise³⁸⁹.

Os riscos evidenciados pela atribuição de natureza de órgão à autoridade administrativa que asseguraria o cumprimento e fiscalização das normas da LGPD são os mesmos observados na hipótese de criação de uma entidade externa responsável por analisar as informações das plataformas de redes sociais e expedir recomendações e normas fiscalizatórias para o cumprimento de obrigações gerais que serão atribuídas às redes sociais.

Por essa razão, parece-nos salutar que eventual regulação da transparência e liberdade de expressão das plataformas seja objeto, preferencialmente de fiscalização por entidade da administração pública de caráter técnico e independente.

³⁸⁸ O direito à proteção de dados fora recentemente reconhecido como direito fundamental à luz da emenda constitucional nº 115/2022, que incluiu o inciso LXXIX ao art. 5º da Constituição Federal.

³⁸⁹ Recentemente foi editada a Medida Provisória nº 1124/2022, encaminhada ao Congresso Nacional em 13/06/2022, prevendo a alteração do art. 55-A da Lei nº 13.709/2018 para transformar a ANPD em uma autarquia de natureza especial.

Frise-se que o Projeto de Lei nº 2630/2020³⁹⁰ previu, em sua redação original, que o conselho responsável por esta fiscalização seria um órgão de caráter multissetorial vinculado ao Poder Legislativo.

Na última versão do substitutivo apresentado pelo Deputado Federal Orlando Silva (PCdoB), ainda não aprovado, tal atribuição passou ao Comitê Gestor da *Internet* no Brasil (CGI.br), com o argumento de que o caráter multissetorial do órgão (*multistakeholder*)³⁹¹, bem como o bom desempenho técnico que apresentou no trabalho de promoção da racionalização da gestão, expansão e uso da *internet* no Brasil (art. 24, inciso II do MCI) sejam suficientes para afastar o temor de sua captura pelo governo.

O modelo corregulatório, portanto, se apresenta como a solução à disposição do Estado mais compatível com o desafio de regulação do poder de governança discursiva das principais plataformas. Evidentemente, a consagração de um controle indireto dos procedimentos de moderação e curadoria da plataforma por um órgão/entidade estatal impõe alguns cuidados ao legislador para evitar que tal medida se torne uma ferramenta de promoção de censura, tais como, a atribuição de tal função a um órgão/entidade primordialmente técnico, com representação multissetorial e funcionalizável mediante procedimentos públicos de deliberação

³⁹⁰ A última versão do projeto de lei pode ser conferida em: <https://www.camara.leg.br/propostas-legislativas/2256735>.

³⁹¹ O órgão fora criado pela Portaria Interministerial nº 147, de 31 de maio de 1995 e teve seu funcionamento e atribuições disciplinados pelo Decreto Federal nº 4.829/2003 com o propósito primordial de estabelecer diretrizes estratégicas relacionadas ao uso e desenvolvimento da *internet* no Brasil e diretrizes para execução dos registros de Nomes de Domínio, alocação de Endereço IP (*Internet Protocol*) e administração pertinente ao Domínio de Primeiro Nível “.br”. Diz-se que seu caráter multissetorial decorre da presença de representantes dos setores da sociedade civil que o compõem. *Vide* art. 2º do referido Decreto:

Art. 2º. O CGI.br será integrado pelos seguintes membros titulares e pelos respectivos suplentes: I - um representante de cada órgão e entidade a seguir indicados: a) Ministério da Ciência e Tecnologia, que o coordenará; b) Casa Civil da Presidência da República; c) Ministério das Comunicações; d) Ministério da Defesa; e) Ministério do Desenvolvimento, Indústria e Comércio Exterior; f) Ministério do Planejamento, Orçamento e Gestão; g) Agência Nacional de Telecomunicações; e h) Conselho Nacional de Desenvolvimento Científico e Tecnológico; II - um representante do Fórum Nacional de Secretários Estaduais para Assuntos de Ciência e Tecnologia; III - um representante de notório saber em assuntos de *Internet*; IV - quatro representantes do setor empresarial; V - quatro representantes do terceiro setor; e VI - três representantes da comunidade científica e tecnológica.

O PL nº 2630/2020 ainda tem o cuidado de exigir que todas as reuniões do CGI.br referentes ao tema da liberdade de expressão e transparência das redes sociais devam ser transmitidas na *internet*, de modo a conferir maior publicidade às deliberações, exigindo também um quórum mínimo de participação dos membros dos setores que o compõem para seu regular funcionamento.

4 A CONSTITUCIONALIZAÇÃO DO ESPAÇO DIGITAL E OS DEVERES DAS REDES SOCIAIS

4.1 Considerações gerais

Encerra-se esta pesquisa com uma proposta de regulação para as plataformas de redes sociais.

Como já se antecipou, arranjos corregulatórios (ou de autorregulação regulada) mostram-se mais eficazes como resposta para o duplo objetivo que justifica a criação de pautas de vinculação aos agentes privados que gerenciam tais plataformas: (I) assegurar o real compromisso dessas plataformas no combate à desinformação e outras formas de assédios virtuais que limitam a autonomia privada e distorcem o exercício da liberdade de expressão no âmbito da *internet*, gerando prejuízos sociais de ordem diversa; (II) conferir legitimidade ao exercício da moderação de conteúdo.

Muitos são os argumentos evocados para justificar a criação limites à autonomia das plataformas de redes sociais. Para Balkin, a limitação à autonomia privada das plataformas na gestão de seus negócios decorre do que ele denomina de obrigações dos “fiduciários informacionais” da *internet*, a partir de uma analogia do papel das plataformas no gerenciamento de dados privados de seus usuários com o dever especial de certas carreiras, tais como advogados, médicos e psicólogos, de não divulgar dados e informações sigilosas em prejuízo aos seus clientes/pacientes³⁹².

A despeito de o mérito da analogia de inculir parâmetros de boa-fé para atuação das plataformas, há quem critique tais premissas pela ingenuidade e prefira usar o argumento regulatório como contraponto à poluição decorrente da atividade de coleta de dados e construção de perfis.³⁹³

³⁹² A extensão destes deveres às plataformas das redes sociais seria decorrência de duas premissas centrais. Primeira, o reconhecimento da essencialidade da atividade de coleta de dados, para fins de exploração principalmente de publicidades direcionadas. Segunda, a necessidade de estabelecer algum parâmetro normativo para conferir limites éticos ao uso indiscriminado de informações privadas para fins manipulatórios. A existência de uma relação notoriamente desequilibrada (entre particulares e redes sociais) marcada por profunda assimetria informacional, exige das empresas alguma constrição na administração de tais dados (BALKIN, Jack. Information Fiduciaries and First Amendment. *UC Davies Law Review*, [s.l.], v. 4, n. 49, 2016).

³⁹³ Segundo estes autores, há um excesso de otimismo na visão de Balkin ao equiparar redes sociais a fiduciários, quando sabidamente os interesses lucrativos da empresa as movem em sentido contrário ao dos

As duas justificativas são úteis e se completam, mas perdem o foco do principal: a inviabilidade já comprovada de um modelo autorregulatório para se construir uma perspectiva de limitação material e procedimental ao poder de governança dos discursos em geral que estes entes privados possuem, fenômeno que a doutrina tem identificado como o avanço de um modelo de *constitucionalismo digital*.

O termo busca congrega o conceito de “digital” usualmente evocado para se referir às tecnologias desenvolvidas na *internet*, tais como ferramentas de processamento de dados e moderação de conteúdo, com a ideologia política construída no século XVIII, de influência *lockeana*, segundo a qual o poder deveria ser legalmente limitado e sua legitimidade dependeria de um controle recíproco acerca do cumprimento destas limitações.³⁹⁴

A lógica é a seguinte: as regras do Estado de Direito surgem em uma época em que o Estado-Nação era o agente mais poderoso para influenciar e governar a vida dos cidadãos. Estes fatos não correspondem mais à realidade contemporânea, pois as plataformas de redes sociais possuem capacidade similar ao Estado de influenciar e moldar a esfera pública, fazendo-o muitas vezes com uma ampla margem de discricionariedade e sem que haja uma efetiva responsabilização por suas ações, uma vez que tais agentes não se submetem regularmente ao escrutínio público.³⁹⁵

Nos dizeres de Nicolas Suzor, a constitucionalização em última instância envolve “um paradoxo pelo qual o poder de governança se torna legítimo porque aqueles que exercem o poder acreditam que estão eles próprios vinculados a suas próprias regras”.³⁹⁶

Em síntese, o movimento doutrinário em sentido a um constitucionalismo digital³⁹⁷ representa a internalização da lógica da eficácia horizontal dos direitos fundamentais³⁹⁸ para o âmbito da *internet*.

usuários/consumidores do serviço (como exigir confiança neste contexto?). Para alguns, a limitação da autonomia das plataformas decorre mais dos riscos decorrentes da “poluição de dados” (*data pollution*) do que propriamente de um compromisso ético de atuação destas plataformas. Neste sentido, POZEN, David; KHAN, Lina. A Skeptical View of Information Fiduciaries. *Harvard Law Review*, [s.l.], v. 133, n. 497, dez. 2019; e BEM-SHAHAR, Omri. Data Pollution. *Journal of Legal Analysis*, [s.l.], v. 11, 2019.

³⁹⁴ GREGORIO, Giovanni de. The Rise of Digital Constitutionalism in the European Union. *International Journal of Constitutional Law*, [s.l.], v. 19, n. 1, p. 41-70, jun. 2021.

³⁹⁵ Cf. NITRINI, Rodrigo Vidal. *Liberdade de Expressão nas Redes Sociais*. São Paulo: Dialética, 2021, p. 132.

³⁹⁶ SUZOR, Nicolas. *Lawless: the secret rules that govern our digital lives*. [S.l.]: Cambridge University Press, 2019, p. 110-120.

³⁹⁷ Nos dizeres de Nicolas Suzor, o principal mérito do termo “constitucionalismo” estaria no enfoque na questão da avaliação da “legitimidade (das decisões tomadas pelas plataformas) em contextos de discricionariedade”. Cf. Id. Digital Constitutionalism: Using the Rule of Law to Evaluate the Legitimacy of Governance by Platforms. *Social Media + Societ*, [s.l.], jul./set. 2018.

³⁹⁸ Em sentido similar, MENDES, Gilmar Ferreira; FERNANDES, Victor Oliveira. Constitucionalismo digital e jurisdição constitucional: uma agenda de pesquisa para o caso brasileiro. *Revista Brasileira de Direito*, Passo Fundo, v. 16, n. 1, p. 1-33, jan./abr. 2020.

O curioso da construção do constitucionalismo digital é que seu surgimento se deu como um argumento retórico defendido pelas principais *big techs* no intuito de legitimar o emprego de mecanismos autorregulatórios como o do *Oversight Board*, que confeririam maior racionalidade à sua atuação a partir de adesão a certos valores humanísticos³⁹⁹, mas que no seu desenvolvimento acabou avançando para um reconhecimento de confrontação do poder das plataformas, mediante a criação de mecanismos de *compliance* externos.

Não é exagero afirmar que a defesa de arranjos corregulatórios implica o reconhecimento do fim da fase libertária da *internet*, no qual se reconhecia a ampla autonomia dos intermediários como corolário da liberdade de expressão, para uma etapa que se caracteriza por debate sobre questões de transparência e *accountability*⁴⁰⁰⁻⁴⁰¹ por parte das plataformas e a busca por mecanismos regulatórios que limitem, de forma adequada, o poder dos intermediários na era digital.⁴⁰²

É razoável concluir que a busca e o aperfeiçoamento de arranjos corregulatórios, com a identificação de um núcleo mínimo de deveres destas plataformas é consequência lógica do avanço da tese do constitucionalismo digital.⁴⁰³

4.2 Os deveres das redes sociais

Quais seriam os deveres básicos das plataformas de redes sociais no exercício do seu poder de governança discursiva?

³⁹⁹ Neste sentido, afirma Keller: “(...) a ideia de um constitucionalismo digital – rótulo usado para se referir à constelação de iniciativas privadas que procuram articular um conjunto de direitos políticos, normas e limites de governança sobre o exercício do poder no âmbito da *Internet* – é uma das principais retóricas da governança da *internet*, que buscar certificar (ou ao menos tentar) que essa estrutura se mova no sentido do interesse público” (KELLER, Clara Iglesias. *Regulação Nacional de Serviços na Internet: Exceção, Legitimidade e o Papel do Estado*. Rio de Janeiro: Lumen Juris, 2019, p. 101).

⁴⁰⁰ Cf. CELESTE, Edoardo. Digital Constitutionalism: a new systematic theorisation. *International Review of Law, Computers and Technology*, [s.l.], v. 33, n. 1, p. 76-99, 2019.

⁴⁰¹ Quanto a este ponto, bons padrões de transparência são mencionados nos Princípios de Santa Clara, documento escrito em 2018, em conjunto, por acadêmicos e representantes da sociedade civil, no qual se consignou a construção de parâmetros mínimos para assegurar a transparência e prestação de contas na moderação de conteúdo por parte das plataformas privadas. Em síntese, o documento previa a necessidade de ampla divulgação dos números sobre postagens removidas; necessidade de notificação ao usuário afetado sempre que tiver uma postagem derrubada ou conteúdo suspenso e, por fim, a necessidade de disponibilização de um instrumento de “apelação” em prazo razoável. Disponível em: <https://santaclaraprinciples.org/>. Acesso em 27 jun. 2022.

⁴⁰² GREGORIO, Giovanni de. The Rise of Digital Constitutionalism in the European Union. *International Journal of Constitutional Law*, [s.l.], v. 19, n. 1, p. 41-70, jun. 2021.

⁴⁰³ Não se nega, portanto, que o termo “constitucionalismo digital” tem sofrido mutações e, não raro, sentidos diversos lhe são atribuídos (o próprio *Oversight Board*, como se viu, é defendido por alguns autores como concretização do fenômeno do constitucionalismo digital). O alcance que aqui se atribui, contudo, é o que se destacou acima. Não há constitucionalismo sem participação estatal e o que se pretende é defender a incorporação de limites ao exacerbado poder de governança discursiva das plataformas, considerando a natureza constitucional do direito à liberdade de expressão.

Defende-se aqui que a pauta de vinculação sujeita à fiscalização externa de um órgão independente e multissetorial se concentre em três deveres nucleares: o dever de transparência, o dever de observância ao devido processo legal e a proteção à boa-fé dos usuários.

Em que pese a nomenclatura adotada seja parcialmente original⁴⁰⁴, a construção do conteúdo de cada um dos parâmetros é fruto de recomendações doutrinárias que foram totais ou parcialmente abarcadas por legislações estrangeiras (em especial, a legislação alemã e o recente *Digital Service Act* da União Europeia⁴⁰⁵), que se anteciparam na regulação do tema ou hipóteses ventiladas em projetos de lei em estágios mais avançados, como é o caso do Projeto de Lei nº 2630/2020 no Brasil. Faz-se, portanto, de um esforço de compilação de boas ideias e práticas regulatórias. Vamos a cada um deles.

4.2.1 O dever de transparência

O dever de transparência consiste no reconhecimento de uma obrigação das plataformas de redes sociais em relação ao Estado e de usuários de compartilhamento de prestar informações tidas como essenciais para compreender se os processos de curadoria algorítmica (definição *a priori* dos critérios de visibilidade/invisibilidade das postagens) e de moderação (elaboração e aplicação dos termos de uso e normas comunitárias) estão ocorrendo com observância dos direitos fundamentais dos usuários, em especial, do direito fundamental à liberdade de expressão.

Trata-se da mais importante das obrigações das plataformas,⁴⁰⁶ pois é a partir dela que o Poder Público e a sociedade civil verificarão o cumprimento das obrigações que as redes sociais possuem de garantir o bom funcionamento do livre mercado de ideias e de uma esfera pública efetivamente democrática, viabilizando a existência do próprio arranjo corregulatório.

Há dois evidentes ganhos na criação de um procedimento de troca de informações entre plataformas de redes sociais e Poder Público. O primeiro é destacado por George Abboud e Ricardo Campo, que veem nestas regras “uma dimensão de abertura para um constante processo

⁴⁰⁴ Nenhuma legislação menciona propriamente um dever de “proteção à boa-fé dos usuários”, embora seja comum a referência ao dever de transparência e a um dever de diligência ou cuidado atrelado à observância do devido processo legal.

⁴⁰⁵ A norma foi aprovada em 05 de julho de 2022 pelo Parlamento Europeu e entrará em vigor em 2024.

⁴⁰⁶ Nenhuma das propostas corregulatórias criadas ou em concepção prescindem da necessidade de se reconhecer que as plataformas de redes sociais deverão compartilhar informações relevantes sobre o processo de moderação e curadoria algorítmica que utilizam.

de aprendizagem para temas complexos”⁴⁰⁷ que permite a construção de um “direito administrativo aberto à aprendizagem”⁴⁰⁸. Em outros termos, aprimora-se a eficiência da regulação por meio de uma troca constante de informação entre agentes público e privados, que permite, gradualmente, o aprimoramento da regulação pública à medida em que se reduz a assimetria informacional sobre o funcionamento das novas tecnologias.

O outro ganho decorre do fato de que a disponibilização de relatórios informacionais públicos torna a relação entre Estado e intermediários mais *republicana*. Como se percebeu da análise das decisões do *Oversight Board*, parte das recusas da empresa Meta em informar detalhes técnicos sobre os critérios que orientaram o apagamento ou redução do alcance de certos conteúdos se deve a suspeitas de pressão que certos agentes governamentais fazem sobre o exercício do poder de moderação destas empresas.⁴⁰⁹

O dever de prestação de contas, se de um lado impõe um ônus às plataformas privadas, também confere maior transparência à relação entre Estados e redes sociais. Ao criar um canal de comunicação oficial entre governo e plataformas privadas, a medida viabiliza um processo participativo, que permitirá ao usuário/cidadão avaliar não só eventuais falhas das plataformas, mas também aferir criticamente se as pressões da máquina estatal sobre a governança das redes sociais podem estar ocorrendo de forma exacerbada, o que justificaria inclusive pedidos de informações complementares,⁴¹⁰ no caso de suspeitas do uso destas tecnologias para exercício de censura. Tal lógica parece não só em conformidade com o postulado fundamental republicano (art. 1º, *caput* da CRFB), mas também com garantias constitucionais básicas como a da publicidade (art. 37, *caput* da CRFB), o direito de acesso à informação (art. 5º, inciso XIV da CRFB) e o dever de transparência estatal (art. 5º, inciso XXXIII da CRFB).

⁴⁰⁷ ABBOUD, Georges; CAMPOS, Ricardo. A autorregulação regulada como modelo do Direito proceduralizado. In: ABBOUD, Georges; CAMPOS, Ricardo; NERY JR., Nelson (Orgs.). *Fake News e Regulação*. 2. ed. São Paulo: Revista dos Tribunais, 2020, p. 129.

⁴⁰⁸ *Ibid.*, p. 134.

⁴⁰⁹ Um exemplo dessa relação pouco transparente entre Estado e redes sociais pode se observar pela própria dinâmica do episódio que resultou no banimento de Trump de diversas redes sociais após a invasão do Capitólio, em 06 de janeiro de 2021. Durante todo o seu mandato, o Facebook e demais redes sociais se recusaram em exercer a moderação dos discursos do então presidente, sob o argumento de que tais informações eram um conteúdo de “interesse público” (*newsworthy*). A despeito de Trump testar constantemente os limites das plataformas, a barreira do “interesse público” só foi superada após a sua derrota na eleição norte-americana. Neste sentido, João Archegas sugere que a decisão de banimento só fora tomada porque o presidente, ao perder a eleição, deixou de se tornar uma “ameaça política para os barões do Vale do Silício”. Cf. ARCHEGAS, João Victor. Admirável 2020 Novo: banimento de Trump de redes sociais e moderação de conteúdo. *Jota*, [s.l.], 13 jan. 2021. Disponível em: <https://www.jota.info/opiniao-e-analise/artigos/admiravel-2020-novo-banimento-de-trump-de-redes-sociais-e-moderacao-de-conteudo-13012021>. Acesso em: 24 jun. 2022.

⁴¹⁰ A partir do momento em que o Poder Público constrói, efetivamente, uma política pública de controle do poder de governança das plataformas, expedientes como o da Lei de Acesso à Informação podem ser utilizados pelo cidadão/usuário para rogar informações complementares sobre o tratamento que o Poder Público tem dado a tais relatórios. É dever do Poder Público ter elementos, passíveis de disponibilização ao público, que embasem eventual sanção e/ou recomendação às redes sociais.

O desenvolvimento de trabalhos acadêmicos sobre a governança de discursos em rede viabilizou a identificação de alguns elementos que devem constar de tais relatórios, quais sejam:

- a) número total de usuários que acessam os provedores a partir de conexões do país;
- b) número de total de reclamações relativas à violação dos termos e políticas de uso próprio dos provedores;
- c) número e discriminação das medidas aplicadas a contas e conteúdo, em especial àquelas que influenciem a visibilidade/disponibilidade dos conteúdos, indicando se oriundas de mecanismos automatizados ou atuação humana;
- d) o tempo decorrido entre as reclamações e adoções das medidas por parte das plataformas;
- e) o número total de pedidos de revisão apresentados por usuários às medidas aplicadas às contas e conteúdo, em razão de suposta violação aos termos e políticas de uso, indicando ainda o percentual de reforma discriminado por conteúdo;
- f) número total de medidas aplicadas em decorrência de decisão judicial por força de violação da legislação local com indicação do prazo para execução das medidas;
- g) informações sobre a equipe envolvida na aplicação dos termos de uso, com indicativos sobre competência linguística, critérios de contratação, nacionalidade, atributos demográficos e índice de precisão na aplicação das medidas de moderação (por exemplo, por meio da indicação do percentual de casos revistos, segmentando por fundamento da intervenção).
- h) explicação sobre os meios automatizados utilizados para moderação de conteúdo, acompanhado de informações relevantes, como: desempenho dos sistemas automatizados na compreensão de aspectos próprios da língua portuguesa; identificação dos objetivos pretendidos com a automatização, índice de precisão no cumprimento destes objetivos e demais informações úteis à proteção dos direitos dos usuários;
- i) alterações ocorridas nas políticas próprias dos termos de uso feitos no período abarcado pelos relatórios, com indicação das razões que motivaram sua alteração.
- j) considerações gerais a respeito dos esforços adotados pelos provedores de redes sociais na estruturação da arquitetura das plataformas, para impedimento de ações ilícitas, incluindo ainda avaliação de riscos sistêmicos decorrente do funcionamento e utilização de seus serviços de perfilamento e publicidade direcionada.
- k) repositório com indicação dos anúncios patrocinados e pagos distribuídos nas plataformas, com indicação do titular anunciante, valor pago e critérios de direcionamento por perfis selecionados por anunciantes.⁴¹¹

Este rol, evidentemente, está em constante construção, visto que o conhecimento sobre o funcionamento das plataformas de redes sociais é um processo ainda em curso, mas tais elementos têm sido identificados de forma mais ou menos consensual por boa parte dos projetos regulatórios, em diversos países, como informações minimamente úteis para compreender se as ações das redes sociais no combate às desordens informacionais tem se de dado de forma eficiente e com observância de garantias mínimas dos usuários.

⁴¹¹ Parte deste rol é decorre de um esforço de compilação de deveres de transparência atribuídos às plataformas por documentos oficiais diversos (DSA, *NetzDG*, PL2630, *White Online Harms Papers*), mas em doutrina pode-se destacar o seguinte trabalho (MACCARTHY, Mark. Transparency Requirements for digital social media platforms: recommendations for policy makers and industry. *Transatlantic Working Group*, [s.l.], fev. 2020. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3615726. Acesso em: 01 nov. 2022).

De forma geral, duas críticas têm sido apresentadas à consolidação de um dever de transparência: a) de que o excesso de informação e a aparente transparência podem vir a ser manobrados para ocultar informações e confundir a opinião pública, a chamada “transparência opaca”;⁴¹² b) de que o excesso de transparência possa implicar violações a garantias constitucionais como a privacidade e o segredo empresarial.

A crítica quanto ao excesso de transparência é pertinente: mais transparência não necessariamente significa maior conhecimento⁴¹³. Só há transparência de fato se os elementos sejam compreensíveis tanto ao Poder Público quanto à sociedade civil. Para que isso ocorra, alguns cuidados precisam ser tomados.

Primeiro, é preciso que haja concordância quanto aos conceitos técnicos, critérios e métricas que embasarão os relatórios e que são comumente utilizados no combate às desordens informacionais, tais como perfilhamento, microdirecionamento, contas falsas, contas automatizadas e operações inautênticas coordenadas. Esta dificuldade é destacada por Douek e Pottier, em interessante artigo no qual ressaltam a impossibilidade de se comparar o desempenho das plataformas no combate às campanhas coordenadas de desinformação, visto que cada rede social usa um conceito diverso para se referir a esta espécie de articulação.⁴¹⁴

É dever do legislador, portanto, buscar a unificação de critérios e conceitos para que seja possível não apenas a comparação de atuação entre as várias plataformas, mas a redução da assimetria informacional que impede uma adequada proteção dos direitos básicos dos usuários do serviço.

É fundamental também que haja alguma prestação de contas. Transparência sem *accountability* é quase sempre ineficaz. As últimas eleições nacionais do país, tanto em 2018 quanto em 2020, revelam que compromissos públicos de transparência firmados entre o Tribunal Superior Eleitoral e principais plataformas de redes sociais para redução do alcance

⁴¹² FOX, Jonathan. The uncertain relationship between transparency and accountability. *Development in practice*, [s.l.], v. 17, n. 4-5, p. 663-671, 2007, p. 247.

⁴¹³ O prejuízo decorrente do excesso de informações é abordado de forma mais ampla em recente obra do professor Cass Sunstein. Cf. SUNSTEIN, Cass. *Too Much Information: understanding what you don't want to know*. Cambridge, MA: The Mit Press, 2020.

⁴¹⁴ As autoras citam exemplo dos conceitos “*Coordinated Inauthentic Behavior*” (CIB) do Facebook; “*Information Operations*” (IO), no Twitter e “*Coordinated Inauthentic Behavior*”, do Google. Conceitos criados para mesma finalidade, mas com características diversas, o que inviabiliza a comparação da eficiência das plataformas na remoção de conteúdos que violem seus termos de uso ou sejam ilícitos. Cf. DOUEK, Evelyn; FRANÇOIS, Camille. The Accidental Origins, Underappreciated Limits and Enduring Promises of Platform. Transparency Reporting About Informations Operations. *Journal of Online Trust and Safety*, [s.l.], out. 2021.

de campanhas de desinformação durante o período eleitoral tiveram, no mínimo, uma eficácia questionável.⁴¹⁵

Compromissos públicos de transparência, ao menos em um campo de conhecimento tão opaco como o poder de governança das redes sociais, não prescindem de mecanismos de fiscalização externos, o que reforça a necessidade de construção de um padrão corregulatório como algo fundamental para assegurar alguma ferramenta de coerção sobre tais plataformas privadas. Sugestão interessante para reduzir a opacidade das plataformas advém de regra constante no art. 28 do *Digital Services Act* (DSA), que cria para as plataformas de “dimensão muito grande” um dever de contratação anual de auditoria independente, para fins de avaliação do cumprimento das obrigações estabelecidas naquela legislação ou quaisquer outras condutas assumidas em códigos de conduta e protocolos de crise das plataformas.⁴¹⁶

É preciso, ainda, que os relatórios sejam acessíveis não somente aos especialistas, mas aos usuários e à sociedade civil como um todo. A título de sugestão, poder-se-ia cogitar que os próprios relatórios de transparência fossem divididos em dois documentos distintos, a exemplo do que ocorre na seara ambiental, na elaboração dos estudos prévios de impacto ambiental (EIA) e no relatório de impacto ao meio ambiente (RIMA). O primeiro, mais completo e significativamente mais extenso, poderia ser mais útil ao Estado, pesquisadores e terceiros especialmente interessados na análise dos dados sobre moderação das plataformas; ao passo que o segundo, mais sucinto, poderia ser distribuído nas redes sociais para conhecimento dos usuários em geral mediante uma rápida leitura. Prestigia-se, assim, o processo de ampla “ativação da esfera pública”, que é essencial para a manutenção de um contínuo escrutínio sobre o poder de governança das plataformas privadas.

⁴¹⁵ A impressão é compartilhada por ex-funcionária da plataforma que capitaneou o processo de vazamento de documentos internos da Meta, escândalo que ficou conhecido como Facebook Papers. Informação disponível em: DAROS, Gabriel. Frances Haugen, do Facebook Papers: ‘Meta não protege eleições no Brasil’. *Tilt Uol*, [s.l.], 6 jul. 2022. Disponível em: <https://www.uol.com.br/tilt/noticias/redacao/2022/07/06/frances-haugen-do-facebook-papers-meta-nao-protege-eleicoes-no-brasil.htm>. Acesso em: 20 dez. 2022. Essa impressão fora confirmada por estudos recentes que atestam a negligência das principais plataformas na retirada de conteúdo eleitoral falso. A respeito, confira-se: MELLO, Patrícia Campos. Facebook permite conteúdo eleitoral classificado como falso, apesar de acordo com TSE. *Folha de S. Paulo*, [s.l.], 20 out. 2022. Disponível em: <https://www1.folha.uol.com.br/poder/2022/10/facebook-permite-conteudo-eleitoral-classificado-como-falso-apesar-de-acordo-com-tse.shtml>. Acesso em: 01 jan. 2022.

⁴¹⁶ Considera-se como plataformas de “muito grande dimensão” aquelas que prestam seus serviços a um número médio mensal de usuários ativos na União Europeia igual ou superior a 45 milhões, calculados de acordo com a metodologia estabelecida na norma (art. 25).

A segunda crítica, comumente evocada pelas empresas que gerenciam tais plataformas para negar acesso a dados sensíveis que possuem à sua disposição, pode ser superado mediante o que Luna Barroso denomina de uma “abordagem em camadas”⁴¹⁷ do dever de transparência.

Se os dados acima enumerados podem ser divulgados por meio de relatórios públicos acessíveis aos usuários e pesquisadores em geral, é possível que documentos mais sensíveis, que possam resvalar em violação a questões de privacidade e/ou segredo empresarial⁴¹⁸, sejam ofertados mediante “um sistema seguro, supervisionado por um órgão independente, que seria responsável por um processo de habilitação destes pesquisadores ou órgãos regulatórios estatais”⁴¹⁹, de modo a assegurar que o fornecimento de dados sensíveis ocorra sempre de forma desagregada e anonimizada, mediante assunção de compromissos específicos de responsabilidade por parte dos interessados. Trata-se de solução que resguarda de forma proporcional o interesse dos usuários e os direitos de privacidade de terceiros, bem como os legítimos interesses econômicos das plataformas de redes sociais.

4.2.2 O devido processo legal

O dever de observância ao devido processo legal⁴²⁰ consiste na exigência de que a decisão de moderação destas plataformas, ainda que fundada em seus termos de uso, sejam motivadas, por implicarem restrições a um direito fundamental de liberdade.

A motivação legitimadora de tais decisões demanda principalmente a construção de procedimentos decisórios, que disponibilizem aos usuários/cidadãos não só um mecanismo facilmente compreensível de denúncias, mas que também lhes permita se defender quando denunciados, conferindo compreensão sobre as razões da decisão da plataforma, garantindo a aplicação de sanções proporcionais e, ainda, viabilizando recursos, especialmente nos casos de decisão automatizada.

⁴¹⁷ BARROSO, Luna Van Brussel. *Liberdade de Expressão e Democracia na Era Digital*. Belo Horizonte: Fórum, 2021, p. 247.

⁴¹⁸ Poder-se-ia cogitar informações mais específicas sobre o código-fonte da plataforma, por exemplo.

⁴¹⁹ *Ibid.*, p. 250.

⁴²⁰ O Reino Unido propõe um modelo de corregulação das plataformas digitais fundada em um dever geral de cuidado (*duty of care*), que em síntese implica observância de garantias procedimentais mínimas na tomada das decisões de moderação. Relatório disponível em: Online Harms White Paper - GOV.UK (www.gov.uk). Acesso em: 24 jun. 2022.

Tal exigência consubstancia o primado da *eficácia horizontal dos direitos fundamentais*⁴²¹, teoria de origem alemã⁴²², já consagrada na jurisprudência do Supremo Tribunal Federal.⁴²³

Em síntese, a teoria sustenta que particulares estão vinculados aos direitos fundamentais em suas relações quando “exercçam alguma atividade de caráter público, ainda que sem qualquer dependência formal em relação ao Estado”⁴²⁴ ou, também, quando estejam em uma posição de notória superioridade em relação a outro particular, de modo que a incidência dos direitos fundamentais, ainda que de forma limitada, faça-se necessária para evitar que a relação bilateral se torne um mecanismo de opressão.⁴²⁵

É possível sustentar a eficácia dos direitos fundamentais na relação entre usuários e redes sociais com base nos dois argumentos. As principais redes sociais possuem capacidade de influenciar e moldar a esfera pública, por meio do seu exercício de poder de governança privado, impactando debates sociais e políticos que transcendam o universo do *online*⁴²⁶, o que

⁴²¹ As principais considerações sobre o tema podem ser encontradas em SARMENTO, Daniel. *Direitos Fundamentais e Relações Privadas*. Rio de Janeiro: Lumen Juris, 2004; SILVA, Virgílio Afonso da. *A constitucionalização do direito: os direitos fundamentais nas relações entre particulares*. São Paulo: Malheiros, 2005.

⁴²² A certidão de nascimento da referida teoria remonta ao caso Lüth, julgado no qual o Tribunal Constitucional Alemão considerou como legítima a medida de boicote promovida por Eric Lüth ao filme “Amada Imortal”, do cineasta Veit Harlen, sob o argumento de o diretor ter sido um dos maiores cineastas que atuaram na promoção do governo nazista. O Tribunal Constitucional Alemão julgou que a medida de boicote fora proporcional à luz do caso concreto, destacando que os interesses de Lüth não eram meramente egoísticos, mas em conformidade com o interesse fundamental do povo alemão. Na ementa do julgado, é destacado que, embora os direitos fundamentais sejam em primeira linha um direito de defesa do cidadão face ao Estado, ele também se corporifica em uma ordem axiológica objetiva, que vale para todas as áreas do direito como uma decisão fundamental constitucional.

⁴²³ O *leading case* é de 2005 e remonta a precedente que julgou como inconstitucional medida de expulsão automática de particular de associação de proteção de direitos autorais. No caso, reforçou-se a necessidade de que a expulsão de uma associação privada deve observar o devido processo legal. Cf. REExt 201.819-8/RJ, Segunda Turma, Min. Rel. para o acórdão Gilmar Mendes. Julgado em 11/10/2005.

⁴²⁴ SARMENTO, Daniel. Op. cit., p. 317.

⁴²⁵ Pode-se citar como exemplo de incidência destas hipóteses o direito do trabalho e o direito do consumidor. Mesmo o direito civil, em sua acepção geral, sofre algumas derrogações de ordem pública por força da eficácia horizontal dos direitos fundamentais, fenômeno que ficou reconhecido pela doutrina civilista como “constitucionalização do direito civil”.

⁴²⁶ Nos Estados Unidos, tese similar é adotada sob o nome de *function public theory*, que reconhece a possibilidade de se estender o regime dos direitos fundamentais a particulares que exercem funções tipicamente estatais. O caso mais emblemático de aplicação desta teoria foi o caso Marsh vs Alabama, julgado em 1946, onde se discutia se uma empresa privada, que possuía terras no interior das quais se localizavam ruas, residências e estabelecimentos comerciais poderia proibir Testemunhas de Jeová de pregar em seu interior. A Suprema Corte Norte-Americana entendeu que a empresa se equiparava a uma cidade, exercendo funções municipais, o que a sujeitaria à primeira emenda (326. U.S 501 (1946)). Vale frisar que há precedente da SCOTUS (*Packingham vs North Carolina*), que utilizou a teoria da *function public theory* como argumento para justificar a inconstitucionalidade de uma lei da Carolina do Norte que impedia criminosos sexuais de se registrarem em plataformas como Facebook e Twitter. No voto condutor do Justice Kennedy, foi destacado que essas plataformas são a “praça pública moderna”, essencial ao exercício da liberdade de expressão, e impedir alguém em absoluto de acessá-la implicaria, talvez, o “mais poderoso mecanismo disponível para privar um cidadão de ter sua voz escutada”. Embora a força vinculante do precedente esteja relacionada, tão somente, ao impedimento de acesso em definitivo a tais plataformas (não adentrando no debate sobre os limites do exercício da sua autonomia gerencial), o julgado se tornou um importante

aproxima sua função do caráter público. Por outro lado, há um evidente desequilíbrio na relação jurídica entre usuários dos serviços das plataformas e redes sociais, uma vez que o exercício da moderação pelas plataformas pode não só implicar silenciamento injusto do usuário penalizado com efeitos similares ao da censura estatal, como eventual exclusão ou suspensão de perfil, como pode também produzir prejuízos significativos de ordem material, já que na sociedade contemporânea tem sido cada vez mais habitual o uso das redes sociais como meio de divulgação de atividade profissional ou econômica.

Vale ressaltar que este talvez seja o menos controvertido dos deveres aos quais as plataformas de rede sociais estão sujeitas, visto que regulações usualmente propostas, por partidos de matizes ideológicos diversos, reforçam a necessidade de transparência e fundamentação dos métodos decisórios das plataformas privadas⁴²⁷. O ponto é aquele que encontra menor resistência, inclusive, das plataformas.⁴²⁸

Sob o ponto de vista legal, é possível inclusive afirmar que o Marco Civil da *Internet* já apresenta uma norma legal capaz de fundamentar a exigência de um devido processo legal na tomada de decisões por parte das aplicações de conteúdo; notadamente, o art. 20 da referida lei menciona que ao se deparar com conteúdo ilícito caberá às aplicações de *internet* “comunicar-lhe os motivos e informações relativos à indisponibilização de conteúdo, com informações que permitam o contraditório e a ampla defesa em juízo”.

Alguns questionamentos podem ser suscitados quanto se pensa na criação de mecanismos de reclamação e recursos internos a estas plataformas que assegurem a adoção de decisões razoavelmente fundamentadas. O primeiro é se a ideia de um devido processo legal substancial não implicaria a aniquilação da autonomia destas plataformas e, em segundo lugar, se a exigência de criação destes canais de reclamação e recursos não gerariam um ônus

argumento retórico na defesa da função quase-pública das redes sociais nos Estados Unidos (137 S. Ct. 1730 (2017)). No Brasil, houve já alguma sinalização neste sentido quando a Ministra Rosa Weber, em decisão monocrática, suspendeu os efeitos da Medida Provisória nº 1068/2021 por considerar seu impacto sobre questões de cidadania e direitos políticos (MC-ADI 6991/DF).

⁴²⁷ Como já se apontou logo no início da dissertação, mesmo libertários têm defendido a necessidade de se regular o procedimento de moderação das redes sociais. A diferença está no escopo da proposta desta regulação. Enquanto modelos libertários buscam restringir também substancialmente o poder de moderação das plataformas, limitando os conteúdos que podem ser retirados das páginas pelos entes privados que administram tais canais de comunicação, os defensores de uma concepção democrática da liberdade de expressão buscam dar ênfase à transparência argumentativa e procedimental, não adentrando em aspectos de conteúdo.

⁴²⁸ A regulamentação de um procedimento para a racionalização da decisão das plataformas foi defendida por Mark Zuckerberg em artigo escrito para o *Washington Post*. Informação disponível em: ZUCKERBERG, Mark. The Internet needs new rules. Let's start in these four áreas. *The Washington Post*, [s.l.], 30 mar. 2019. Disponível em: https://www.washingtonpost.com/opinions/mark-zuckerberg-the-internet-needs-new-rules-lets-start-in-these-four-areas/2019/03/29/9e6f0504-521a-11e9-a3f7-78b7525a8d5f_story.html. Acesso em: 10 nov. 2022.

excessivo, que poderia implicar insuportável custo de entrada a novos *players* em um mercado já altamente concentrado.

Sobre o primeiro ponto, cumpre relembrar os ensinamentos de Virgílio Afonso da Silva de que não é possível demandar dos entes privados que promovam apenas as limitações necessárias a direitos fundamentais nas relações contratuais, visto que “se aos particulares não resta outra solução que não a adoção das medidas estritamente necessárias, não se pode mais falar em autonomia”.⁴²⁹ Em outras palavras, é preciso que as plataformas tenham discricionariedade para criar regras materiais e procedimentais em seus ambientes, que não precisam guardar identidade com as normas legais.

Ainda que tal ideia possa parecer uma obviedade para muitos, várias propostas regulatórias debatidas na atualidade parecem esquecer tais premissas. A título de exemplo, cumpre relembrar os comentários já tecidos sobre a Medida Provisória nº 1068/2021 (*vide* capítulo 2.5.2). Para além dos problemas pertinentes à censura reversa, exigir que a exclusão de conteúdo das plataformas somente decorra de violação da lei ou dependa de decisão judicial é obrigação que asfixia por completo a autonomia privada.

A própria Procuradoria-Geral da República, no bojo do recurso extraordinário nº RE 1.037.396/SP-RG, no qual se debate a constitucionalidade do art. 19 do MCI, sustenta a constitucionalidade da norma sob o aparente argumento de que a exclusão de conteúdo das redes sociais depende exclusivamente de decisão judicial, não tendo as plataformas autonomia para suprimir discurso de suas páginas. A manifestação, além de evidenciar desconhecimento do próprio conteúdo do art. 19 do MCI, revela uma interpretação em desconformidade com aquilo que preconiza a teoria da eficácia horizontal dos direitos fundamentais.⁴³⁰

O que se exige das plataformas de redes sociais é transparência decisória; assim, os critérios devem ser isonômicos — a princípio — e mesmo em caso de distinções na sua aplicação, a depender da categoria dos usuários, tais critérios devem ser de conhecimento geral. As aplicações das regras de moderação devem se dar sempre de forma fundamentada, ainda que se embases em termos, atos e procedimentos próprios.

⁴²⁹ SILVA, Virgílio Afonso da. *A constitucionalização do direito: os direitos fundamentais nas relações entre particulares*. São Paulo: Malheiros, 2005, p. 163-164.

⁴³⁰A íntegra do parecer pode ser conferida em: <https://portal.stf.jus.br/processos/downloadPeca.asp?id=15338769612&ext=.pdf> Acesso em: 11 nov. 2022.

Um exemplo do que *não* deve ser feito pelas plataformas das redes sociais à luz do ideal de devido processo legal pode ser obtido a partir exemplo da aplicação da suposta regra da figura pública ou do “interesse noticioso” (*newsworthy*) pelo Facebook.⁴³¹

Desde 2009, executivos do Facebook relatavam a criação de parâmetros de exceção para justificar, dentre outras coisas, a concessão de maior liberdade discursiva para autoridades públicas, justificando um menor rigor na aplicação das normas de moderação para contas administradas por políticos em geral. Parâmetro similar foi seguido por plataformas como Twitter. De modo geral, tal parâmetro se sustentava na necessidade de se proteger materiais considerados noticiosos, por envolverem figuras públicas cujas manifestações eram de interesse geral. Especulou-se que tal parâmetro, por exemplo, teria justificado a manutenção de inúmeros comentários do ex-Presidente Donald Trump, que incitavam o ódio (tome-se por exemplo, quando afirmou no Twitter “quando começam os saques, começam os tiros”, em referência a alguns episódios de violência que se observaram durante o ápice do movimento *Black Lives Matter*)⁴³².

Para surpresa geral, contudo, no mais emblemático dos julgamentos do *Oversight Board*, o Facebook informou que nunca usara a exceção para proteger qualquer manifestação na plataforma do ex-presidente, o que levou o Comitê de Supervisão a exigir maior transparência e clareza da empresa na aplicação deste parâmetro para evitar suspeitas quanto ao seu uso injustificado.⁴³³ Recentes informações divulgadas por ex-funcionários do Facebook, no *Wall Street Journal* (“*Facebook Papers*”), revelam, contudo, que a informação prestada ao comitê, possivelmente, não foi verdadeira, já que segundo os documentos vazados o critério teria sido utilizado de forma indiscriminada pela plataforma para proteção de perfis produtores de maior engajamento, mesmo em caso de discursos de assédio, incitação à violência ou desinformação danosa, condutas vedadas pelos termos de uso da plataforma.

O uso da exceção é questionável, visto que atribui maior liberdade discursiva àqueles que deveriam ter maior cuidado em suas manifestações por força da posição de liderança que exercem.⁴³⁴ Entretanto, o ponto mais grave não é a criação da distinção em si, mas sim não saber exatamente quando tal critério foi aplicado e quais publicações foram mantidas em razão

⁴³¹ Um histórico da criação do parâmetro pode ser obtido em KADRI, Thomas; KLONICK, Kate. Facebook v. Sullivan: public figures and newsworthiness in online speech. *Southern California Law Review*, [s.l.], v. 93, 2019.

⁴³² Informação disponível em: TRUMP sugere atirar em manifestantes, e Twitter inclui aviso sobre glorificação da violência. *Folha de S. Paulo*, [s.l.], 29 mai. 2020. Disponível em: <https://www1.folha.uol.com.br/mundo/2020/05/trump-sugere-atirar-em-manifestantes-e-twitter-inclui-aviso-sobre-glorificacao-da-violencia.shtml>. Acesso em: 11 nov. 2022.

⁴³³ O caso é abordado com maior profundidade no capítulo 3.4.

⁴³⁴ Aborda-se o tema no capítulo 3.1.

da sua utilização, inviabilizando um escrutínio público sobre o parâmetro. Sabe-se que o critério existe, mas as plataformas (em especial, a Meta) têm feito reserva mental dos casos no qual a aplicaram.

A criação de regras diferenciadas para autoridades públicas, inclusive, pode ser bem-vinda, desde que a incidência se dê de forma fundamentada. Neste ponto, anda bem, por exemplo, a última versão do Projeto de Lei nº 2630/2020, que previu parâmetros diferenciados e, no geral, mais rigorosos para fins de moderação de contas administradas por autoridades políticas.⁴³⁵

O segundo questionamento, por sua vez, é passível de superação pela inclusão de um limitativo material à exigência de criação de canais de reclamação e recursos em face de medidas de moderação, afastando o dever de criação de uma burocracia de governança discursiva para aplicações de rede de menor expressão. Tal norma somente se aplicaria às plataformas de redes sociais de grande alcance.

A limitação à autonomia para supressão de perfis e discursos em redes aplicar-se-ia somente às plataformas que cumprem o papel de *gatekeeper* do espaço informacional da *internet*. Por essa lógica seria impossível aos provedores de conexão censurar qualquer espécie de conteúdo, devendo guardar estrita observância do parâmetro de neutralidade a que aludimos no capítulo segundo. Como só há acesso à *internet* por meio destas plataformas, não pode lhe ser facultado realizar qualquer espécie de controle discursivo.

Plataformas de redes sociais de grande alcance também têm o potencial de restringir a liberdade de expressão no espaço público, embora sem impedir o acesso à *internet* em si. Isto significa que ainda que tais aplicações tenham alguma liberdade para moderar conteúdos conforme seus termos de uso, devem respeitar algumas garantias substanciais e procedimentais mínimas. Por sua vez, plataformas de menor alcance teriam uma autonomia quase que plena na supressão dos conteúdos que violem seus termos de uso e normas comunitárias, dado que sua atuação não implica um limite substantivo à circulação de discursos na praça pública da *internet*. Assim, o ônus de criação de verdadeiras burocracias tecnológicas para moderação não subsistiria.⁴³⁶

⁴³⁵ O ponto é tratado no artigo 22 presente na última versão do projeto de lei, e prevê a submissão da conta de entidades públicas, agentes políticos e determinadas categorias de servidores a princípios próprios da Administração Pública, enumerados no art. 37 da Constituição Federal. A nota negativa fica por conta da dúbia redação do § 8º do art. 22, que pode induzir à interpretação de que as imunidades parlamentares impedem ou limitam a aplicação de medidas de moderação por parte das plataformas de redes sociais, o que pode legitimar a atuação desses usuários como poderosos disseminadores de desinformação.

⁴³⁶ O argumento é desenvolvido por LAIDLAW, Emily. A Framework for Identifying Internet Information Gatekeepers. *International Review of Law, Computers & Technology*, [s.l.], v. 24, n. 3, 2010.

É essa lógica que tem embasado regras que limitem a incidência de regimes corregulatórios somente a plataformas que alcancem um número específico de usuários indicados pela lei como necessário para configurar aquela plataforma de rede social como de grande impacto.⁴³⁷

É ainda relevante indagar se o direito a um devido processo legal deveria abarcar o direito de o usuário ter sua decisão revisada por uma instância humana, visto que a ampla maioria das medidas de moderação são aplicadas de forma automática por algoritmos construídos pelas plataformas de redes sociais.

O ponto costuma ser abordado a partir do que se denomina de direito à explicação, que engloba o direito de o particular de compreender os critérios que pautaram decisões tomadas por sistemas automatizados de informação e que afetam sua vida diretamente.

O Brasil já perdeu a oportunidade de regulamentar esse direito quando da elaboração da Lei Geral de Proteção de Dados (LGPD).⁴³⁸ O § 3 do art. 20 da referida lei foi encaminhado à sanção com a previsão de que a decisão tomada unicamente com base em tratamento automatizado de dados pessoais e que afetem os interesses do titular de dados deveria ser revisada por pessoa natural.⁴³⁹

O dispositivo legal, no entanto, foi vetado por razões de interesse público pelo então Presidente Michel Temer, sob o argumento de que tal exigência inviabilizaria “os modelos atuais de plano de negócios de muitas empresas, notadamente das *startups* (...)”. O país perdeu com isso a oportunidade de seguir uma tendência regulatória que fora inaugurada pela União Europeia no art. 22 do GDPR (Regulamento Geral de Proteção de Dados da União Europeia)⁴⁴⁰, reduzindo a proteção aos titulares de dados.

⁴³⁷ O critério usualmente referenciado tem sido a quantidade de usuários com conta ativa em cada plataforma. A lei alemã, como já indicado, considera suas regras aplicáveis apenas às redes sociais com mais de dois milhões de usuários registrados no país. No Brasil, o PL nº 2630/2020 tem trabalhado com um referencial de dez milhões de usuários ativos, em sua última versão (art. 2º).

⁴³⁸ O histórico do debate legislativo sobre o tema é sucintamente apresentado em artigo do Professor Marco Aurélio Marraffon, em conjunto com Filipe Medon, publicado no site Conjur. A respeito, confira-se: MARRAFON, Marco Aurélio; MEDON, Filipe. Importância da revisão humana das decisões automatizadas na Lei Geral de Proteção de Dados. *Consultor Jurídico*, [s.l.], 9 set. 2019. Disponível em: https://www.conjur.com.br/2019-set-09/constituicao-poder-importancia-revisao-humana-decisoes-automatizadas-igpd#_ftn15. Acesso em: 14 nov. 2022.

⁴³⁹ A íntegra da mensagem de veto pode ser conferida em: http://www.planalto.gov.br/ccivil_03/_Ato2019-2022/2019/Msg/VEP/VEP-288.htm. Acesso em 14 nov. 2022.

⁴⁴⁰ Consoante a regra geral, “o titular dos dados tem o direito de não ficar sujeito a nenhuma decisão tomada exclusivamente com base no tratamento automatizado, incluindo a definição de perfis, que produza efeitos na sua esfera jurídica ou que o afete significativamente de forma similar” (Cf. texto completo em: <https://gdpr.algolia.com/gdpr-article-22>. Acesso em 14 nov. 2022).

O argumento que embasa um direito à justificação é usualmente a superação da opacidade do uso de mecanismos tecnológicos, que afetem oportunidade de terceiros, possibilitando que a decisão algorítmica seja minimamente inteligível⁴⁴¹.

No caso de moderação de conteúdo, o problema é diverso, mas ainda relevante. A maioria dos sistemas automatizados utilizados são relativamente simples e não há propriamente um risco de ininteligibilidade dos critérios que orientam a moderação automatizada, mesmo quando equivocada.⁴⁴² Como já se expôs, as decisões automatizadas de moderação tomam por base, em regra, o enquadramento do caso analisado a uma lista pretérita de casos negativos com exemplos de nudez, terrorismo, violência, casos de racismo, entre outros.

O principal problema desta metodologia é que por trabalhar com uma lógica de categorização — não raro — acaba por falhar na identificação de contextos, ampliando a insegurança jurídica e o risco de *chilling effect*.⁴⁴³

O direito a uma revisão humana destas decisões automatizadas torna-se, portanto, essencial para a correção de tais falhas, viabilizando o aprimoramento destes mecanismos automatizados de decisões por meio da incorporação de novos contextos ao parâmetro decisório do sistema automatizado.

Para ilustrar melhor o ponto, um exemplo. O Conselho de Supervisão (*Oversight Board*) já anulou decisão da empresa Meta de supressão do Instagram de conteúdo com nudez que fora publicizado com o intuito de promover campanha do “outubro rosa” no Brasil (Caso 2020-004-IG-BR). A decisão, além de anulada, levou à expedição de recomendação da necessidade de se incluir uma exceção aos mecanismos de moderação automatizada que vedam a promoção de conteúdo com nudez para as hipóteses de “promoção de campanha de conscientização contra o câncer”, e de revisão das Diretrizes de Comunidade do Instagram a fim de esclarecer que a proibição à nudez, na plataforma, não é absoluta. Essa recomendação foi de fato seguida pela rede social, que hoje traz exceções à vedação à nudez em suas diretrizes de comunidade.⁴⁴⁴ Sem a revisão humana, dificilmente a mudança dos parâmetros de sistemas automatizados teria ocorrido de forma pública, como no caso.

⁴⁴¹ O intuito do GDPR foi criar um conjunto de normas que imponham processos de transparência, responsabilidade e supervisão ao uso de algoritmos para tomada de decisões que afetem seres humanos. Para aprofundamento sobre o ponto, cf. KAMINSKI, Margot E. The right to explain, explained. *Berkeley Technology Law Journal*, [s.l.], v. 34, n. 189, 2019.

⁴⁴² Para usar um termo de Frank Pasquale, a “caixa preta” das redes sociais consiste principalmente nos critérios que orientam as decisões de perfilhamento que embasam o encaminhamento de publicidade direcionada aos usuários.

⁴⁴³ A limitação dos mecanismos automatizados para compreensão de contextos é fato confessado pelos provedores de redes sociais nos relatórios apresentados à autoridade alemã com fundamento na *NetzDG*. Vide nota de rodapé nº 377.

⁴⁴⁴ Veja-se: <https://pt-br.facebook.com/help/instagram/477434105621119>. Acesso em 14 nov. 2022.

Em suma, as plataformas possuem ampla autonomia para definir quais condutas e comportamentos são vedados em suas páginas à luz dos seus termos de uso e normas comunitárias, o que viabiliza a criação de procedimentos próprios para revisão. O direito ao devido processo legal demanda, contudo, transparência na aplicação das normas comunitárias e na criação de condutas passíveis de punição, regras de exceção e protocolos de crise, sem olvidar que o procedimento de revisão não dispensa ao menos uma instância de ação humana, especialmente nos casos de tomada de decisões de moderação por sistemas automatizados.⁴⁴⁵

4.2.3 A proteção à boa-fé dos usuários

O dever de proteção à boa-fé dos usuários se inspira em uma constatação de que as redes sociais não podem atuar de forma contraditória com os objetivos e missões institucionais que anunciam em seus atos constitutivos. Exige-se, assim, que o design de funcionamento destas plataformas também seja construído de modo a mitigar os impactos produzidos pelas externalidades algorítmicas.

Com efeito, tais instituições devem agir como instituições intermediárias dignas de confiança cumprindo os objetivos apresentados em suas normas comunitárias e termos de uso, quais sejam, de viabilizar a construção de uma esfera pública segura, vibrante, democrática e difusora de conhecimento⁴⁴⁶. Em contrapartida, as redes sociais não podem usar dados de seus usuários para manipulá-los e explorar vieses de forma ardil, tampouco ser conivente ou contribuir para a prática de ilicitudes em suas plataformas.

Há alguns fundamentos para sustentar a existência de tal dever. O primeiro é a supracitada ideia de eficácia horizontal dos direitos fundamentais, que exige das plataformas de redes sociais boa-fé na interação com a base de usuários com os quais mantém contrato, o que implica transparência quanto aos riscos produzidos pela desinformação e outras espécies de discursos danosos propagados em suas páginas.

⁴⁴⁵ Rebeca Tushnet defende, ainda, que o devido processo legal requer a adoção de procedimentos que criem mecanismos de empoderamento aos usuários destes serviços, tanto sob o ponto de vista substancial quanto procedimental, diluindo o controle centralizado da atividade de moderação por parte dos órgãos de cúpula destas plataformas. Cf. TUSHNET, Rebecca. Power without responsibility: intermediaries and the first amendment. *The George Washington Law Review*, [s.l.], v. 76, 2008. Disponível em: <https://www.gwlr.org/wp-content/uploads/2012/08/76-4-Tushnet.pdf>. Acesso em: 14 nov. 2022.

⁴⁴⁶ BALKIN, Jack. Information Fiduciaries and First Amendment. *UC Davies Law Review*, [s.l.], v. 4, n. 49, 2016.

O segundo é a necessidade de proteção da própria liberdade de expressão, não só na sua acepção meramente subjetiva (o direito do usuário de se expressar), mas na sua dimensão objetiva, enquanto garantia constitutiva da democracia, que confere ao Estado o dever de fomentar tanto quanto possível condições de ampla participação, reflexividade e reciprocidade de todos os cidadãos nas plataformas sociais. Isto implica tanto a promoção de arquiteturas que viabilizem o diálogo, quanto a restrição àquelas que acirrem o conflito na sociedade.

O terceiro argumento decorre do princípio da dignidade da pessoa humana (art. 1º, inciso III e art. 170, *caput* da CRFB), na condição de limite legítimo imposto a qualquer atividade econômica privada. Isto demanda que as plataformas de redes sociais sejam transparentes na relação com usuários, incrementando a consciência e o grau de informação dos cidadãos quanto à forma como exploram vieses de seus usuários e gerenciam as plataformas, de modo a lhes conferir maior autonomia.

Este é, entretanto, o mais incerto dos deveres atribuídos às redes sociais. Ainda que o conhecimento sobre os critérios que norteiam a curadoria algorítmica das plataformas tenha ampliado substancialmente nos últimos anos, muitos detalhes são desconhecidos⁴⁴⁷. Daí é possível extrair dois riscos: a instituição de um dever excessivamente abstrato pode resultar numa regulação ineficiente onde o Poder Público continue à mercê das intenções altruísticas dos agentes privados para compreender melhor como as arquiteturas das plataformas podem ser utilizadas em prejuízo aos próprios usuários. De outro, uma resposta muito contundente pode aniquilar este tipo de negócio, aproximando a intervenção estatal de um sofisticado instrumento de censura de conteúdos reputados como nocivos, tal qual o modelo chinês apelidado de *Great Firewall of China*.

A resposta está em um caminho intermediário. Mesmo que uma regulação exclusivamente principiológica⁴⁴⁸ não seja recomendável por traduzir menor poder coercitivo a um setor que historicamente tem se mostrado pouco responsivo à proteção de direitos de seus usuários, em alguma medida se faz necessária para viabilizar futuras revisões de seu alcance conforme novos conhecimentos sejam adquiridos e, ainda, para evitar que as plataformas

⁴⁴⁷ O Facebook, por exemplo, alega usar mais de 100.000 fatores personalizados para sua análise de curadoria algorítmica, embora já se saiba que a popularidade do conteúdo e a proximidade do usuário que o compartilha sejam os elementos com maior peso nesta avaliação. Cf. SUMPTER, David. *Dominado pelos números. Do Facebook e Google às fake news. Os algoritmos que controlam nossas vidas*. São Paulo: Bertrand, 2019, p. 151-153.

⁴⁴⁸ Por princípios, adota-se o conceito alexyano que as define como normas que estabeleçam que algo seja realizado na maior medida do possível, diante das possibilidades existentes. Para aprofundamento, cf. ALEXY, Robert. *Teoria dos Direitos Fundamentais*. 2. ed. Trad. Virgílio Afonso da Silva. São Paulo: Malheiros, 2017.

fiquem sobrecarregadas com exigências excessivamente burocráticas por parte do Poder Público.

Neste sentido, a construção de princípios que impliquem valorização da construção de arquiteturas e *designs* que incentivem uma esfera plural, segura e dialógica, reduzindo o impacto de bolhas informacionais — uma espécie de princípio da liberdade de expressão pelo *design (freedom of expression by design)*⁴⁴⁹ — parece ser um compromisso de partida adequado para uma legislação que pretenda regular o dever das plataformas de proteção às legítimas expectativas que criou em seus usuários.⁴⁵⁰

A maior concretização deste dever geral dependerá de alguns fatores. Primeiro, é preciso que os relatórios de transparência a que aludimos alguns tópicos acima tragam informações relevantes sobre a estruturação da arquitetura das plataformas, especialmente com indicação dos riscos evidenciados na utilização de serviços de perfilhamento e publicidade direcionada. Segundo, com a maior compreensão dos limites de cada um dos sistemas de curadoria e moderação destas plataformas, será necessário um esforço conjunto não só dos programadores das principais redes, mas também da sociedade civil e de setores de pesquisas para idealizar arquiteturas capazes de mitigar os danos decorrentes de ferramentas de *microtargeting* e afins, avaliando em cada caso aquelas que se mostraram mais ou menos eficientes.

Ganha especial ênfase, neste campo, os *insights* das ciências comportamentais para construção de *nudges*. Os *nudges*, ou cutucões são mecanismos de criação de arquiteturas de escolhas que dão um “empurrãozinho” à adoção de decisões em um determinado sentido, sem privar o indivíduo de sua liberdade de escolha. Quando utilizados pelo Estado, o objetivo destas arquiteturas deve ser a conformidade com os fins constitucionalmente estabelecidos⁴⁵¹; no caso, a preservação da dimensão objetiva da liberdade de expressão (a criação de um livre mercado de ideias e uma esfera pública reflexiva e participativa).

⁴⁴⁹ A expressão é cunhada como uma das diretrizes gerais do relatório produzido pelo Grupo de Trabalho Transatlântico, oriundo de uma parceria entre a Universidade da Pensilvânia, *Annenberg Foundation Trust* e o Instituto de Direito da Informação da Universidade de Amsterdam, sobre moderação de conteúdo *online* e liberdade de expressão. O relatório elaborado em conjunto pelos pesquisadores Emma Llansó, Joris Van Hoboken, Paddy Leersen e Jaron Harambam, publicado em fevereiro de 2020, refere-se à necessidade de construção de um princípio análogo à ideia de *privacy by design*, que implique a criação de um dever geral por parte das plataformas de redes sociais de melhor incorporação dos ideais da liberdade de expressão e direitos humanos associados à criação de novas configurações, práticas de etiquetagem de notícias fraudulentas e sistemas de recomendação algorítmica. A íntegra do relatório se encontra disponível em: <https://www.ivir.nl/publicaties/download/AI-Llanso-Van-Hoboken-Feb-2020.pdf>. Acesso em 16 nov. 2022.

⁴⁵⁰ É passível de críticas a consagração no âmbito do PL 2630/2020 de um princípio que assegure um direito a uma “visão de mundo pessoal” (art. 3º, III), que poderia justificar a manutenção de arquiteturas nas plataformas que se orientem pela “personalização” do conteúdo disponibilizado ao usuário.

⁴⁵¹ Cf. SUNSTEIN, Cass. *Why nudges? The Politics of Libertarian Paternalism*. New Haven, Londres: Yale Press, 2015.

Alguns modelos de *nudges* vêm demonstrando alguns resultados relevantes no âmbito de construção de novas arquiteturas para plataformas de redes sociais.

A etiquetagem de informações veiculadas nas mídias sociais, conhecida como alertas de *fact checking* talvez seja a mais famosa das iniciativas. Trata-se de mecanismo que começou a ser utilizado, após forte pressões sociais subsequentes ao escândalo *Cambridge Analytica*, em redes sociais como Facebook, Twitter e Youtube, desde 2018, com especial repercussão desde 2020, em razão da pandemia provocada pelo novo coronavírus e as eleições norte-americanas.

A etiquetagem de uma informação como falsa ou descontextualizada decorre de uma avaliação formulada por esses provedores após análise de notícias ou declarações por agências de *fact-checking*⁴⁵². A partir da avaliação por essas terceiras partes, os provedores de redes sociais poderão rotular a mensagem como potencialmente falsa, reduzindo sua circulação nas redes ou desmonetizando o conteúdo, ao invés de meramente censurar ou derrubá-lo.

Para assegurar a credibilidade da atuação dessas agências, os principais provedores de redes sociais têm buscado fechar parcerias com agências de *fact-checking* cujas práticas, serviços e métodos sejam reconhecidos por alguma instituição, que lhe confira um certificado de reputação.⁴⁵³

A etiquetagem, mais do que apostar na censura do conteúdo falacioso proposto, alerta ao usuário sobre ausência de credibilidade para a informação. Apela, assim, ao sistema 2 para que reflita sobre a veracidade da informação antes de divulgá-la e reproduzi-la acriticamente pelas redes sociais. A lógica não é muito distinta da utilizada em caso de eventual declaração falsa ou fraudulenta feita perante um veículo tradicional de imprensa. Usualmente, após essas manifestações, é comum que o editor (em jornais ou revistas impressos) ou jornalistas (nos canais televisivos) teçam breves comentários sobre a falta de comprovação do afirmado. Esse filtro alerta o receptor da notícia quanto à necessidade de avaliar com cuidado o conteúdo que fora apresentado sem implicar negativa à circulação da informação.

Veja-se que a iniciativa não impede o direito de manifestação do usuário, mas amplia o conhecimento sobre os fatos, na medida em que torna claro que aquela afirmação, tal como apresentada, não é corroborada por inúmeras fontes, agências e canais de mídias de matizes ideológicos distintos.

⁴⁵² A criação de parcerias com agências de *fact-checking* faz parte de um projeto de resposta do Facebook às demandas sociais de maior controle de qualidade sobre as notícias circuladas nas redes, denominado *Third Party Fact Checker Project*, iniciado em 2018.

⁴⁵³ Um exemplo é a *International Fact-Checking Network (IFCN)*, instituição criada em 2015 com o intuito de reunir agências de *fact-checking* em todo mundo, sob um código de ética e princípios de atuação que garanta um padrão mínimo de qualidade para o funcionamento dessas agências.

Há discussões acerca da efetividade da medida⁴⁵⁴. Como já antecipado, pelo caráter inovador do uso de *nudges* para redirecionamento das plataformas das redes a fins democráticos, são poucos os estudos sobre o tema. Há, contudo, algumas pesquisas empíricas já realizadas, que trazem evidências de que expor os usuários a alertas de *fact-checking* reduz substancialmente a probabilidade de compartilhamento de notícias se comparado àqueles que não foram expostos a esse tipo de alerta,⁴⁵⁵ mesmo para quem alimenta certo ceticismo quanto a fontes tradicionais da mídia.⁴⁵⁶

No mais, é difícil questionar a juridicidade de tais ferramentas. Como já exposto, não se trata de censura, já que as mensagens de pessoas, agentes públicos ou de fontes das mídias (especialmente as pouco conhecidas), por regra, permanecerão com seu conteúdo imodificável nas redes. O alerta apenas sinaliza para o caráter controverso da afirmação, à luz de fatos analisados por agências de *fact-checking*. Com efeito, produz-se mais informação ao usuário. E se há mais informação, é reduzido o enviesamento produzido pelas bolhas filtros e câmaras de eco, aprimorando a formação de pensamento do cidadão.

A etiquetagem é o exemplo de *nudge* mais conhecido das redes sociais, mas não é o único. Há ainda medidas de transparência, que também agregam informações simples, porém fundamentais, ao espaço virtual. Plataformas como Facebook e Instagram admitem, por exemplo, o uso de ferramentas denominadas como conteúdos impulsionados.

O impulsionamento de conteúdo é uma ferramenta disponível nessas plataformas, que aumenta o alcance original de uma página ou conteúdo publicado. Em vez de o conteúdo impactar apenas seus contatos, ele será exibido a mais possíveis interessados, conforme análise preditiva dos algoritmos utilizados por essas plataformas. O problema é que nem sempre esses conteúdos são claramente identificados como impulsionados, o que impede o indivíduo de perceber que a popularidade daquele conteúdo é alavancada artificialmente.

⁴⁵⁴ Solução interessante sugerida por Ivar Alberto Hartmann para aumentar a eficácia das medidas de checagem de fatos é a associação deste instrumento à ferramenta da biblioteca de anúncios. O autor propõe que todas as plataformas de redes sociais sejam obrigadas a manter bibliotecas de todo o conteúdo pago ou patrocinado, incluindo metadados sobre identidade do anunciante, URL, valor pago e opções de direcionamento de usuários selecionados pelo anunciante. A ideia é que a plataforma indique quais conteúdos daquela biblioteca já foram objeto de checagem de fatos, o que viabilizaria uma visão mais ampla sobre o ecossistema da desinformação. Cf. HARTMANN, Ivar Alberto Martins. Combinando Biblioteca de Anúncios com Checagem de Fatos para aumentar a transparência sobre a desinformação. *RDP*, Brasília, v. 18, n. 99, p. 460-485, jul./set. 2021.

⁴⁵⁵ NEKMAT, Elmie. Nudge effect of fact-check alerts: Source influence and media skepticism on sharing of News information in social media. *Social media + Society*, [s.l.], v. 6, n. 1, 2020. No mesmo sentido, RIDKIN, Kim; KENNEY, Patrick J.; WINTERSIECK, Amanda. Liar, liar, pants on fire: how fact-checking influences citizens' reactions to negative advertising. *Political Communication*, [s.l.], v. 32, n. 1, p. 139, 2015.

⁴⁵⁶ Na verdade, o que se observou, para além da redução dos compartilhamentos, é que para os céticos quanto à mídia tradicional, eventual etiquetagem de uma notícia de uma mídia tradicional como falsa produz um efeito sobre a reputação desta mídia maior do que aquele que é observado para canais de mídia não tradicionais.

Esses conteúdos são inteligentemente explorados para fins políticos ou econômicos e suscitam o efeito manada (também um viés). Muitos usuários são surpreendidos com um conteúdo cuja popularidade é artificial, mas não tem plena consciência disso e afetos pelo poder impressionista da suposta popularidade, passam a seguir a tendência daquele comentário, aumentando sua projeção.

A transparência quanto aos conteúdos impulsionados na rede e a identificação de quem pagou por este serviço é relevante para esclarecer o caráter propagandístico do conteúdo veiculado e reduzir o impacto que este tipo de informação possa ter sobre nossos vieses, em especial o de efeito manada, que pode levar usuários a confiarem acriticamente na popularidade de um determinado posicionamento político.⁴⁵⁷ O Projeto de Lei nº 2630/2020, em trâmite na Câmara dos Deputados, que visa regulamentar a Liberdade, Transparência e Responsabilidade nas redes sociais, abre espaço para o uso deste *nudge*, ao exigir transparência quanto à identificação de conteúdo patrocinado e impulsionado nas redes (arts. 16 a 21). Apenas em setembro de 2020, o Facebook passou a disponibilizar ferramentas de transparência para anúncios políticos em suas redes.⁴⁵⁸

Na versão mais recente da obra de Cass Sunstein, já no apogeu da era das redes sociais, o autor sugere ainda um peculiar *nudge* para propiciar maiores pontos de contato entre pessoas, visões de mundo, grupos e fóruns no âmbito das redes sociais. São os *serendipity buttons* (ou botões de aleatoriedade).⁴⁵⁹

O *feed* de notícias de diversas redes sociais, notadamente do Facebook, Instagram, Twitter e Youtube, é comumente abastecido com informações de seus amigos, conteúdos patrocinados ou impulsionados e notícias. Esse painel, por óbvio, é adaptado às preferências do usuário, por meio de análises produzidas pelos algoritmos destas redes. A sugestão do professor para ampliar o contato com outros pontos de vista e evitar a polarização extrema é manter como configuração padrão nessas redes a “aleatoriedade”, assegurando ao usuário o direito de facilmente alterar essa configuração — desligando o *serendipity button* — para só receber notícias e conteúdo personalizados.

⁴⁵⁷ O *Digital Services Act* (DSA) trouxe uma série de inovações sobre este ponto. Em seu art. 30, estabeleceu às plataformas o dever de as redes sociais de “muito grande dimensão” preverem para cada publicidade veiculada, de forma explícita e não ambígua: a) que se trata de anúncio publicitário; b) a pessoa ou entidade em nome de quem o anúncio é veiculado; c) informações sobre os principais parâmetros utilizados para determinar o destinatário daquele conteúdo, indicando se ele foi distribuído por perfilhamento.

⁴⁵⁸ Informação disponível em: FACEBOOK disponibiliza ferramentas de transparência para anúncios políticos no Brasil. *GI*, [s.l.], 22 set. 2020. Disponível em: <https://g1.globo.com/economia/tecnologia/noticia/2020/09/22/facebook-disponibiliza-ferramentas-de-transparencia-para-anuncios-politicos-no-brasil.ghtml>. Acesso em: 16 nov. 2022.

⁴⁵⁹ SUNSTEIN, Cass. *#Republic: Divided Democracy in the Age of Social Media*. Nova Jersey: Princeton Press, 2017, p. 232-233.

Trata-se de uma estratégia regulatória que busca construir uma arquitetura de escolha voltada à pluralidade e à multiplicidade de pontos de vista, valendo-se do viés da inércia ou do *status quo* (tendemos a seguir a configuração padrão, pois nosso sistema 2 é preguiçoso)⁴⁶⁰.

A sugestão parece não violar a liberdade do indivíduo de dispor sobre quais notícias gostaria de receber, mas estabelece como padrão uma configuração que, potencialmente, reduziria os impactos negativos de só ter acesso àquilo que confirma pré-compreensões enraizadas. É difícil imaginar, contudo, que tal mecanismo seja voluntariamente construído pelas redes sociais — ao contrário das demais medidas informativas —, pois focar na ampliação de informações de matizes ideológicas distintas afeta, em alguma medida, o modelo de negócio fundado na economia de atenção. Se imposto pelo Estado, é possível antever questionamentos judiciais à luz do princípio da livre-iniciativa (art. 170, *caput* da Constituição da República). A manutenção do modelo pode vir a demandar comprovação, portanto, de sua adequação à finalidade pretendida (evidências empíricas dos efeitos positivos) e, especialmente, a ausência de outro meio menos oneroso para atingir resultados similares.⁴⁶¹

A proteção à boa-fé dos usuários exige das plataformas de redes sociais um dever geral de informar os riscos decorrentes de seus sistemas de recomendação algorítmico e o compromisso com a construção de *designs* e ferramentas que minorem estes riscos, em geral, por meio do estímulo à construção de arquiteturas que facilitem a construção de uma esfera pública efetivamente democrática.

A evolução do conhecimento viabiliza, ainda, a consolidação de deveres específicos capazes de tornar a governança algorítmica dos agentes privados mais responsiva aos riscos decorrente de suas atividades, tais como: a) *identificação clara aos usuários dos conteúdos pagos, impulsionados ou patrocinados no âmbito destas plataformas*; b) *manutenção de biblioteca de anúncios com informações gerais sobre anunciantes e critérios de perfilamento utilizados*; c) *incentivo à construção de parcerias com institutos de fact checking para*

⁴⁶⁰ Este tipo de viés é o que fundamenta as regras de adesão automática a regimes de previdência complementar, a exemplo do que consta no art. 4º da Lei Federal nº 13.183/2015. O viés comportamental da regra foi objeto de defesa do Laboratório de Regulação Econômica da UERJ (UERJ-REG), no bojo da ADI 5502/DF, ajuizada pelo Partido Socialismo e Liberdade (PSOL). A íntegra da manifestação pode ser conferida em: LABORATÓRIO DE REGULAÇÃO ECONÔMICA. Adesão automática à previdência complementar é constitucional. *Jota*, [s.l.], 6 out. 2017. Disponível em: <https://www.jota.info/opiniao-e-analise/colunas/reg/adesao-automatica-a-previdencia-complementar-06102017>. Acesso em: 16 nov. 2022.

⁴⁶¹ O *Digital Services Act* (DSA) trouxe, em seu art. 29, uma previsão que possibilita uma solução necessária (menos onerosa) para reduzir a influência em seus sistemas de recomendação. De um lado, passou a exigir que a rede social informe de forma clara, acessível e facilmente compreensível, os principais parâmetros que orientam o funcionamento deste sistema, incluindo ao menos uma opção de configuração que não se baseie na definição de perfis. De outro, exige que esta opção de configuração alternativa esteja facilmente acessível ao usuário, de modo a permiti-lo que altere a qualquer momento a sua opção preferida.

*aprimoramento de medidas de moderação de etiquetagem, desmonetização ou redução da distribuição de peças de informação falsas ou descontextualizadas; d) elaboração de configurações que prestigiem a “aleatoriedade” (em detrimento de conteúdos exclusivamente personalizados) e que sejam inseridas, ao menos, como opções alternativas facilmente disponíveis e acessíveis aos usuários.*⁴⁶²

4.3 Da responsabilidade civil à responsabilidade administrativa: reflexões sobre o papel do Poder Judiciário em um contexto corregulatório

O protagonismo do controle judicial da desinformação e da atuação das plataformas no combate de conteúdos reputados como danosos à democracia tem sido uma tendência dos últimos anos no Brasil. Em meio a inúmeras decisões polêmicas, destaca-se o Inquérito Policial nº 4781, instaurado pelo então Presidente do Supremo Tribunal Federal (Min. Dias Toffoli), por meio da Portaria GP nº 69/2019, para apuração da disseminação de notícias fraudulentas (*fake news*), que atinjam a honra e segurança do Supremo Tribunal Federal e de seus Ministros.

⁴⁶² Outro ponto relevante sobre o *design* destas plataformas, mas que por limitações do âmbito da pesquisa acabou por não ser aprofundado, é a necessidade de um compromisso com a identificação de perfis anônimos, que se valham destas redes para prática de crimes. Neste ponto, é um imperativo que tais plataformas adotem uma postura colaborativa com autoridades judiciais e administrativas investigativas. Não se coaduna, contudo, com projetos de lei que buscam alterar o Marco Civil da *Internet* para exigir que o cadastro de usuários nestas plataformas dependa de indicação de identidade e CPF (esta espécie de regra consta em cerca de 14% de todos os projetos de lei que buscam alterar o Marco Civil da *Internet*). Cf. HARTMANN, Ivar Alberto Martins; MONTEIRO, Júlia Iunes. Fake News no Contexto de Pandemia e Emergência Social: os Deveres e Responsabilidades das Plataformas de Redes Sociais na Moderação de Conteúdo Online: entre a Teoria e as Proposições Legislativas. *RDP*, Brasília, v. 17, n. 94, p. 388-414, jul./ago. 2020). Nosso entendimento se aproxima, ainda que parcialmente, do defendido por André Farah — a partir de uma interpretação teleológica de diversos precedentes do Supremo Tribunal Federal — que sustenta a incompatibilidade de exigências de identificação civil, por haver no Marco Civil da *Internet* instrumentos proporcionais capazes de garantir um anonimato rastreável e apto a afastar a incidência da regra constante no inciso IV do art. 5º da CF. Explica-se. No âmbito da *internet*, como há meios de se identificar usuários a partir dos registros de acesso a aplicações da *internet* (art. 5º, inciso VIII do MCI), cabendo ao provedor de aplicação da *internet* o dever de armazená-las (art. 15), entende-se que não há um anonimato absoluto que aniquile a regra do art. 5º, inciso IV da CF. Como o STF tem interpretado a vedação ao anonimato como um instrumento necessário para evitar a ausência de responsabilização por abuso no exercício da liberdade de se expressar, as regras do MCI são ao menos *a priori* uma ferramenta proporcional para apuração da responsabilidade daqueles que usam o aparente escudo do anonimato para praticar atos ilícitos. A respeito, cf. FARAH, André. *Liberdade de expressão e remoção de conteúdo*. Rio de Janeiro: Lumen Juris, 2018, p. 156-163. A imposição estatal de apresentação de identidade ou CPF para eventual ingresso em plataforma de rede social traz ainda outros problemas, que não podem ser ignorados quando da proposição destas regras: (I) a identificação por documentos pessoais pode ser utilizada como instrumento de perseguição de opositores, produzindo um efeito censor; (II) confere às plataformas de redes sociais ainda mais controle sobre dados sensíveis dos particulares. Por esta razão — a nosso entender —, a regra constitucional do inciso IV do art. 5º da CF somente restaria violada por plataformas de redes sociais que não assegurem qualquer sistema de verificação de responsabilidade, inclusive quanto a metadados de registro de acesso a aplicações da *internet* (uma espécie de anonimato absoluto que evite futura responsabilização por atos ilícitos), estando, portanto, também em desacordo com a normativa geral do MCI.

A instauração do inquérito se deu com base em uma controvertida interpretação do art. 43 do RISTF⁴⁶³, e teve como fundamento fático matéria jornalística veiculada no site O Antagonista e na Revista Crusoé, intitulada “Amigo do amigo do meu pai”, no qual era citado o Min. Dias Toffoli.

O Min. Alexandre de Moraes, então sorteado relator do inquérito, determinou em 13 de abril de 2019 a supressão imediata da referida matéria, sob pena de instituição de multa diária no valor de R\$ 100.000,00 (cem mil reais). Na visão do Min. Relator, haveria claro abuso do exercício da liberdade de informação, no caso. A decisão foi objeto de crítica por diversos setores da sociedade civil, por representar uma espécie de “censura”, implicar abusos por parte do Poder Judiciário, que passaria a atuar como investigador e julgador de uma mesma causa e, ainda, por se tratar de um inquérito sem objeto específico.

Dias após a primeira decisão, o Min. Alexandre de Moraes a revogou sob o argumento de que o conteúdo da delação premiada que embasou a matéria jornalística de fato não era adulterado, mas manteve em vigor o inquérito por considerá-lo um instrumento que não configuraria hipótese de censura prévia, mas sim de responsabilização posterior daqueles que se escudam na liberdade de expressão para praticar ilícitos.⁴⁶⁴

O partido REDE SUSTENTABILIDADE ajuizou, então, a ADPF nº 572, requerendo liminarmente a suspensão da Portaria do Presidente do STF que instaurou o referido inquérito.⁴⁶⁵

Em maio de 2021, o inquérito fora considerado válido pelo Supremo Tribunal Federal, que votou para mantê-lo, embora tenha limitado expressamente seu objeto a “*manifestações que denotam risco efetivo à independência do Poder Judiciário, pela via de ameaça de seus membros e, assim, risco aos Poderes instituídos, ao Estado de Direito e à democracia*”. Foram excluídos expressamente do seu escopo matérias jornalísticas, compartilhamento de outras

⁴⁶³ RISTF, art. 43. “Ocorrendo infração à lei penal na sede ou dependência do Tribunal, o Presidente instaurará inquérito, se envolver autoridade ou pessoa sujeita à sua jurisdição ou delegará esta atribuição a outro Ministro”.

⁴⁶⁴ Informação disponível em: BOMFIM Camila. Alexandre de Moraes revoga decisão que censurou reportagens de ‘Crusoé’ e ‘O Antagonista’. *GI*, [s.l.], 18 abr. 2019. Disponível em: <https://g1.globo.com/politica/noticia/2019/04/18/alexandre-de-moraes-revoga-decisao-que-censurou-reportagens-de-crusoe-e-antagonista.ghtml>. Acesso em: 24 nov. 2022.

⁴⁶⁵ Após a propositura da ação e posterior mudança do objeto do inquérito (que passou a se destinar ao combate da atuação do denominado “Gabinete do Ódio”), houve pedido de desistência da ação por parte do partido autor. Em observância à jurisprudência da Suprema Corte, a desistência não foi aceita pelo relator ante caráter objetivo deste tipo de ação.

manifestações (inclusive pessoais) na *internet*, anônimas ou não, desde que não integrem esquemas de financiamento e divulgação em massa nas redes sociais.⁴⁶⁶

Esta, certamente, foi uma das mais controvertidas atuações do Poder Judiciário nos últimos anos, mas não foi única, sendo acompanhada por outras decisões polêmicas, dentre as quais cumpre mencionar: condenação do ex-Deputado Daniel Silveira, no bojo da Ação Penal nº 1044, ajuizada pelo Ministério Público Federal, por divulgação de vídeos em suas redes sociais nos quais aparece fazendo ameaças ao Supremo Tribunal Federal e seus Ministros⁴⁶⁷; a cassação do mandato do ex-Deputado Fernando Francischini pelo Tribunal Superior Eleitoral, por divulgar notícias falsas contra o sistema eletrônico de votação (condenação posteriormente ratificada pelo Supremo Tribunal Federal)⁴⁶⁸; e a decisão que determinou a prisão preventiva do ex-Deputado Roberto Jefferson, por atacar integrantes de instituições públicas, desacreditar o processo eleitoral brasileiro e reforçar o discurso de polarização e de ódio.⁴⁶⁹

Em meio à escalada de desinformação das eleições presidenciais em 2022, o Tribunal Superior Eleitoral editou também polêmica resolução, que ampliou seu poder de combate à desinformação durante o segundo turno eleitoral (Resolução nº 23.714/2022). Dentre as normas da resolução, destaca-se os exíguos prazos conferidos para derrubada de conteúdo fraudulento, produtor de desordens informacionais (até 2 horas, como regra)⁴⁷⁰ e a ampliação dos poderes do Presidente do TSE para, de ofício, poder determinar a extensão da decisão colegiada do tribunal sobre desinformação, para outras situações com idêntico conteúdo, indicando as

⁴⁶⁶ STF, ADPF nº 572/DF, Min. Rel. Edson Fachin, Plenário, julgado em 07/05/2021. No mesmo julgamento se assentou a constitucionalidade da interpretação conferida ao art. 43 do RISTF pelo Min. Dias Toffoli, quando deflagrou o inquérito. Segundo este entendimento, o art. 43 autorizaria o Supremo Tribunal Federal a instaurar inquérito sempre que o Tribunal ou seus membros fossem ameaçados mesmo fora de suas dependências físicas, tendo em vista que sua jurisdição é una se estendendo por todo território nacional.

⁴⁶⁷ AP 1044/DF, Min. Rel. Alexandre de Moraes, Plenário, Julgado em 20/04/2022.

⁴⁶⁸ A cassação do TSE foi ratificada pelo STF na TPA 39/DF, rel. p/o acórdão Min. Edson Fachin, Segunda Turma, julgado em 07/06/2022. A divulgação de desinformação por parte do então deputado fora reconhecida como espécie de abuso de autoridade e, por analogia, uso indevido dos meios de comunicação social, nos termos da Lei Complementar nº 64/90 – Lei das Inelegibilidades.

⁴⁶⁹ Informação disponível em: CALEGARI, Luiza; GOES, Severino. Roberto Jefferson, aliado de Bolsonaro, é preso por ordem do ministro Alexandre, do STF. *Consultor Jurídico*, [s.l.], 13 ago. 2021. Disponível em: <https://www.conjur.com.br/2021-ago-13/alexandre-manda-prender-roberto-jefferson-aliado-bolsonaro>. Acesso em: 24 nov. 2022.

⁴⁷⁰ Art. 2º É vedada, nos termos do Código Eleitoral, a divulgação ou compartilhamento de fatos sabidamente inverídicos ou gravemente descontextualizados que atinjam a integridade do processo eleitoral, inclusive os processos de votação, apuração e totalização de votos.

§ 1º Verificada a hipótese prevista no caput, o Tribunal Superior Eleitoral, em decisão fundamentada, determinará às plataformas a imediata remoção da URL, URI ou URN, sob pena de multa de R\$ 100.000,00 (cem mil reais) a R\$ 150.000,00 (cem e cinquenta mil reais) por hora de descumprimento, a contar do término da segunda hora após o recebimento da notificação.

§ 2º Entre a antevéspera e os três dias seguintes à realização do pleito, a multa do § 1º incidirá a partir do término da primeira hora após o recebimento da notificação

URL's, URI's ou URN's, que deverão ser removidas.⁴⁷¹ A resolução foi objeto de questionamento imediato pelo Procurador-Geral da República⁴⁷², mas teve sua constitucionalidade chancelada pelo Supremo Tribunal Federal em caráter cautelar, na ADI 7261, de Relatoria do Ministro Edson Fachin.⁴⁷³

Dois são os pontos comuns entre as decisões citadas. Sob o ponto de vista material, todas se valem, expressa ou implicitamente, de uma concepção *militante* de democracia, no qual a atuação do STF se legitimaria por um propósito de autodefesa do tribunal, essencial à preservação do regime democrático tal como preconizado pelo texto constitucional. Do ponto de vista procedimental, caracterizam-se pela ampliação dos poderes da corte (STF ou TSE) ou de seus ministros, mediante a flexibilização de garantias constitucionais procedimentais, em especial a separação de poderes e o juiz natural.

À luz de uma análise estritamente consequencialista, a atuação do Supremo poder-se-ia considerar justificável como salvaguarda de um controle de “normalidade” institucional⁴⁷⁴ em um contexto de quase ruptura alimentada pelo Chefe do Poder Executivo, mas cumpre indagar em que medida tal atuação *proativa* seria, de fato, eficiente para resguardar a legitimidade do tribunal constitucional no exercício de sua função precípua de controle aos impulsos da maioria da vez.⁴⁷⁵ Afinal, se a própria corte constitucional ignora os limites procedimentais que lhe foram impostos pelo constituinte, o que resta de jurídico na sua atuação?

A observância rigorosa das garantias constitucionais procedimentais é fundamental para preservar a autoridade moral do tribunal constitucional, resguardando a imagem de

⁴⁷¹ Art. 3º A Presidência do Tribunal Superior Eleitoral poderá determinar a extensão de decisão colegiada proferida pelo Plenário do Tribunal sobre desinformação, para outras situações com idênticos conteúdos, sob pena de aplicação da multa prevista no art. 2º, inclusive nos casos de sucessivas replicações pelo provedor de conteúdo ou de aplicações.

§ 1º Na hipótese do caput, a Presidência do Tribunal Superior Eleitoral apontará, em despacho, as URLs, URIs ou URNs com idêntico conteúdo que deverão ser removidos.

§ 2º A multa imposta em decisão complementar, proferida na forma deste artigo, não substitui a multa aplicada na decisão original

⁴⁷² PGR questiona resolução editada pelo TSE a menos de duas semanas do segundo turno. *MPF*, [s.l.], 21 out. 2022. Disponível em: <https://www.mpf.mp.br/pgr/noticias-pgr/pgr-questiona-resolucao-editada-pelo-tse-ha-menos-de-duas-semanas-do-segundo-turno>. Acesso em: 24 nov. 2022.

⁴⁷³ ADI 7261-MC, Min. Rel. Edson Fachin, Plenário, julgado em 26/10/2022.

⁴⁷⁴ A expressão consta em artigo do professor José Vicente Santos de Mendonça disponível em: MENDONÇA, José Vicente Santos de. O novo papel do Supremo é o controle de normalidade. *Jota*, [s.l.], 14 jun. 2022. Disponível em: <https://www.jota.info/opiniao-e-analise/colunas/publicistas/o-novo-papel-do-supremo-e-o-controle-de-normalidade-14062022>. Acesso em: 24 nov. 2022.

⁴⁷⁵ Como bem ensina Rodrigo Brandão, o apoio difuso à corte constitucional é o que define o quanto o tribunal “pode andar com suas próprias pernas”, distanciando-se eventualmente da opinião pública. Cf. BRANDÃO, Rodrigo. *Supremacia Judicial versus Diálogos Constitucionais*: a quem cabe a última palavra sobre o sentido da Constituição. Rio de Janeiro: Lumen Juris, 2018, p. 318.

imparcialidade que sustenta sua legitimidade perante a sociedade.⁴⁷⁶ Um tribunal que flexibiliza, quando não aniquila estas garantias por meio de conduções de inquéritos no qual figura de investigador e julgador se confundem, coloca-se em posição mais vulnerável mediante pretensões antidemocráticas de parte da sociedade.

Sob o ponto de vista material, não se nega que a concepção democrática da Constituição é militante. Tal ponto é defendido nesta dissertação a partir de uma interpretação teleológica do art. 17 da Constituição da República. Sustentou-se, contudo, que a definição de comportamento ou discurso antidemocrático seja cuidadosamente construído por lei. No limite, poder-se-ia até defender a aplicação direta do art. 17 pelo Supremo Tribunal Federal, desde que ao menos a Corte estabelecesse parâmetros objetivos que permitissem identificar o que seria um discurso ou ato antidemocrático, mitigando os riscos de *chilling effect*. Sem esta preocupação com um mínimo de objetividade na definição dos discursos ilícitos, o Tribunal Constitucional flerta com a indesejada alcunha de censor. O discurso antidemocrático seria apenas a informação falsa e dolosamente produzida sobre as urnas? O conceito englobaria uma opinião favorável ao fechamento da Corte constitucional? O pedido de retorno do AI-5 poderia ser considerado antidemocrático? O ataque à reputação dos Ministros do Supremo é ofensivo à democracia? E a incitação à violência contra agentes políticos? Ou, ainda, é possível que tudo isso seja antidemocrático? Ou somente parte destes discursos? Infelizmente, não há resposta para tais questões, visto que o alcance da expressão nunca foi delimitado pelo STF.

É certo que a flexibilização de garantias procedimentais pelo STF é, em último caso, reflexo da tentativa de o Tribunal superar o problema da seletividade trazido no capítulo segundo. Como a atuação do Poder Judiciário é sempre limitada pela provocação de terceiros, amplia-se a capacidade investigativa da corte constitucional para viabilizar uma regulação *ad hoc*, minimamente eficiente, das estruturas dos projetos de disseminação e financiamento de materiais produtores de desordens informacionais. O custo desta prática tem sido o sacrifício de regras procedimentais de estatura constitucional e, por consequência, a própria credibilidade do STF.

A defesa de regime de responsabilização administrativa das redes sociais firme em um compromisso de mitigação dos impactos negativos produzidos pelas desordens informacionais

⁴⁷⁶ Mesmo para Daniel Sarmiento e João Gabriel Madeira Pontes, defensores da teoria da democracia militante, a autocontenção judicial e a garantia do devido processo legal na observância da teoria são fundamentais para assegurar a imagem de imparcialidade do Tribunal. E, citando Bastian Rijpkema, afirmam: “se o juiz resolve vagar por terreno político, ele se torna parte de uma batalha da qual só pode emergir com menos autoridade”. SARMENTO, Daniel; PONTES, João Gabriel Madeira. Democracia Militante e Imunidade Material dos Parlamentares: Limites Constitucionais aos Discursos de Deputados e Senadores. *Revista da AJURIS*, Porto Alegre, v. 47, n. 149, dez. 2020. p.88.

é, neste sentido, uma resposta necessária para direcionar respostas institucionais adequadas aos abusos cometidos sob o manto da liberdade de expressão, afastando dos tribunais o papel de definição do que é verdadeiro ou falso no âmbito das redes sociais.⁴⁷⁷

Por outro lado, isto não significa afirmar que o Poder Judiciário esteja afastado por completo do controle do funcionamento das principais plataformas de redes sociais. O caráter sistêmico do fenômeno das desordens informacionais requer, de certo, uma resposta regulatória complexa e que se viabilize, principalmente por meio de mecanismos administrativos externos de correção ou autorregulação regulada; mas, ao Poder Judiciário, sempre caberá, em último caso, a última palavra sobre higidez dos procedimentos sancionatórios eventualmente aplicados pelos órgãos/entes fiscalizatórios externos.⁴⁷⁸

A liberdade de expressão do século XXI é antes de tudo um problema tecnológico, o que não obsta reconhecer que os impactos substantivos de práticas regulatórias sobre as liberdades discursivas não possam ser avaliados pelos tribunais a quem cumpre assegurar que a ação administrativa se deu de forma proporcional e conforme a legalidade, reduzindo assim os riscos de deslize ao autoritarismo a que aludimos no primeiro capítulo.

Se, a nível sistêmico, a regulação administrativa parece ser a melhor solução, sob o ponto de vista individual, de responsabilização civil, a maioria dos autores que se debruçam sobre o tema insiste em cancelar a constitucionalidade do art. 19 do MCI sem qualquer espécie de ressalva. Assim, a doutrina nacional tem sustentado a necessidade de criação de mecanismos de responsabilização administrativa no atacado, mas defende a regra de responsabilização civil construída pelo art. 19 do MCI, no varejo.⁴⁷⁹

Usualmente, alega-se a necessidade de preservação de tal norma para afastar o efeito resfriador da regulação sobre o debate público (*chilling effect*) e por se considerar que as plataformas de redes sociais têm sido responsivas a outras espécies de pressão social para aprimoramento das suas normas e medidas de moderação.⁴⁸⁰

⁴⁷⁷ Lembrando que a hipótese regulatória aqui defendida foca mais na arquitetura das redes sociais com assunção de deveres de transparência, respeito ao devido processo legal e prestígio à boa-fé dos seus usuários, do que propriamente em regulações orientadas ao conteúdo, cujo papel seria residual.

⁴⁷⁸ Em último caso, o próprio art. 5º, inciso XXXV da Constituição Federal respalda a possibilidade de atuação posterior do Poder Judiciário.

⁴⁷⁹ Por todos, cf. BARROSO, Luna Van Brussel. *Liberdade de Expressão e Democracia na Era Digital*. Belo Horizonte: Fórum, 2021; e NITRINI, Rodrigo Vidal. *Liberdade de Expressão nas Redes Sociais*. São Paulo: Dialética, 2021.

⁴⁸⁰ Nesta linha, afirma Nitri: “(...) mesmo sob regimes legais que não implicam responsabilidade civil direta, plataformas digitais cedem a outros tipos de incentivos (...) essenciais para enfrentarem conteúdos problemáticos (ilícitos ou não)” (NITRINI, Rodrigo Vidal. Op. cit., p. 175). Diverge-se deste argumento, visto que a fiscalização social não se dá de forma contínua e estruturada. Na verdade, o que se percebe é que boa parte das reformas ocorridas no funcionamento das plataformas — incrementando transparência da sua atuação — foram frutos de

O que se sugere agora, contudo, é uma revisão parcial do dispositivo legal. A regra do art. 19 do MCI tem como principal objetivo evitar um efeito resfriador do debate na *internet*; assim, o provedor de aplicação de conteúdo só poderia ser responsabilizado por dano gerado por terceiros em duas hipóteses: (I) se após ordem judicial específica não tomar as providências necessárias no prazo assinalado pela decisão; (II) se, por sua conta e risco, adotar providências com base em seus termos de uso e tais medidas serem reputadas como excessivas ou ilegais por eventual decisão judicial posterior. Há, portanto, um claro desincentivo à aplicação de medidas de moderação por parte das plataformas, visto que a plataforma que não derruba conteúdos ilícitos danosos, de maneira autônoma, não poderá ser responsabilizada pela inércia.

A regra, no que tange especificamente as redes sociais, apresenta dois problemas. O primeiro decorre de sua incompatibilidade com a percepção atual da capacidade de influência de tais plataformas sobre o comportamento de seus usuários.

A construção da regra do art. 19 fazia sentido em um contexto no qual ainda se nutria uma visão de isenção das redes sociais quanto aos conteúdos que circulavam em suas páginas, não havendo qualquer tipo de medida de controle editorial. O que se viu, contudo, é que existe, sim, um controle do fluxo de informações por parte destas redes, viabilizado por um procedimento de curadoria algorítmica, que define quais conteúdos terão alcance amplificado ou não. Se a experiência tem comprovado que tal curadoria tem contribuído para a produção não só de desinformação, mas também de proliferação de discursos odiosos, resta evidente que há um *risco criado*⁴⁸¹ pela atividade econômica das próprias plataformas que justificaria a responsabilização civil no caso de não derrubada de conteúdos potencialmente danosos mesmo sem prévia decisão judicial.

Em segundo lugar, é preciso esclarecer que a importação da cláusula 230 do CDA se deu de forma incompleta. Como já se antecipou, a norma norte-americana criou uma isenção de responsabilidade por danos gerados por terceiros, mas também criou uma cláusula do “bom samaritano”, que respaldava as medidas de moderação ou controle exercidas por plataformas de redes sociais de boa-fé.

A importação desta passagem final não ocorreu, possivelmente, pelo justificado temor dos legisladores de que o parâmetro da “boa-fé” criasse uma insegurança jurídica desnecessária

escândalos e vazamentos ocasionais de informações sigilosas internas, que demandaram atuações paliativas de empresas do setor para assegurar uma boa reputação junto aos usuários e acionistas.

⁴⁸¹ Segundo Caio Mário da Silva Pereira, a teoria do risco criado “aumenta os encargos do agente; é, porém, mais equitativa para a vítima, que não tem de provar que o dano resultou de uma vantagem ou de um benefício obtido pelo causador do dano”, uma vez que é decorrência lógica da atividade explorada. PEREIRA, Caio Mário da Silva. *Responsabilidade civil*. 5. ed. Rio de Janeiro: Forense, 1994, p. 285.

ante nossa lamentável tradição de expandir o uso do argumento principiológico para (tentar) justificar arbitrariedades judiciais de toda ordem.⁴⁸²

Nada obstante, em um contexto normativo no qual um regime de responsabilização administrativa seja criado, parece possível se recorrer ao parâmetro da “boa-fé” ou similar, desde que se defina que a presunção de boa-fé cessa para as plataformas que descumpram os deveres estabelecidos na regulação administrativa.

O benefício legal da norma de imunidade de responsabilização civil por parte das plataformas de redes sociais seria uma espécie de *sanção premial*, tanto pela transparência da atuação de tais provedores de aplicação na condução de suas políticas comunitárias, protocolos de crise e termos de uso, quanto pelo atendimento ao compromisso de combate aos discursos ilícitos ou danosos produzidos por usuários nestas plataformas.

Entende-se que tal construção — um meio termo entre o modelo de *notice and takedown* e o de imunidade substancial das plataformas — mitiga o legítimo receio de *chilling effect*, visto que o regime de responsabilização civil ordinário seria aplicável apenas aos provedores de redes sociais claramente negligentes, que ignoram ou criam empecilho à aplicação dos parâmetros regulatórios fixados em lei. Por consequência, também se prestigia as redes sociais que têm se esforçado para se adequar às diretrizes normativas do mercado nacional e assumido, inclusive, compromissos de boa-conduta com órgãos e entes públicos.

4.4 Os limites à regulação pública das novas redes de comunicação

A regulação de novos canais de comunicação, como as redes sociais, também implica o reconhecimento de certas limitações. O que não significa afirmar que a regulação estatal é um remédio jurídico inútil, mas sim que confiar exclusivamente ao Estado a solução para todos os males produzidos na esfera digital levará a resultados certamente improdutivos e potencialmente censórios.

Para melhor sistematização do ponto, é possível identificar três grandes espécies de limites à eventual pretensão regulatória estatal: os limites de natureza *jurídica*, *institucional* e *temporal*.

⁴⁸² Neste sentido, Cf. SUNDFELD, Carlos Ari. Princípio é preguiça? In: _____. *Direito Administrativo para Céticos*. 2. ed. São Paulo: SBDP, Malheiros, 2014, p. 205-229.

O primeiro dos limites foi, em grande parte, objeto de exame ao longo da dissertação. Os limites jurídicos têm como enfoque os impactos que eventual pretensão regulatória pode ter sobre o exercício de direitos fundamentais tanto de usuários (em geral, a liberdade de expressão e privacidade) quanto das empresas que a administram as plataformas de redes sociais (em geral, liberdades comunicativas e econômicas). Tais pontos, com ênfase na liberdade de pensamento e informação, foram enfrentados com maior detalhamento nos dois primeiros capítulos deste trabalho.

Os limites institucionais, por sua vez, têm correlação com as restrições políticas e fáticas às quais um Estado está sujeito na tentativa de regulamentação de serviços no âmbito da *internet*, uma rede mundial de redes, que não observa limitações de ordem geográfica e que se encontra submetida a uma série de influxos de governança privada.⁴⁸³

Ao pensar no Estado, enquanto ente soberano, não é possível ignorar as inúmeras tensões que regulações nacionais podem ensejar a nível mundial, podendo tornar problemática o cumprimento de determinações (administrativas ou judiciais) a nível nacional.

Para ficarmos em um famoso exemplo brasileiro, cabe citar que apesar da aplicabilidade inequívoca do MCI a provedores de conexão e aplicação que prestam serviço em território nacional, não há ainda segurança jurídica quanto à forma de se assegurar o devido cumprimento das obrigações estabelecidas na norma, mesmo oito anos após a sua edição, no caso em que a ordem judicial ou administrativa implicar fornecimento de informações sobre dados armazenados no exterior por empresa estrangeira, mesmo quando relativos a nacionais ou a serviços prestados no país.

O tema está pendente de julgamento pelo STF. No bojo da ADC nº 51, a Federação das Empresas Brasileiras de Tecnologia e Informação (ASSESSPRO) busca o reconhecimento da constitucionalidade do Decreto nº 3.810/2001, que promulgou o Acordo de Assistência Judiciária em Matéria Penal (MLAT), entre Brasil e Estados Unidos. Por este decreto, a entrega de dados e comunicações de dados de empresas cujos *data centers* estejam nos Estados Unidos deve, necessariamente, obedecer aos trâmites burocráticos indicados no tratado internacional.

⁴⁸³ O termo “governança privada” apresenta um conteúdo mais amplo do que o de “regulação”, enquanto nesta há um elemento de intencionalidade, isto é, constrói-se de antemão um conjunto de regras públicas e privadas direcionadas a uma finalidade específica, a governança se caracteriza por envolver um plexo de atividades e ordens de caráter executivo ou não que embora influenciem o comportamento de terceiros, não possui propriamente uma intencionalidade. A governança, em verdade, percebe-se no curso do seu exercício como um procedimento de coordenação reflexiva. Cf. HOFMANN, Jeanette; KATZENBACH, Christian; GOLLATZ, Kirsten. Between Coordination and regulation: Finding the governance in Internet Governance. *New Media & Society*, [s.l.], v. 19, n. 9, p. 1406-1423, mar. 2016. Em âmbito nacional o tema é aprofundado por KELLER, Clara Iglesias. *Regulação Nacional de Serviços na Internet: Exceção, Legitimidade e o Papel do Estado*. Rio de Janeiro: Lumen Juris, 2019. op.cit.

O ajuizamento da ADC tem por objetivo evitar, em especial, requisições diretas do Poder de Judiciário pela obtenção destes dados, em descompasso com o procedimento administrativo fixado no MLAT. A despeito de o julgamento não ter se encerrado, os votos proferidos até o momento⁴⁸⁴ vêm seguindo a linha do relator Ministro Gilmar Mendes, em assentar a constitucionalidade parcial do referido decreto, com a ressalva de que é possível a solicitação direta de dados e comunicações eletrônicas às empresas de tecnologia estrangeiras, nas hipóteses do art. 11 do MCI e do art. 18 da Convenção de Budapeste.⁴⁸⁵

Tais obstáculos relativos ao mero cumprimento de decisões judiciais corriqueiras é uma pequena prova de todas as dificuldades que permeiam a regulação de serviços no âmbito da *internet* e serve de exemplo para o argumento *ciberlibertário* da insuscetibilidade da *internet* aos limites políticos e jurídicos do Estado. A dita natureza *excepcional* da *internet* era o fundamento usualmente evocado para afastar qualquer tentativa de condicionamento estatal à expansão da nova rede informacional.⁴⁸⁶

A tese do excepcionalismo, contudo, é passado. Como já se arguiu neste trabalho, ainda que o caráter global e aberto da *internet* exija a busca por novas estratégias regulatórias, suas peculiaridades não são suficientes, por si só, para afastar a *internet* dos influxos do Direito.⁴⁸⁷

A lembrança, contudo, das insuficiências de eventuais parâmetros regulatórios nacionais⁴⁸⁸, sempre que a ordem administrativa ou judicial possa resvalar na soberania de Estado estrangeiro (por exemplo, com determinação de derrubadas de páginas de outro país ou,

⁴⁸⁴ Foram proferidos apenas três votos (Ministros Gilmar Mendes, Nunes Marques e André Mendonça). Os dois últimos votaram pelo não conhecimento da ADC, mas, se vencidos, seguiram no mérito o Ministro Gilmar Mendes, relator do processo. As informações referentes ao julgamento podem ser obtidas em: <https://portal.stf.jus.br/processos/detalhe.asp?incidente=5320379>. Acesso em: 30 nov. 2022.

⁴⁸⁵ A Convenção de Budapeste foi incorporada ao ordenamento jurídico nacional pelo Decreto Legislativo nº 37, aprovado em 16 de dezembro de 2021, e tem como objetivo facilitar a cooperação internacional para combater o cibercrime. Elaborado pelo Comitê Europeu para os Problemas Criminais, com o apoio de uma comissão de especialistas, o documento lista os principais crimes cometidos por meio da rede mundial de computadores.

⁴⁸⁶ Esta é a tese implícita na declaração de *ciberindependência* de Joe Perry Barlow.

⁴⁸⁷ O ponto é enfrentado principalmente por Lawrence Lessig em obra já citada neste trabalho. Em sentido similar, com críticas ao argumento do excepcionalismo e detalhamento sobre o posicionamento ciberlibertário, pode ser citado também WU, Tim. Is Internet Exceptionalism Dead? In: SZOKA, Berim; MARCUS, Adam (Eds.). *The Next Digital Decade: Essays on the Future of Internet*. Washington, DC: TechFreedom, 2010, p.179-188.

⁴⁸⁸ Após o segundo turno das eleições para presidente de 2022, como forma de driblar as restrições contidas na supracitada Resolução nº 23.714/2022, apoiadores do atual presidente, vencido nas urnas, passaram a divulgar uma *live* produzida pelo canal *La Derecha Diário* (página eletrônica argentina), que *supostamente* apresentou provas de fraude no processo eleitoral brasileiro. A despeito da ausência de argumentos sólidos no vídeo, ante sua ampla repercussão em âmbito nacional, foi preciso que o Tribunal Superior Eleitoral emitisse uma nota explicativa sobre a higidez das urnas eletrônicas. Informação disponível em: VALFRÉ, Vinícius; GALZO, Wesley; FRAZÃO, Felipe. Live que distorce informações sobre urnas é assistida por 415 mil pessoas. *Uol Notícias*, [s.l.], 4 nov. 2022. Disponível em: <https://noticias.uol.com.br/ultimas-noticias/agencia-estado/2022/11/04/live-que-distorce-informacoes-sobre-urnas-e-assistida-por-415-mil-pessoas.htm>. Acesso em: 30 nov. 2022.

como citado acima, para demandar obtenção de dados armazenados no exterior), indica um risco de parcial inefetividade das pautas de vinculações estabelecidas aqui no Brasil.

A limitação do poder das redes sociais requer, portanto, mais que normas nacionais, mas também acordos de cooperação a nível internacional que tragam pautas mais ou menos homogêneas sobre desordens informacionais, opacidade algorítmica e transparência das plataformas.

Também no campo institucional, é preciso destacar que resta superado o mito de que o Estado seria capaz por si só de influenciar o comportamento da sociedade civil, orientando-a a propósitos positivos e democráticos⁴⁸⁹. No campo da *internet*, tais limitações são ainda mais dramáticas, visto que não raro mesmo medidas mais contundentes (inclusive as censórias) são facilmente contornadas no âmbito digital. Valendo-se das palavras de Michael Fromkin, a *internet* é uma espécie de Hidra Moderna. No mito grego, toda vez que alguém corta a cabeça desta criatura bestial, outras duas novas nascem. Fenômeno similar ocorre quando reguladores tentam interferir no ambiente *online* (cortando a cabeça da Hidra), e usuários contornam facilmente as novas regras (crescimento das novas cabeças).⁴⁹⁰

Ainda que regulação estatal promotora de transparência dos critérios de governança discursiva das plataformas de rede social seja de todo desejável, a mitigação das externalidades negativas produzidas pela ação privada depende também da assunção de um compromisso de parte substancial da sociedade civil com a democracia e o respeito aos direitos fundamentais.

É imperativo, portanto, que medidas de *literacia digital*⁴⁹¹ sejam fomentadas no campo social, bem como que empresas jornalísticas e agências de *fact-checking* aprimorem o rigor no combate à desinformação adotando táticas que evitem, tanto quanto possível, a amplificação de conteúdos falsos para além das bolhas informacionais existentes nas redes sociais⁴⁹². Somente

⁴⁸⁹ Este é basicamente um dos argumentos centrais da obra de Jacques Chevallier. Cf. CHEVALLIER, Jacques. *O Estado Pós-Moderno*. Trad. Marçal Justen Filho. Belo Horizonte: Fórum, 2009.

⁴⁹⁰ Cf. FROMKIN, A. Michael. The Internet as a Source of Regulatory Arbitrage. In: KAHIN, Brian; NESSON, Charles (Eds.). *Borders in Cyberspace*. Cambridge, MA: MIT Press, 1997.

⁴⁹¹ O termo pode ser utilizado como um sinônimo de “alfabetização digital” e engloba uma ampla gama de habilidades cognitivas, sociais, motoras e emocionais que os usuários precisam ter para navegar no ambiente digital de forma eficaz. Para os fins desta pesquisa, importa principalmente a criação de iniciativas que eduquem os usuários na avaliação do que é desinformação, *misinformation* ou informação maliciosa, bem como aprofundamento sobre os canais à disposição para denúncia de conteúdos ilegais ou danosos, e o conhecimento sobre critérios de monetização utilizados por redes sociais. A nível mundial, a União Europeia notabiliza-se por já ter uma iniciativa política estruturada de diretrizes e apoio à criação de sistemas de educação digital por parte de seus Estados-membros, cuja vigência se estenderá até 2027. A respeito, confira-se: EUROPEAN COMMISSION. Digital Education Action Plan (2021-2027). *European Education Area*, [s.l.], 24 fev. 2022. Disponível em: <https://education.ec.europa.eu/focus-topics/digital-education/action-plan>. Acesso em 30 nov. 2022.

⁴⁹² Exemplos destas táticas podem ser conferidos em entrevista dada pela fundadora da Agência Lupa, Cristina Tardáguila, à jornalista Patrícia Campos Mello da Folha de São Paulo. Dentre outras coisas, a entrevistada defendeu o “silêncio estratégico” para lidar com a desinformação. O *counterspeech* somente se justificaria para

a ampla capacitação dos usuários e o comprometimento das instituições civis com o pacto democrático podem, efetivamente, contribuir para a substancial redução dos impactos do fenômeno da desordem informacional.

Há, por fim, restrições de ordem temporal. Problemas de andamento (*pacing problem*) podem surgir na tentativa de mitigar os efeitos deletério produzidos por novas tecnologias, dado o desafio do regulador de normatizar um mercado em constante mudança onde a informação sobre aquilo que se regula é quase sempre escassa. É este desafio que torna o regulador “constantemente ameaçado pelos riscos de adoção de instrumentos abrangentes demais (...) ou de menos”⁴⁹³, oscilando entre o temor de atuar como censor e inimigo da inovação e o receio de regular de forma ineficiente.

Como já se expôs, ante o potencial inovador da *internet*, houve uma opção consciente do regulador por não estabelecer qualquer parâmetro de vinculação rígido aos particulares que optassem por explorar economicamente sua camada de conteúdo.

Esta opção acabou configurando, na prática, um terreno fértil para a construção de um novo modelo de exploração econômica ao qual Zuboff denomina de “capitalismo da vigilância”, sistema econômico calcado basicamente na coleta maciça de dados de particulares pelas *big techs* com o objetivo de venda das predições sobre comportamentos futuros, obtidas pelo uso de inteligência artificial, a terceiros interessados.⁴⁹⁴

Segundo a professora, a construção deste modelo econômico só foi possível porque os legisladores foram incapazes de interromper o que ela denomina de “ciclos de despossessão” da nossa autonomia. Este ciclo se iniciaria por uma fase de “incursão” sobre espaços juridicamente desprotegidos com objetivos de “sequestrar superávit comportamental de espaços da vida cotidiana não pertencentes ao mercado”.⁴⁹⁵

desinformações com amplo alcance viral. Do contrário, corre risco de a checagem funcionar como um impulsionador acidental do alcance de uma informação fraudulenta que, até então, não tinha furado as bolhas epistêmicas da *internet*. Sustentou ainda a criação de uma “editoria de democracia” nos principais veículos de imprensa para monitorar o risco de que certas manchetes e notícias possam ser descontextualizadas em plataformas de redes sociais na *internet* ou, ainda, para avaliar o risco de se dar espaço demais a certos posicionamentos ou informações claramente falaciosas ou que fletam com radicalismos. O objetivo desta editoria seria garantir a alta qualidade do padrão jornalístico da empresa e evitar que veículos de imprensa sejam manipulados por agentes promotores de desinformação ou discursos de ódio. A íntegra da entrevista pode ser conferida em: <https://www.youtube.com/watch?v=bfutdf951TQ> (Fundadora da Agência Lupa defende o silêncio contra desinformação). Acesso em 30 nov. 2022.

⁴⁹³ BAPTISTA, Patrícia; KELLER, Clara Iglesias. Por que, quando e como regular as novas tecnologias? Os desafios trazidos pelas inovações disruptivas. *Revista de Direito Administrativo*, Rio de Janeiro, v. 273, p. 123-163, set./dez. 2016, p.139.

⁴⁹⁴ ZUBOFF, Shoshana. *A Era do Capitalismo de Vigilância: a luta por um futuro humano na nova fronteira do poder*. Rio de Janeiro: Intrínseca, 2020, p. 113-119.

⁴⁹⁵ *Ibid.*, p. 165.

Posteriormente, a popularização do serviço daria início a uma fase de “habituação”, em que as “práticas contestadas tornam-se consagradas como fatos institucionais”⁴⁹⁶ e as pessoas se acostumam à vulneração de direitos basilares como liberdade e privacidade, dissipando qualquer sensação de perplexidade.

Diante de vazamentos ocasionais de informações internas que evidenciem a má-fé destas empresas ou, no mínimo, a negligência com violação de direitos fundamentais, dá-se início à terceira fase de “adaptações superficiais”, uma “nova retórica, métodos, elementos e projetos que redirecionam as operações (...) apenas o suficiente para parecer compatíveis com as exigências sociais e jurídicas”⁴⁹⁷. E, à medida em que as exigências jurídicas se robustecem e a sociedade toma maior ciência dos riscos envolvidos em determinadas atividades, as principais empresas de tecnologia buscam *redirecionar* suas fontes de suprimento de superávit comportamental mediante criação de novas iniciativas.⁴⁹⁸

A descrição detalhada de Zuboff confere cores ao desafio que David Collingridge já apontava existir quando se pensa na regulação de novas tecnologias. De um lado, antecipar-se no processo regulatório poderia gerar problemas informacionais, visto que o regulador não conhece suficientemente o objeto de regulação ainda em desenvolvimento. De outro, esperar demais para regular limita significativamente o Estado, visto que é especialmente difícil controlar uma tecnologia já difundida socialmente.⁴⁹⁹ Estes desafios ficaram comumente conhecidos como *Dilema de Collingridge*.

Estas considerações são especialmente importantes, pois, à medida em que os debates sobre regulações do poder de governança discursiva das redes sociais avançam no mundo todo, a principal empresa do setor (responsável pelo gerenciamento do Facebook, Instagram e Whatsapp) tem centralizado seus recursos no desenvolvimento de um novo projeto de interação social, que busca mesclar a realidade com mundo virtual em um conceito que ficou conhecido como *metaverso*.⁵⁰⁰

Não se sabe se o projeto será, de fato, bem-sucedido, mas a mudança dos rumos da empresa parece indicar justamente uma tentativa de *redirecionamento* das fontes de suprimento

⁴⁹⁶ Ibid., loc. cit.

⁴⁹⁷ ZUBOFF, Shoshana. *A Era do Capitalismo de Vigilância: a luta por um futuro humano na nova fronteira do poder*. Rio de Janeiro: Intrínseca, 2020, p. 167.

⁴⁹⁸ Ibid., p. 177

⁴⁹⁹ Cf. COLLINGRIDGE, David. *The Social Control of Technology*. Nova Iorque: St. Martin Press, 1980.

⁵⁰⁰ O investimento é tamanho que levou à mudança no nome da própria empresa. A respeito, confira-se: CARVALHO, Lucas. Facebook agora é Meta: entenda a mudança de nome da empresa. *Tilt Uol*, [s.l.], 28 out. 2021. Disponível em: <https://www.uol.com.br/tilt/noticias/redacao/2021/10/28/facebook-agora-e-meta-entenda-a-mudanca-de-nome-da-empresa.htm>. Acesso em: 06 dez. 2022.

de superávit comportamental, que darão reinício ao ciclo de despossessão mencionado por Zuboff.

Embora a doutrina usualmente indique que os riscos de uma regulação tardia sejam normalmente menores de que uma regulação prematura⁵⁰¹, o atual conhecimento que se tem sobre as externalidades produzidas por redes virtuais de interação social parece sugerir a regulação *ex ante* como caminho mais protetivo aos direitos fundamentais diante de inovações que busquem incrementar o potencial de fusão entre mundo virtual e realidade fática.

Isso justificaria inclusive a exigência de autorização prévia para o exercício da atividade econômica⁵⁰², criação de *sandboxes* regulatórios ou, ainda, a construção de modelos regulatórios focados no risco representado pelo desenvolvimento de novas estruturas de inteligência artificial.⁵⁰³

A enumeração destes limites indica obstáculos a eventual projeto de regulamentação nacional dos poderes de governança das redes sociais, mas não afasta a conclusão de que um projeto corregulatório é o primeiro e necessário passo para a constitucionalização deste espaço digital.

A regulação pública das plataformas amplia o grau de consciência dos cidadãos sobre o funcionamento das redes sociais, jogando luz sobre os elementos, critérios, finalidades, taxas de erros e acertos das medidas de moderação ou de curadoria algorítmica aplicadas por estas plataformas. Com mais informações, maior a capacidade da sociedade civil de se mobilizar e cobrar transparência das redes sociais.

Em segundo lugar, constrange estes agentes privados na aplicação de suas próprias normas de moderação, demandando não só o respeito ao devido processo legal na aplicação de seus termos, como também exigindo a conformação de suas normas comunitárias à legislação nacional, mantendo tanto quanto possíveis tais agentes privados limitados pelo ordenamento jurídico pátrio.

E, por fim, indica um caminho não só menos perigoso, mas potencialmente mais eficiente à limitação do poder de governança discursiva das plataformas de redes sociais. O

⁵⁰¹ Cf. BAPTISTA, Patrícia; KELLER, Clara Iglesias. Por que, quando e como regular as novas tecnologias? Os desafios trazidos pelas inovações disruptivas. *Revista de Direito Administrativo*, Rio de Janeiro, v. 273, p. 123-163, set./dez. 2016, p. 145.

⁵⁰² Cf. TUTT, Andrew. An FDA For Algorithms. *Admin. L. Rev.*, [s.l.], v. 83, 2017. Disponível em: <https://ssrn.com/abstract=2747994>. Acesso em: 06 dez. 2022.

⁵⁰³ A regulação pelo risco foi opção adotada pela União Europeia na proposta da Comissão Europeia para regulamentação da inteligência artificial em seu território. A depender do grau de risco que o uso determinado da inteligência artificial possa representar, a regulação será distinta. Os graus de risco variam do inaceitável — hipótese na qual o desenvolvimento da IA é proibido — passando por graus de riscos considerados como elevados, limitados e mínimos. O texto do regulamento pode ser conferido em: <https://eur-lex.europa.eu/legal-content/PT/TXT/HTML/?uri=CELEX:52021PC0206&from=EN>. Acesso em: 06 dez. 2022.

enfoque de regulação sobre plataformas de comunicação na *internet* não deve ser o conteúdo das mensagens ainda que danosas, mas sim o estímulo à mudança de arquiteturas algorítmicas produtoras de bolhas informacionais e câmaras de eco.

Mais do que nunca, faz-se presente a lembrança do filósofo e estudioso dos meios de comunicação Marshall McLuhan, de que o “meio é a mensagem”⁵⁰⁴ e, portanto, novas tecnologias da informação não são meros canais de passagem de conteúdo, mas antes ferramentas que “controlam a forma e a proporção das ações e associações humanas”⁵⁰⁵ influenciando e moldando a cultura e o comportamento humano. Cabe à sociedade civil e suas instituições compreender, discutir e, se necessário, sugerir caminhos alternativos para mitigar os impactos desta mudança sobre nós.

A pesquisa, mesmo ciente de suas limitações, buscou humildemente contribuir para esta empreitada.

⁵⁰⁴ MCLUHAN, Marshall. *Os meios de comunicação como extensões do homem*. São Paulo: Pensamento-Cultrix, 2006, p. 21.

⁵⁰⁵ MCLUHAN, Marshall. *Os meios de comunicação como extensões do homem*. São Paulo: Pensamento-Cultrix, 2006, p. 22.

CONCLUSÕES

As ideias desenvolvidas ao longo do trabalho podem ser sintetizadas por meio das seguintes proposições objetivas.

1. Não há no Brasil qualquer parâmetro regulatório que vincule o poder de governança discursiva das plataformas de redes sociais. Atualmente, a única regra vigente sobre o tema — o art. 19 do Marco Civil da *Internet* — limita-se a prever uma mitigada responsabilidade às aplicações de *internet*, na hipótese de descumprimento de ordem judicial e específica para retirada de conteúdo publicado em seus canais por terceiros usuários de seus serviços;
2. A ausência de um parâmetro regulatório confere, portanto, protagonismo ao Poder Judiciário na imposição de condicionantes às políticas de moderação destas plataformas de redes sociais, o que é especialmente problemático se considerarmos que somente uma minúscula fração dos debates envolvendo os limites ao exercício da liberdade de expressão nas redes sociais chegam ao Poder Judiciário (problema da seletividade), inviabilizando um real condicionamento ao poder das maiores redes sociais;
3. Defende-se, portanto, a necessidade de construção de um parâmetro regulatório legal para as plataformas de redes sociais, proposta que requer ênfase ao aspecto democrático do direito à liberdade de expressão. Ao contrário do contexto norte-americano, marcado por tradição fortemente libertária, a Constituição Brasileira claramente reforça o aspecto democrático e igualitário como elemento a ser considerado na definição dos limites ao exercício deste direito. Partindo desta premissa, quando se pensa em regulação de canais e meios de comunicação como as redes sociais, é possível aventar como objetivo primordial destas normas a criação de um ambiente virtual que impulse a reciprocidade e o caráter reflexivo à construção da esfera pública contemporânea, não havendo um compromisso quase que absoluto com a liberdade de quem discursa;
4. A intervenção estatal na esfera da comunicação enseja, contudo, riscos que não podem ser ignorados. O receio do autoritarismo estatal demanda, por um lado, a necessidade de se priorizar a lei na construção de regimes jurídicos que imponham deveres aos canais de comunicação social, ante o maior potencial da esfera

legislativa para deliberação e veiculação pública de razões, cabendo sempre ao Poder Judiciário o papel de revisor eventual de tais propostas regulatórias em razão da natureza fundamental do direito à liberdade de expressão. Por outro lado, o *chilling effect* é elemento a ser considerado na construção de parâmetros regulatórios que possam inibir, direta ou incidentalmente, certos tipos de discursos, exigindo do Poder Público cautela na elaboração de condutas proibitivas a fim de reduzir a margem de discricionariedade ao intérprete da norma;

5. Em linhas gerais, leis regulatórias no campo da comunicação devem observar os seguintes parâmetros: a) amparar-se pelo filtro da razão pública ; b) jamais podem se converter em medidas de censura prévia (art. 5º, IX da CRFB), tampouco censura por razões ideológicas, políticas ou artísticas (art. 220, § 2º CRFB); c) devem, em regra, priorizar a capacidade de influir na “arquitetura de escolhas” criadas pelos detentores dos meios de comunicação; d) podem, excepcionalmente, incidir diretamente sobre o conteúdo, desde que haja consensos solidamente estabelecidos e justificáveis acerca dos riscos graves de danos que certos discursos podem produzir à democracia e à sociedade como um todo; e) os atos regulatórios passíveis de restrição da liberdade de comunicar devem, tanto quanto possível, pautarem-se por critérios objetivos, evitando o uso de expressões vagas e ambíguas capazes de produzir efeito resfriador;
6. Considerando, especificamente, uma proposta regulatória para as redes sociais, é fundamental compreender como estas novas plataformas passaram a influenciar a relação entre Estado e sociedade civil. Segundo teoria de Balkin, a liberdade de expressão hoje seria um triângulo, onde, de um lado, estaria o Estado, capaz de condicionar a liberdade discursiva por meio de leis, regulamentos e decisões judiciais; de outro, os indivíduos da sociedade civil, que exercem sua liberdade discursiva de forma direta ou organizada por meio de associações coletivas, instituições e canais de mídia em massa; e, na terceira ponta, os intermediários digitais (dentre os quais se incluem as redes sociais), aplicações da *internet* que são responsáveis pela circulação de conteúdo na *internet*. Tais plataformas detêm hoje verdadeiros aparatos tecnológicos e burocráticos, que lhe conferem a capacidade de governar discursos e monitorar expressões e comportamentos de seus usuários;
7. Quando se pensa em limites à liberdade de expressão e autonomia do indivíduo há, portanto, novos riscos em jogo. Ao lado do receio da censura e do *chilling effect*

estatal, há novas formas de censura, como a colateral (quando o Estado usa o aparato tecnológico privado para cercear a capacidade discursiva dos cidadãos) e a reversa (quando o Estado usa a estrutura algorítmica das redes sociais para inundar a esfera pública de discursos que sejam favoráveis ao governo da vez, mesmo quando possam configurar falsidades, ataques odiosos ou incitação à violência);

8. Por outro lado, as plataformas de redes sociais são produtoras de novas externalidades, que contaminam o “livre mercado de ideias”. A manipulação algorítmica, produtora de câmaras de eco e filtros-bolha, cria um cenário fértil para a produção de *desordens informacionais* que, em último caso, representam uma ameaça não só à autonomia individual (não há mais segurança para distinguir o verdadeiro do falso), como também à própria democracia (a quebra da confiança mútua dificulta a preservação de uma base de valores em comum);
9. Por regra, a moderação de conteúdo tem sido a resposta dada pelas plataformas para mitigar os impactos decorrentes das externalidades oriundas das desordens informacionais. Tal moderação se concretiza por meio do aprimoramento dos termos de uso e normas comunitárias das redes sociais de modo a compatibilizá-los com as exigências legais de cada país, mas sua aplicação tem sido alvo de crítica, tanto pela falta de transparência quanto aos critérios utilizados, quanto pela ineficiência no cerceamento das desordens informacionais, apresentando elementos que mais a aproximam de uma “censura privada” do que de um instrumento legítimo de limitação a manifestações abusivas no âmbito da *internet*;
10. A partir deste cenário, é possível aventar que uma proposta regulatória das redes sociais deve ter dois objetivos centrais: (a) mitigar os impactos negativos das desordens informacionais, externalidades negativas oriundas, principalmente, das técnicas de perfilhamento e microdirecionamento que garantem o funcionamento lucrativo das plataformas de redes sociais; (b) conferir legitimidade às técnicas de moderação de conteúdo aplicadas por estas plataformas;
11. Dentre as estratégias regulatórias possíveis, sustenta-se que a regulação orientada à mudança de arquitetura das plataformas de redes sociais, mediante criação de deveres que propiciem a alteração dos algoritmos que definem quais conteúdos terão maior ou menor visibilidade na plataforma, seja uma forma mais eficiente e potencialmente menos onerosa à liberdade de expressão se comparada às estratégias voltadas ao conteúdo (proibição mediante ameaça de sanção de

veiculação de certos tipos de discurso), cujo papel seria residual, limitado a prevenir uso de estratégias de censura reversa por agentes políticos;

12. Tais estratégias podem, ainda, concretizar-se por meio de arranjos distintos. Há quem defenda que a autorregulação orientada por compromissos públicos de maior transparência seja suficiente para conferir transparência e legitimidade ao poder de governança nas redes sociais;
13. Defende-se, contudo, que mecanismos corregulatórios ou de autorregulação regulada seriam mais eficientes para cumprimento de tal finalidade, visto que, ao atribuir a um agente externo estatal o papel de fiscalizador da atuação destas plataformas de redes sociais, ainda que indiretamente, tem-se aptidão coercitiva maior sobre o agente privado regulado;
14. Um regime corregulatório das plataformas de redes sociais teria por base a exigência de cumprimento de três deveres por parte destes agentes privados;
15. O primeiro é o de transparência. O dever de transparência consiste no reconhecimento de obrigação das plataformas de redes sociais em relação ao Estado e usuários de compartilhamento de informações tidas como essenciais para compreender se os processos de curadoria algorítmica (definição *a priori* dos critérios de visibilidade/invisibilidade das postagens) e de moderação (elaboração e aplicação dos termos de uso e normas comunitárias) estão ocorrendo de forma a observar os direitos fundamentais dos usuários; em especial, o direito fundamental à liberdade de expressão;
16. Trata-se da mais importante das obrigações das plataformas, pois é a partir dela que o Poder Público e a sociedade civil verificarão se as principais redes sociais estão cumprindo as obrigações de promover o funcionamento do livre mercado de ideias e de assegurar uma esfera pública efetivamente democrática;
17. O segundo é o de observância do devido processo legal. Consiste na exigência de que a decisão de moderação destas plataformas, ainda que fundada em seus termos de uso, sejam motivadas, por implicarem restrições a um direito fundamental de liberdade. A motivação legitimadora de tais decisões demanda principalmente a construção de procedimentos decisórios, que disponibilizem aos usuários/cidadãos não só um mecanismo facilmente compreensível de denúncias, mas que também lhes permitam se defender quando denunciados, conferindo compreensão sobre as razões da decisão da plataforma, garantindo a aplicação de sanções proporcionais

e, ainda, viabilizando mecanismos de recorribilidade humana, especialmente nos casos de decisão automatizada;

18. O terceiro é o de proteção à boa-fé dos usuários, que se inspira em uma constatação de que as redes sociais não podem atuar de forma contraditória com os objetivos e missões institucionais que anunciam em seus atos constitutivos. É o dever que impõe às redes sociais a obrigação de criar arquiteturas que incentivem a reflexão e o contato com múltiplos pontos de vista, estimulando a promoção da liberdade de expressão desde a concepção da plataforma tecnológica. Em contrapartida, as redes sociais não podem usar dados de seus usuários para manipulá-los e explorar vieses de forma ardil, tampouco ser conivente ou contribuir para a prática de ilicitudes em suas plataformas;
19. Neste contexto de responsabilização administrativa, caberia ao Poder Judiciário duas funções. A primeira seria assegurar a higidez da ação sancionatória do órgão ou entidade administrativa fiscalizatória. No campo da responsabilização civil, sugere-se a revisão parcial do art. 19 do MCI. A previsão de limitada responsabilidade, condicionada a prévia ordem judicial, seria extensível apenas às plataformas de redes sociais “de boa-fé”, isto é, aquelas que estejam em consonância com seus deveres de transparência, respeito ao devido processo legal e proteção à boa-fé de seus usuários definidos na esfera administrativa;
20. Há, por fim, limites a serem considerados na criação de um novo regime regulatório das plataformas de redes sociais. Para além dos limites jurídicos, que limitam as opções regulatórias à disposição do Estado, ante a necessidade de preservação de valores como a liberdade de expressão, autonomia e privacidade dos usuários e a autonomia negocial das empresas que operam as plataformas, os limites de ordem institucional exigem comprometimento com a preservação dos direitos fundamentais no âmbito da *internet* não só do Estado, mas de outros atores relevantes a nível internacional e de setores privados da sociedade;
21. Há, também, limitações de ordem temporal, haja vista a dificuldade de os debates regulatórios acompanharem a sempre rápida evolução tecnológica. O avanço do conhecimento sobre a *internet* e o funcionamento dos algoritmos das plataformas de redes sociais parece recomendar modelos regulatórios calcados no potencial de risco das ferramentas de inteligência artificial como meio de prevenção ao surgimento de novas ameaças ao exercício dos direitos fundamentais.

22. Nada disso afasta a consideração central deste trabalho de que arranjos corregulatórios orientado primariamente ao controle da arquitetura algorítmicas das redes sociais é o primeiro e necessário passo para a efetiva constitucionalização deste espaço digital. Primeiro, por ampliar o grau de consciência dos cidadãos sobre o funcionamento destes espaços, jogando luz sobre os elementos, critérios, finalidades, taxas de erros e acertos das medidas de moderação ou de curadoria algorítmica aplicadas por estas plataformas. Segundo, por constranger tais plataformas a cumprir não só o pactuado nos seus termos de adesão e normas comunitárias, como também para viabilizar que as redes sociais sejam efetivamente deferentes ao ordenamento jurídico nacional e aos comandos emanados do Poder Judiciário. Terceiro, viabiliza uma resposta equilibrada à necessidade de imposição de controle à autonomia das principais redes sociais, sem esquecer do risco sempre presente de que regulações públicas que impactem no exercício ao direito da liberdade de expressão possam se converter em novas ferramentas de censura estatal.

REFERÊNCIAS

ABBOUD, Georges; CAMPOS, Ricardo. A autorregulação regulada como modelo do Direito proceduralizado. In: ABBOUD, Georges; CAMPOS, Ricardo; NERY JR., Nelson (Orgs.). *Fake News e Regulação*. 2. ed. São Paulo: Revista dos Tribunais, 2020.

ABRANCHES, Sérgio. *O tempo dos Governantes Incidentais*. São Paulo: Companhia das Letras, 2020.

ALCOTT, H.; GENTZKOW, M. Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, [s.l.], v. 31, n. 2, p. 211-236, 2017.

ALEMANHA multa Facebook em 2 milhões de euros por violação à lei de transparência online. *GI*, [s.l.], 2 jul. 2019. Disponível em: <https://g1.globo.com/economia/tecnologia/noticia/2019/07/02/alemanha-multa-facebook-em-2-milhoes-de-euros-por-violacao-a-lei-de-transparencia-online.ghtml>. Acesso em: 14 jun. 2022.

ALEXY, Robert. *Teoria dos Direitos Fundamentais*. 2. ed. Trad. Virgílio Afonso da Silva. São Paulo: Malheiros, 2017.

ALMEIDA, Guilherme da Franca Couto Fernandes de; LEITE, Fábio Carvalho; HANNIKAINEN, Ivar Allan Rodriguez. Liberdade de Expressão e Direito à honra: Medindo Atitudes e Prevendo Decisões. *Espaço Jurídico Journal of Law*, Joaçaba, v. 21, n. 1, p. 191-216, jan./jun. 2020.

ANDRADE, André. *Liberdade de expressão em tempos de cólera*. Rio de Janeiro: GZ, 2020.

AGÊNCIA tentou contratar youtubers para desacreditar vacinas contra covid. *Exame*, [s.l.], 26 jul. 2021. Disponível em: <https://exame.com/tecnologia/agencia-tentou-contratar-youtubers-para-desacreditar-vacinas-contracovid/>. Acesso em: 16 dez. 2021.

AGENTE da CIA diz que Putin interferiu nas eleições americanas de 2016. *Exame*, [s.l.], 10 set. 2019. Disponível em: <https://exame.com/mundo/russia-confirma-que-empregou-espiao-dos-eua-em-gabinete-presidencial/>. Acesso em: 16 dez. 2021.

ARAGÃO, Alexandre Santos de. *Curso de Direito Administrativo*. 2. ed. Rio de Janeiro: Forense, 2013.

_____. *Direito dos Serviços Públicos*. 4. ed. Belo Horizonte: Fórum, 2017.

ARAL, Sinan; ROY, Deb; VOSOUGHI, Soroush. The spread of true and false news online. *Science*, [s.l.], v. 359, n. 6380, p. 1146-1151, mar. 2022. Disponível em: <https://www.science.org/doi/full/10.1126/science.aap9559>. Acesso em 04 fev. 2022.

ARCHEGAS, João Victor. Admirável 2020 Novo: banimento de Trump de redes sociais e moderação de conteúdo. *Jota*, [s.l.], 13 jan. 2021. Disponível em:

<https://www.jota.info/opiniao-e-analise/artigos/admiravel-2020-novo-banimento-de-trump-de-redes-sociais-e-moderacao-de-conteudo-13012021>. Acesso em: 24 jun. 2022.

_____.; BOTTINO, Celina; PERRONE, Christian. Moderação de Conteúdo em 2021: quem regula o moderador? *Jota*, [s.l.], 17 jan. 2022. Disponível em: <https://www.jota.info/opiniao-e-analise/artigos/moderacao-de-conteudo-2021-quem-regula-moderador-17012022>. Acesso em: 17 mai. 2022.

ARCHEGAS, João Victor; ESTARQUE, Marina. *Redes Sociais e Moderação de Conteúdo: criando regras para o debate público a partir da esfera privada*. Rio de Janeiro: ITS, 2020.

ARENDDT, Hannah. *As Origens do Totalitarismo*. Trad. Roberto Raposo. São Paulo: Companhia das Letras, 2012.

ARUN, Chinmaayi. Facebook's Face. *Harvard Law Review*, [s.l.], v. 135, abr. 2021. Mimeografado.

ATUALIZAÇÕES no Centro de Transparência do Twitter. *Blog Twitter*, [s.l.], 14 jul. 2021. Disponível em: https://blog.twitter.com/pt_br/topics/company/2021/atualizacoes-no-centro-de-transparencia-do-twitter. Acesso em: 19 abr. 2022.

ÁVILA, Humberto. *Teoria dos princípios: da definição à aplicação dos princípios jurídicos*. 16. ed. São Paulo: Malheiros, 2015.

BALDI, Vania. A construção viral da realidade: ciberpopulismos e polarização dos públicos em rede. *Observatorio (OBS*)*, [s.l.], special issue, p. 4-20, 2018.

BALDWIN, Robert; CAVE, Martin; LODGE, Martin. *Understanding Regulation*. 2. ed. Nova Iorque: Oxford University Press, 2012.

BALKIN, Jack M. Digital Speech and Democratic Culture: A Theory of Freedom of Expression for the Information Society. *N.Y.U. L. Rev.*, [s.l.], v. 79, n. 1, 2004.

_____. Free Speech is a triangle. *Columbia Law Review*, [s.l.], v. 118, p. 2011-2056, mai. 2018.

_____. Free Speech in the Algorithmic Society: Big Data, Private Governance and New School Speech Regulation. *UC Davis Law Review*, Davis, v. 51, n. 3, p. 1149-1210, fev. 2018. Mimeografado.

_____. How to regulate social media. *Journal of Free Speech Law*, [s.l.], v. 1, n. 71, nov. 2019. Mimeografado.

_____. Information Fiduciaries and First Amendment. *UC Davies Law Review*, [s.l.], v. 4, n. 49, 2016.

_____. O Futuro da Liberdade de Expressão na Era Digital. In: SAMPAIO, José Adércio Leite (Coord.). *Liberdade de Expressão no Século XXI*. Belo Horizonte: Del Rey, 2016.

BAPTISTA, Patrícia; KELLER, Clara Iglesias. Por que, quando e como regular as novas tecnologias? Os desafios trazidos pelas inovações disruptivas. *Revista de Direito Administrativo*, Rio de Janeiro, v. 273, p. 123-163, set./dez. 2016.

BARBERÁ, Pablo. Social Media, Echo Chambers and Political Polarization. In: _____. *Social media and democracy*. [S.l.]: Cambridge University Press, 2020. Disponível online em: <https://www.cambridge.org/core/books/social-media-and-democracy/social-media-echo-chambers-and-political-polarization/333A5B4DE1B67EFF7876261118CCFE19> Acesso em: 06 mai. 2022.

BARCELLOS, Ana Paula de. *Ponderação, Racionalidade e Atividade Jurisdicional*. Rio de Janeiro: Renovar, 2005.

BARLOW, John Perry. A Declaration of Independence of Ciberspace. *Duke Law and Technology Review*, [s.l.], v. 18, n. 5-7, 2019.

BAROCAS, Solon; SELBST, Andrew D. Big Data's Disparate Impact. *Calif. L. Rev.*, [s.l.], v. 104, n. 671, 2016.

BARROSO, Luís Roberto. *A Judicialização da Vida e o Papel do Supremo Tribunal Federal*. Belo Horizonte: Fórum, 2021.

_____. Aqui, lá e em todo lugar: A dignidade da pessoa humana no direito contemporâneo e no discurso transnacional. *Revista dos Tribunais*, São Paulo, v. 919, p. 127-196, 2012.

_____. Colisão entre Liberdade de expressão e Direitos da Personalidade: critérios de ponderação. Interpretação constitucionalmente adequada do Código Civil e da Lei de Imprensa. *Revista de Direito Administrativo*, Rio de Janeiro, v. 235, p. 1-36, jan./mar. 2014.

_____. Interpretação constitucionalmente adequada do Código Civil e da Lei de Imprensa. *Revista de Direito Administrativo*, Rio de Janeiro, v. 235, p. 1-36, jan./mar. 2014.

_____. Liberdade de expressão, censura e controle da programação na Constituição de 1988. In: _____. *Temas de Direito Constitucional*. Rio de Janeiro: Renovar, 2001. Tomo 1.

_____. Revolução tecnológica, crise da democracia e mudança climática: limites do direito num mundo em transformação. *Revista Estudos Institucionais*, [s.l.], v. 5, n. 3, p. 1234-1313, 2019. Mimeografado.

_____. *Sem data venia: um olhar sobre o Brasil e o Mundo*. [S.l.]: História Real, 2020.

BARROSO, Luna Van Brussel. *Liberdade de Expressão e Democracia na Era Digital*. Belo Horizonte: Fórum, 2021.

BBC. META vai bloquear conteúdo pró-invasão em Brasília das Redes Sociais. 09 de janeiro de 2023. Disponível em: <https://g1.globo.com/tecnologia/noticia/2023/01/09/meta-vai-bloquear-conteudo-pro-invasao-em-brasilia-das-redes-sociais.ghtml>.

BEDI, Sunéal. The Mith of Chilling Effect. *Harvard Journal of Law & Technology*, [s.l.], v. 35, n. 1, p. 267-307, 2021. Mimeografado.

BELLI, Luca; FRANCISCO, Pedro Augusto; ZINGALES, Nicolo. Lei do Estado ou lei da Plataforma? Cuidado com a privatização da regulação e da polícia. In: BELLI, Luca; CAVALLI, Olga (Orgs.). *Governança e Regulações da Internet na América Latina: análise sobre infraestrutura, privacidade, cibersegurança e evoluções tecnológicas em homenagem aos dez anos da South School on Internet Governance*. Rio de Janeiro: FGV, 2019.

BEM-SHAHAR, Omri. Data Pollution. *Journal of Legal Analysis*, [s.l.], v. 11, 2019.

BHARGAVA, Vikram; VELASQUEZ, Manuel. Ethics of Attention Economy: The Problem of Social Media Addiction. *Business Ethics Quarterly*, [s.l.], v. 31, n. 3, p. 321-359, jul. 2021.

BINENBOJM, Gustavo. Fake news como externalidades negativas. *Jota*, [s.l.], 23 jun. 2020. Disponível em: <https://www.jota.info/opiniao-e-analise/colunas/publicistas/fake-news-como-externalidades-negativas-23062020>. Acesso: 20 abr. 2022.

_____. *Poder de Polícia, Ordenação e Regulação: Transformações Político-Jurídicas, Econômicas e Institucionais do Direito Administrativo Ordenador*. 3. ed. Belo Horizonte: Fórum, 2020.

BLACK, Julia. Decentring Regulation: Understanding the Role of Regulation and Self-Regulation in a “Post-Regulatory” World. *Current Legal Problems*, Oxford: Oxford University Press, v. 54, n. 1., p. 103-146, jan. 2001.

BOLSONARO compartilha conteúdo falso, e Instagram oculta postagem do presidente. *GI*, [s.l.], 12 mai. 2020. Disponível em: <https://g1.globo.com/economia/tecnologia/noticia/2020/05/12/bolsonaro-compartilha-conteudo-falso-e-instagram-oculta-postagem-do-presidente.ghml>. Acesso em: 20 dez. 2022.

BOMFIM Camila. Alexandre de Moraes revoga decisão que censurou reportagens de ‘Crusoé’ e ‘O Antagonista’. *GI*, [s.l.], 18 abr. 2019. Disponível em: <https://g1.globo.com/politica/noticia/2019/04/18/alexandre-de-moraes-revoga-decisao-que-censurou-reportagens-de-crusoe-e-antagonista.ghml>. Acesso em: 24 nov. 2022.

BRANDÃO, Rodrigo. *Supremacia Judicial versus Diálogos Constitucionais: a quem cabe a última palavra sobre o sentido da Constituição*. Rio de Janeiro: Lumen Juris, 2018.

CADWALLADR, Carole. The great British Brexit robbery: how our democracy was hijacked. *The Guardian*, [s.l.], 7 mai. 2017. Disponível em: <https://www.theguardian.com/technology/2017/may/07/the-great-british-brex-it-robbery-hijacked-democracy>. Acesso em: 21 abr. 2022.

CALEGARI, Luiza; GOES, Severino. Roberto Jefferson, aliado de Bolsonaro, é preso por ordem do ministro Alexandre, do STF. *Consultor Jurídico*, [s.l.], 13 ago. 2021. Disponível em: <https://www.conjur.com.br/2021-ago-13/alexandre-manda-prender-roberto-jefferson-aliado-bolsonaro>. Acesso em: 24 nov. 2022.

“CARNAVAL da esquerda”, diz Carlos após Instagram marcar posts como falsos. *Poder 360*, [s.l.], 8 nov. 2021. Disponível em: <https://www.poder360.com.br/brasil/carnaval-da-esquerda-diz-carlos-apos-instagram-marc-ar-posts-como-falsos/>. Acesso em: 20 dez. 2022.

CARVALHO, Ana Cristina Azevedo P. *Marco Civil da Internet no Brasil: análise da Lei nº 12.965/14 e do Direito de Informação*. Editora Rio de Janeiro: Alta Books, 2014.

CARVALHO, Lucas. Facebook agora é Meta: entenda a mudança de nome da empresa. *Tilt Uol*, [s.l.], 28 out. 2021. Disponível em: <https://www.uol.com.br/tilt/noticias/redacao/2021/10/28/facebook-agora-e-meta-entenda-a-mudanca-de-nome-da-empresa.htm>. Acesso em: 06 dez. 2022.

_____; MILITÃO, Eduardo. Bolsonaro cita decreto para regular internet, mas assessores negam. *Tilt Uol*, [s.l.], 5 mai. 2021. Disponível em: <https://www.uol.com.br/tilt/noticias/redacao/2021/05/05/bolsonaro-fala-em-regulamentar-marco-civil-da-internet-somos-cerceados.htm>. Acesso em: 17 dez. 2021.

CASTELLS, Manuel. *A ruptura: a crise da democracia liberal*. Rio de Janeiro: Zahar, 2018.

_____. *A Sociedade em rede*. São Paulo: Paz e Terra, 2013. v.1.

CAVALLAZI, Vanessa Wendhausen. *E-democracia deliberativa: a criação de espaços de deliberação em rede para a implementação de direitos sociais*. São Paulo: Juspodium, 2020.

CELESTE, Edoardo. Digital Constitutionalism: a new systematic theorisation. *International Review of Law, Computers and Technology*, [s.l.], v. 33, n. 1, p. 76-99, 2019.

CHEQUER, Cláudio. *A Liberdade de Expressão como Direito Preferencial Prima Facie: análise crítica de revisão ao padrão jurisprudencial brasileiro*. 2. ed. Rio de Janeiro: Lumen Juris, 2017.

CHEVALLIER, Jacques. *O Estado Pós-Moderno*. Trad. Marçal Justen Filho. Belo Horizonte: Fórum, 2009.

COMPARATO, Fábio Konder. A Democratização dos Meios de Comunicação de Massa. In: GRAU, Eros Roberto; GUERRA FILHO, Willis Santiago (Orgs.). *Direito Constitucional: Estudos em Homenagem a Paulo Bonavides*. São Paulo: Malheiros, 2001.

CONTEXTO: as manifestações do 7 de Setembro. *G1*, [s.l.], 7 set. 2021. Disponível em: <https://g1.globo.com/politica/noticia/2021/09/07/contexto-as-manifestacoes-do-7-de-setembro.ghtml>. Acesso em: 17 dez. 2021.

COLLINGRIDGE, David. *The Social Control of Technology*. Nova Iorque: St. Martin Press, 1980.

CORTE INTERAMERICANA DE DIREITOS HUMANOS. *Caso Herzog e outros vs. Brasil*. Autores: Centro pela Justiça e o Direito Internacional e outros. Juiz Presidente: Eduardo Ferrer Mac-Gregor Poisot, 15 de março de 2018. Disponível em: https://www.corteidh.or.cr/docs/casos/articulos/seriec_353_por.pdf. Acesso em: 28 mar. 2022.

CRUZ, Bruna Souza. Veja motivos do Facebook, Twitter e outras plataformas para bloquear Trump. *Tilt Uol*, [s.l.], 7 jan. 2021. Disponível em:

<https://www.uol.com.br/tilt/noticias/redacao/2021/01/07/trump-bloqueado-saiba-os-motivos-da-acao-do-facebook-instagram-e-twitter.htm>. Acesso em: 14 dez. 2021.

CRUZ, Francisco Brito. *Novo Jogo Velhas Regras: Democracia e Direito na era da nova propaganda política e das fake news*. Belo Horizonte: Casa do Direito, 2020.

DAMASIO, Antônio R. *O erro de Descartes: Emoção, Razão e o Cérebro Humano*. São Paulo: Companhia das Letras, 2020.

DAROS, Gabriel. Frances Haugen, do Facebook Papers: ‘Meta não protege eleições no Brasil’. *Tilt Uol*, [s.l.], 6 jul. 2022. Disponível em: <https://www.uol.com.br/tilt/noticias/redacao/2022/07/06/frances-haugen-do-facebook-papers-meta-nao-protege-eleicoes-no-brasil.htm>. Acesso em: 20 dez. 2022.

DELGADO, Richard; YUN, David. The Speech We Hate: First Amendment Totalism, the ACLU, and the Principle of Dialogic Politics. *Alabama Law*, Tuscaloosa: University of Alabama, v. 27, p 1281-1300, 1995.

DOUEK, Evelyn; FRANÇOIS, Camille. The Accidental Origins, Underappreciated Limits and Enduring Promises of Platform. Transparency Reporting About Informations Operations. *Journal of Online Trust and Safety*, [s.l.], out. 2021.

DOUEK, Evelyn. Governing Online Speech: From 'Posts-As-Trumps' to Proportionality and Probability. *Colum. L. Rev.*, [s.l.], v. 121, n. 759, ago. 2020. Disponível em: <https://ssrn.com/abstract=3679607>. Acesso em: 20 dez. 2022.

DWORKIN, Ronald. Why Speech must be free. In: _____. *Freedom's Law: the Moral Reading of the American Constitution*. Cambridge: Harvard University Press, 1996.

EIFERT, Martin. Lei Alemã para a Melhoria da Aplicação nas Redes Sociais (Lei de aplicação da Lei nas redes – NetzDG). Trad. Pedro Henrique Ribeiro e Revisão de Júlio Carvalho. In: ABOUD, Georges; CAMPOS, Ricardo; NERY JR., Nelson (Orgs.). *Fake News e Regulação*. 2. ed. São Paulo: Revista dos Tribunais, 2020.

EMPOLI, Giuliano da. *Os Engenheiros do Caos: como as Fake News, as Teorias da Conspiração e os Algoritmos estão sendo utilizados para disseminar o ódio, medo e influenciar as eleições*. Belo Horizonte: Vestígio, 2019.

ESTADOS UNIDOS DA AMÉRICA. *Miami Herald Publishing Co v. Tornillo*, 418, U.S., 241, 25 de junho de 1974. Disponível em: <https://supreme.justia.com/cases/federal/us/418/241/#tab-opinion-1950903>. Acesso em: 09 abr. 2022.

_____. *Red Lion Broadcasting Co., Inc v. FCC*, 395, U.S. 367, 9 de junho de 1969. Disponível em: <https://supreme.justia.com/cases/federal/us/395/367>. Acesso em: 9 abr. 2022.

EUA mudam regras da internet e favorecem grandes operadoras. *Tilt Uol*, [s.l.], 14 dez. 2017. Disponível em: <https://www.uol.com.br/tilt/noticias/redacao/2017/12/14/estados-unidos-decidem-pelo-fim-da-neutralidade-da-rede.htm>. Acesso em: 18 mar. 2022.

EUROPEAN COMMISSION. Digital Education Action Plan (2021-2027). *European Education Area*, [s.l.], 24 fev. 2022. Disponível em: <https://education.ec.europa.eu/focus-topics/digital-education/action-plan>. Acesso em 30 nov. 2022.

EXECUTIVE order on preventing online censorship. *White House*, [s.l.], 28 mai. 2020. Disponível em: <https://www.whitehouse.gov/presidential-actions/executive-order-preventing-online-censorship/>. Acesso em: 25 12. 2022.

EYAL, Nir. *Hooked: como construir produtos e serviços formadores de hábitos*. Trad. Edson Furmankiewicz. Cascavel: Alfacon, 2020.

FACEBOOK censura fotos de Sebastião Salgado de tribo indígena da Amazônia. *Folha de S. Paulo*, [s.l.], 4 set. 2018. Disponível em: <https://www1.folha.uol.com.br/poder/2018/09/facebook-censura-fotos-de-sebastiao-salgado-de-tribo-indigena-da-amazonia.shtml>. Acesso em 20/04/2022. Acesso em: 20 abr. 2022.

FACEBOOK disponibiliza ferramentas de transparência para anúncios políticos no Brasil. *GI*, [s.l.], 22 set. 2020. Disponível em: <https://g1.globo.com/economia/tecnologia/noticia/2020/09/22/facebook-disponibiliza-ferramentas-de-transparencia-para-anuncios-politicos-no-brasil.ghtml>. Acesso em: 16 nov. 2022.

FACEBOOK remove live em que Bolsonaro relaciona vacina à Aids. *Tecmundo*, [s.l.], 25 out. 2021. Disponível em: <https://www.tecmundo.com.br/redes-sociais/227472-facebook-remove-live-bolsonaro-relaciona-vacina-aids.htm>. Acesso em: 20 dez. 2022.

FADEL, Anna Laura Maneschy. *O discurso de ódio é um limite legítimo ao exercício da liberdade de expressão? Uma análise das teorias de Ronald Dworkin e Jeremy Waldron a partir da herança do liberalismo de John Stuart Mill*. Rio de Janeiro: Lumen Juris, 2018.

‘FAKE News’ é eleita palavra do ano e vai ganhar menção em dicionário britânico. *GI*, [s.l.], 2 nov. 2017. Disponível em: <https://g1.globo.com/educacao/noticia/fake-news-e-eleita-palavra-do-ano-e-vai-ganhar-mencao-em-dicionario-britanico.ghtml>. Acesso em: 24 mai. 2022.

FARAH, André. *Liberdade de expressão e remoção de conteúdo*. Rio de Janeiro: Lumen Juris, 2018.

FARKAS, Johan; SCHOU, Jannick. *Post-Truth, Fake News and Democracy: Mapping the Politics of Falsehood*. Londres: Routledge, 2020.

FISS, Joele; MCHANGAMA, Jacob. *The Digital Berlin Wall: How Germany (Accidentally) Created a Prototype for Global Online Censorship*. Copenhagen: Justitia, 2020. Mimeografado.

FISS, Owen. *A Ironia da Liberdade de Expressão: Estado, Regulação e Diversidade na Esfera Pública*. Trad. Gustavo Binenbojm e Caio Mário da Silva Pereira Neto. Rio de Janeiro: Renovar, 2005.

_____. *A ironia da liberdade de expressão: Estado, regulação e diversidade na esfera pública*. Trad. Gustavo Binenbojm e Caio Mario da Silva Pereira Neto. São Paulo: Editora FGV, 2022.

FOGG, B. J. *Persuasive technology: using computers to change what we think and do*. Burlington: Morgan Kaufmann, 2003.

FOX, Jonathan. The uncertain relationship between transparency and accountability. *Development in practice*, [s.l.], v. 17, n. 4-5, p. 663-671, 2007.

FRAGA, Fernando. Telegram assina acordo de adesão a programa do TSE. *Agência Brasil*, [s.l.], 25 mar. 2022. Disponível em: <https://agenciabrasil.ebc.com.br/justica/noticia/2022-03/telegram-assina-acordo-de-adesao-programa-do-tse>. Acesso em: 28 mar. 2022.

FRASER, Nancy. Rethinking the public sphere: a contribution to the critique of actually existing democracy. *Social Text*, [s.l.], n. 25/26, 1990.

FROOMKIN, A. Michael. The Internet as a Source of Regulatory Arbitrage. In: KAHIN, Brian; NESSON, Charles (Eds.). *Borders in Cyberspace*. Cambridge, MA: MIT Press, 1997.

FUNG, Brian; VOGUE, Ariane de. Texas asks Supreme Court to keep state social media censorship law in effect. *CNN*, [s.l.], 19 mai. 2022. Disponível em: <https://www.cnn.com/2022/05/18/tech/texas-supreme-court-social-media/index.html>. Acesso em: 01 jun. 2022.

GIACHETTA, André Zonaro. Atuação e Responsabilidade dos Provedores Diante das Fake News e da Desinformação. In: RAIS, Diogo (Coord.). *Fake News a conexão entre desinformação e o direito*. 2. ed. São Paulo: Thomsom Reuters Brasil, 2020.

GILLEPSIE, Tarleton. Content moderation, AI, and the question of scale. *Big Data & Society*, [s.l.], jul./dez. 2020.

GORZIZA, Amanda. Uma Mentira Sobre Vacina Enganou 46% dos Brasileiros. *Piauí*, [s.l.], 22 mar. 2021. Disponível em: <https://piaui.folha.uol.com.br/quase-metade-dos-brasileiros-acredita-em-boatos-sobre-vacinas/>. Acesso em: 20 abre. 2022.

GOVERNANCE. *Oversight Board*, [s.l.], [2018]. Disponível em: <https://oversightboard.com/governance/>. Acesso em 08 mar. 2022.

GRAHAM, David A. The Age of Reverse Censorship. *The Atlantic*, [s.l.], 26 jun. 2018. Disponível em: <https://www.theatlantic.com/politics/archive/2018/06/is-the-first-amendment-obsolete/563762/>. Acesso em: 16 dez. 2021.

GREGORIO, Giovanni de. The Rise of Digital Constitutionalism in the European Union. *International Journal of Constitutional Law*, [s.l.], v. 19, n. 1, p. 41-70, jun. 2021.

HABERMAS, Jürgen. *Mudança estrutural da esfera pública: investigações sobre uma categoria da sociedade burguesa*. Trad. Denilson Luís Werle. São Paulo: Editora Unesp, 2014.

HAGGART, Blayne; KELLER, Clara Iglesias. Democratic Legitimacy in global platform governance. *Telecommunications Policy*, [s.l.], v. 45, n. 6, jul. 2021. Disponível em: <https://doi.org/10.1016/j.telpol.2021.102152>. Acesso em: 20 dez. 2022.

HAI DT, Jonathan. *A Mente Moralista: por que as pessoas boas se separam por causa da política e da religião?* [S.l.]: Alta Cult, 2019.

_____; ROSE-STOCKWELL, Tobias. The dark psychology of social networks. *The Atlantic*, [s.l.], dez. 2019. Disponível em: <https://www.theatlantic.com/magazine/archive/2019/12/social-media-democracy/600763/>. Acesso em: 11 abr. 2022.

HARARI, Yuval Noah. *Homo Deus: uma breve história do amanhã*. Trad. Paulo Geiger. São Paulo: Companhia das Letras, 2016.

_____. *Homo Sapiens: uma breve história*. Trad. Janaína Marcoantonio. 25. ed. Belo Horizonte: L&PM, 2017.

HARTMANN, Ivar Alberto Martins. A crise dos precedentes no supremo: o caso dos precedentes sobre liberdade de expressão. *Revista Estudos Institucionais*, [s.l.], v. 6, n. 1, p. 109-128, jan./abr. 2020.

_____. A liberdade de expressão nas decisões de primeira instância do TJ-RJ. *Opin. Jur.*, Fortaleza, ano 18, n. 27, p. 13-44, jan./abr. 2020.

_____. Combinando Biblioteca de Anúncios com Checagem de Fatos para aumentar a transparência sobre a desinformação. *RDP*, Brasília, v. 18, n. 99, p. 460-485, jul./set. 2021.

_____; MONTEIRO, Júlia Iunes. Fake News no Contexto de Pandemia e Emergência Social: os Deveres e Responsabilidades das Plataformas de Redes Sociais na Moderação de Conteúdo Online: entre a Teoria e as Proposições Legislativas. *RDP*, Brasília, v. 17, n. 94, p. 388-414, jul./ago. 2020.

HAYEK, Friedrich von August. The Use of Knowledge in Society. *The American Economic Review*, [s.l.], v. 35, n. 4, set. 1945.

HEBERT SMITH FREEHILLS. *Race to Regulate: Online Harms*. [S.l.]: Herbert Smith Freehills, 2019. Disponível em: <https://www.herbertsmithfreehills.com/latest-thinking/online-harmful-content>. Acesso em: 23 mai. 2022.

HELDT, Amélie. Reading Between the lines and the numbers: an analysis of the first NetzDG reports. Internet Policy Review. *Journal on Internet Regulation*, [s.l.], v. 8, n. 2, jun. 2019.

HOFMANN, Jeanette; KATZENBACH, Christian; GOLLATZ, Kirsten. Between Coordination and regulation: Finding the governance in Internet Governance. *New Media & Society*, [s.l.], v. 19, n. 9, p. 1406-1423, mar. 2016.

HOLTER, Mikael. Facebook remove foto da 'menina do napalm' e gera protestos. *Tilt Uol*, [s.l.], 9 set. 2016. Disponível em: <https://www.uol.com.br/tilt/noticias/redacao/2016/09/09/facebook-remove-foto-da-menina-do-napalm-e-gera-protestos.htm>. Acesso em: 20 dez. 2022.

HONNETH, Axel. *Luta por Reconhecimento: A Gramática Moral dos Conflitos Sociais*. Trad. Luiz Repa. São Paulo: Editora 34, 2003.

HUMAN RIGHTS COUNCIL. *Report of the independent international fact-finding mission on Myanmar*. [S.l.]: HRC, 2018. Disponível em: https://www.ohchr.org/Documents/HRBodies/HRCouncil/FFM-Myanmar/A_HRC_39_64.pdf. Acesso em: 20 dez. 2022.

ISAAC, Mike. Facebook hampers do-it-yourself mask efforts. *The New York Times*, [s.l.], 5 abr. 2020. Disponível em: <https://www.nytimes.com/2020/04/05/technology/coronavirus-facebook-masks.html>. Acesso em: 20 abr. 2022.

JUSTIÇA dos EUA bloqueia lei do Texas que restringe empresas de redes sociais. *CNN Brasil*, [s.l.], 31 mai. 2022. Disponível em: <https://www.cnnbrasil.com.br/tecnologia/justica-dos-eua-bloqueia-lei-do-texas-que-restringe-empresas-de-redes-sociais/>. Acesso em: 01 jun. 2022.

KADRI, Thomas; KLONICK, Kate. Facebook v. Sullivan: public figures and newsworthiness in online speech. *Southern California Law Review*, [s.l.], v. 93, 2019.

KAHNEMAN, Daniel. *Rápido e Devagar: Duas Formas de Pensar*. São Paulo: Objetiva, 2018.

KAISER, Brittany. *Manipulados: Como a Cambridge Analytica e o Facebook invadiram a privacidade de milhões e botaram a democracia em xeque*. Trad. Roberta Clapp e Bruno Fiuza. Rio de Janeiro: Harper Collins, 2020.

KAKUTANI, Michiko. *A morte da Verdade: notas sobre a mentira na Era Trump*. Trad. André Czarnobai e Marcela Duarte. Rio de Janeiro: Intrínseca, 2018.

KELLER, Clara Iglesias. Don't Shoot the Message: Regulate Disinformation Beyond Content. *RDP*, Brasília, v. 18, n. 99, p. 486-515, jul./set. 2021.

_____. Policy by judicialisation: the institutional framework for intermediary liability in Brazil. *International Review of Law, Computers and Technology*, [s.l.], 13 jul. 2020. DOI: <https://doi.org/10.1080/13600869.2020.1792035>. Acesso em: 26 abr. 2022.

_____. *Regulação Nacional de Serviços na Internet: Exceção, Legitimidade e o Papel do Estado*. Rio de Janeiro: Lumen Juris, 2019.

KAMINSKI, Margot E. The right to explain, explained. *Berkeley Technology Law Journal*, [s.l.], v. 34, n. 189, 2019.

KEMP, Simon. Digital 2022: global overview report. *Datareportal*, [s.l.], 26 jan. 2022. Acesso em: <https://datareportal.com/reports/digital-2022-global-overview-report>. Acesso em: 07 fev. 2022.

KENDRICK, Leslie. Speech, Intent and the Chilling Effect. *William & Mary Law Review*, [s.l.], v. 54, n. 5, p. 1633-1691, 2013.

KING, Gary; PAN, Jennifer; ROBERTS, Margaret E. How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, Not Engaged Argument. *Am. Pol. Sci. Rev.*, [s.l.], v. 111, n. 484, 2017.

KLONICK, Kate. The Facebook Oversight Board: Creating an Independent Institution to Adjudicate Online Free Expression. *Yale Law Journal*, [s.l.], v. 129, n. 2418, jul. 2020.

_____. The New Governors: The People, Rules, and Processes Governing Online Speech. *Harv. L. Rev.*, [s.l.], v. 131, n. 1598, mar. 2017.

KOSINSKI, Michal, STIWELL, David; GRAEPEL, Thore. Private traits and attributes are predictable from digital records of human behavior. *PNAS*, [s.l.], v. 110, n. 15, mai. 2013.

LABORATÓRIO DE REGULAÇÃO ECONÔMICA. Adesão automática à previdência complementar é constitucional. *Jota*, [s.l.], 6 out. 2017. Disponível em: <https://www.jota.info/opiniao-e-analise/colunas/reg/adesao-automatica-a-previdencia-complementar-06102017>. Acesso em: 16 nov. 2022.

LAIDLAW, Emily. A Framework for Identifying Internet Information Gatekeepers. *International Review of Law, Computers & Technology*, [s.l.], v. 24, n. 3, 2010.

LEERSEN, Paddy; TWOREK, Heidi. *An analysis of Germany's NetzDG law*. Amsterdam: Transatlantic High Level Working Group on Content Moderation Online and Freedom of Expression, 2019.

LEITE, Fábio Carvalho. Por que juízes não aplicam o art. 19 do marco civil da Internet? *PLEB*, [s.l.], 22 jul. 2019. Disponível em: <https://www.plebpucc.com.br/post/por-que-ju%C3%ADzes-n%C3%A3o-aplicam-o-art-19-do-marco-civil-da-internet>. Acesso em: 26 abr. 2022.

LEMOS, Ronaldo; SOUZA, Carlos Affonso de. A Bill of Rights for the Brazilian Internet ('Marco Civil'): A Multistakeholder Policymaking Case. In: URS, Gasser; SCHULZ, Wolfgang (Orgs.). *Governance of Online Intermediaries: observations from a Series of National Case Studies*. Cambridge, MA: Berkman Center Research Publication, 2015.

_____. Marco Civil da Internet: construção e aplicação. Juiz de Fora: Editar, 2016.

LESSIG, Lawrence. *Code and Others Laws of Cyberspace, version 2.0*. Nova Iorque: Basic Books, 2006.

_____. The Law of the Horse: What Cyber Law Might Teach. *Harvard Law Review*, [s.l.], v. 113, p. 501-546, mar. 1999.

_____. The limits in Open Code: Regulatory Standards and Future of Net. *Berkeley Law Technology Journal*, [s.l.], n. 759, 1999.

LEVITSKY, Steven; ZIBLATT, Daniel. *Como as Democracias Morrem*. Trad. Renato Aguiar. Rio de Janeiro: Zahar, 2018.

MACCARTHY, Mark. Transparency Requirements for digital social media platforms: recommendations for policy makers and industry. *Transatlantic Working Group*, [s.l.], fev. 2020. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3615726. Acesso em: 01 nov. 2022.

MALÁSIA é o primeiro país do mundo a revogar lei de combate às fake news. *Consultor Jurídico*, [s.l.], 19 ago. 2028. Disponível em: <https://www.conjur.com.br/2018-ago-19/malasia-primeiro-pais-revogar-lei-combate-fake-news>. Acesso em: 20 mai. 2022.

MARDER, Ben et al. The extended chilling‘ effect of Facebook: The cold reality of ubiquitous social networking. *Computers in Human Behavior*, [s.l.], v. 60, 2016.

MARRAFON, Marco Aurélio; MEDON, Filipe. Importância da revisão humana das decisões automatizadas na Lei Geral de Proteção de Dados. *Consultor Jurídico*, [s.l.], 9 set. 2019. Disponível em: https://www.conjur.com.br/2019-set-09/constituicao-poder-importancia-revisao-humana-decisoes-automatizadas-lgpd#_ftn15. Acesso em: 14 nov. 2022.

MATHIESEN, Kay. Censura e Acesso à Expressão. In: SAMPAIO, José Adércio Leite (Coord.). *Liberdade de Expressão no Século XXI*. Belo Horizonte: Del Rey, 2016.

MAZZA, Luigi. O rosto da desinformação. *Piauí*, [s.l.], 10 set. 2021. Disponível em: <https://piaui.folha.uol.com.br/o-rosto-da-desinformacao/>. Acesso em 20 abr. 2022.

MCLUHAN, Marshall. *Os meios de comunicação como extensões do homem*. São Paulo: Pensamento-Cultrix, 2006.

MEDON, Filipe. *Inteligência Artificial e Responsabilidade Civil*. Salvador: Juspodium, 2020.

MEIKLEJONH, Alexander. *Free speech and its relation to self-government*. Nova Iorque: Harper & Brothers Publishers, 1948.

MELLO, Patrícia Campos. *A Máquina do Ódio: notas de uma repórter sobre Fake News e Violência Digital*. São Paulo: Companhia das Letras, 2020.

_____. Facebook permite conteúdo eleitoral classificado como falso, apesar de acordo com TSE. *Folha de S. Paulo*, [s.l.], 20 out. 2022. Disponível em: <https://www1.folha.uol.com.br/poder/2022/10/facebook-permite-conteudo-eleitoral-classificado-como-falso-apesar-de-acordo-com-tse.shtml>. Acesso em: 01 jan. 2022.

MENDES, Conrado Hubner. Freedom Without Responsibility: the Promise of Bolsonaro’s COVID-19 Denial. *Jus Cogens*, [s.l.], v. 3, p. 181-207, 2021.

MENDES, Gilmar Ferreira; BRANCO, Paulo Gustavo Gonet. *Curso de Direito Constitucional*. Editora Saraiva. 8. ed. Saraiva: São Paulo, 2013.

MENDES, Gilmar Ferreira; FERNANDES, Victor Oliveira. Constitucionalismo digital e jurisdição constitucional: uma agenda de pesquisa para o caso brasileiro. *Revista Brasileira de Direito*, Passo Fundo, v. 16, n. 1, p. 1-33, jan./abr. 2020.

MENDONÇA, José Vicente Santos de. *Direito Constitucional Econômico: a intervenção do Estado na economia à luz da razão pública e do pragmatismo*. 2. ed. Belo Horizonte: Fórum, 2018.

MILL, John Stuart. *Sobre a Liberdade*. Trad. Denise Bottmann. Porto Alegre: L&PM, 2019.

MILOŠ, Matija; PELIĆ, Toni. Constitutional Reasoning There and Back Again: The Facebook Oversight Board as a Source of Transnational Constitutional Advice. In: DE POORTER, J. et al. (Eds). *European Yearbook of Constitutional Law 2021: Constitutional Advice*, vol. 3. [s.l.]: Springer & T.M.C. Asser Press, The Hague, 2022, p. 197-22

MILTON, John. *Aeropagítica. Discurso sobre a Liberdade de Expressão*. Trad. Benedita Bettencourt. Coimbra: Almedina, 2009.

MITCHELL-BRADDON, David; WEST, Caroline. What is free speech? *The Journal of Political Philosophy*, [s.l.], v. 12, n. 4, p. 437–460, 2004.

MOUFFE, Chantal. *Por um populismo de esquerda*. São Paulo: Autonomia Literária, 2020.

MOUNK, Yascha. *O povo contra a democracia: porque nossa liberdade corre perigo e como salvá-la*. Trad. Cássio de Arantes Lete e Débora Landsberg. São Paulo: Companhia das Letras, 2019.

NEKMAT, Elmie. Nudge effect of fact-check alerts: Source influence and media skepticism on sharing of News information in social media. *Social media + Society*, [s.l.], v. 6, n. 1, 2020.

NELSON, J. L. Is ‘fake news’ a fake problem? *Columbia Journalism Review*, [s.l.], v. 31, 2017.

NEO, Ric. The Securitisation of Fake News in Singapore. *International Politics*, [s.l.], v. 57, p. 724-740, 2020. Disponível em: <https://link.springer.com/article/10.1057/s41311-019-00198-4>. Acesso em: 20 dez. 2022.

NGUI, Yantoultra. Malaysia Passes ‘Fake News’ Law That Critics Call an Assault on Speech. *The Wall Street Journal*, [s.l.], 3 abr. 2018. Disponível em: <https://www.wsj.com/articles/malaysia-passes-fake-news-law-that-critics-call-an-assault-on-speech-1522757225>. Acesso em 20 mai. 2022.

NITRINI, Rodrigo Vidal. *Liberdade de Expressão nas Redes Sociais*. São Paulo: Dialética, 2021.

NOVINSKY, Anitta. Os regimes totalitários e a censura. In: TUCCI, Maria Luiza (Org.). *Minorias Silenciadas: História da Censura no Brasil*. 2. ed. São Paulo: Editora da Universidade de São Paulo, 2020.

OGUS, Anthony. Rethinking Self-Regulation. *Oxford Journal of Legal Studies*, [s.l.], v. 15, n. 1, 1995.

ORGANIZAÇÃO DOS ESTADOS AMERICANOS et al. *Declaración conjunta sobre la independencia y la diversidad de los medios de comunicación en la era digital*. [S.l.]: [s.n.], 2018. Disponível em: https://www.ohchr.org/sites/default/files/Documents/Issues/Opinion/JointDeclaration2May2018_SP.pdf. Acesso em: 29 mar. 2022.

ORGANIZAÇÃO PARA A COOPERAÇÃO E DESENVOLVIMENTO ECONÔMICO. *The Economic and Social Role of Internet Intermediaries*. [S.l.]: OECD, 2010.

ORWELL, George. *Looking Back on the Spanish War*. [S.l.]: Orwell Foundation, 1943. Disponível em: <https://www.orwellfoundation.com/the-orwell-foundation/orwell/essays-and-other-works/looking-back-on-the-spanish-war/>. Acesso em: 25 mai. 2022.

OSÓRIO, Aline. *Liberdade de expressão e Direito Eleitoral*. Belo Horizonte: Fórum, 2017.

OVERSIGHT BOARD. Former President Trump's suspension. *Oversight Board*, [s.l.], 7 jan. 2021. Disponível em: <https://oversightboard.com/decision/FB-691QAMHJ/>. Acesso em: 06 jun. 2022.

PAIVA, Letícia. PL das Fake News pode tornar plataformas menos seguras, critica Google. *Jota*, [s.l.], 11 mar. 2022. Disponível em: <https://www.jota.info/coberturas-especiais/liberdade-de-expressao/pl-das-fake-news-pode-tornar-plataformas-menos-seguras-critica-google-11032022>". Acesso em: 21 mar. 2022.

PAMPLONA, Danielle Anne. O conteúdo do direito da liberdade de expressão, o discurso de ódio e a resposta democrática. *Revista Brasileira de Direito*, Passo Fundo, v. 14, n. 1, p. 297-316, jan./abr., 2018.

PARISIÉR, Eli. *The Filter Bubble. What the internet is Hiding from you*. Nova Iorque: The Penguin Press, 2011.

PASQUALE, Frank. *The Black Box Society: The Secret Algorithms that Control Money and Information*. Cambridge, MA: Harvard University Press, 2015.

PELA 1ª vez, rede social é mais citada que TV como fonte de notícia no Brasil. *Poder 360*, [s.l.], 16 jun. 2020. Disponível em: <https://www.poder360.com.br/midia/pela-1a-vez-rede-social-e-mais-citada-que-tv-como-fonte-de-noticia-no-brasil/>. Acesso em: 07 fev. 2022.

PENNEY, Jonathon W. Internet Surveillance, Regulation, and Chilling Effects Online: A Comparative Case Study. *Internet Policy Review*, [s.l.], v. 6, n. 2, mai. 2017. Mimeografado.

PEREIRA, Caio Mário da Silva. *Responsabilidade civil*. 5. ed. Rio de Janeiro: Forense, 1994.

PGR questiona resolução editada pelo TSE a menos de duas semanas do segundo turno. *MPF*, [s.l.], 21 out. 2022. Disponível em: <https://www.mpf.mp.br/pgr/noticias-pgr/pgr-questiona-resolucao-editada-pelo-tse-ha-menos-de-duas-semanas-do-segundo-turno>. Acesso em: 24 nov. 2022.

POGREBINSCHI, Thamy. *Pragmatismo: Teoria Social e Política*. Rio de Janeiro: Relume Dumará, 2005.

PONTES, João Gabriel Madeira. *Democracia militante em tempos de crise*. Rio de Janeiro: Lumen Juris, 2020.

POPPER, Karl. *A Sociedade Aberta e Seus Inimigos*. Trad. Milton Amado. Belo Horizonte: Ed. Itatiaia; São Paulo: Ed. da Universidade de São Paulo, 1974.

POSNER, Richard. The Economic Approach to Law. *Chicago Unbound*, Chicago: University of Chicago Law School, v. 53, 1975.

POST, Robert. Participatory Democracy and Free Speech. *Virginia Law Review*, [s.l.], v. 97, n. 3, 2011.

POZEN, David; KHAN, Lina. A Skeptical View of Information Fiduciaries. *Harvard Law Review*, [s.l.], v. 133, n. 497, dez. 2019.

PRESIDÊNCIA DA REPÚBLICA. *Mensagem nº 427, de 1º de setembro de 2021*. Brasília, DF: Presidência da República, 2021. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2019-2022/2021/Msg/VEP/VEP-427.htm#:~:text=%E2%80%9CA%20proposi%C3%A7%C3%A3o%20legislativa%20estabelece%20como,para%20disseminar%20fatos%20que%20sabe. Acesso em: 26 mai. 2022.

PRINCÍPIOS de Manila sobre a responsabilidade dos intermediários. [S.l.]: EFF, 2015. Disponível em: https://www.eff.org/files/2015/07/02/manila_principles_1.0_pt.pdf. Acesso em: 18 mar. 2022.

RAWLS, John. *A Theory of Justice*. Cambridge: Harvard University Press, 1999.

_____. *O Liberalismo Político*. [S.l.]: Presença, 1997.

REINDENBERG, Joel R. Lex informatica: the formulation of information policy rules through technology. *Texas Law Review*, Texas, v. 76, n. 3, 1998, p. 553-584, 1998.

RIBEIRO, Amanda. Falas de Bolsonaro sobre 'tratamento precoce' ferem regras de redes sociais, mas seguem no ar. *Aos Fatos*, [s.l.], 28 jun. 2021. Disponível em: <https://www.aosfatos.org/noticias/falas-de-bolsonaro-sobre-tratamento-precoce-ferem-regras-de-redes-sociais-mas-seguem-no-ar/>. Acesso em: 17 dez. 2021.

RIBEIRO, Leonardo Coelho. O direito administrativo como caixa de ferramenta e suas estratégias. *Revista de Direito Administrativo*, Rio de Janeiro, v. 272, p. 209-249, mai./ago. 2016.

RIDKIN, Kim; KENNEY, Patrick J.; WINTERSIECK, Amanda. Liar, liar, pants on fire: how fact-checking influences citizens' reactions to negative advertising. *Political Communication*, [s.l.], v. 32, n. 1, p. 139, 2015.

RÚSSIA criminaliza cobertura da guerra na Ucrânia e bane BBC, Twitter e Facebook. *Época Negócios*, [s.l.], 4 mar. 2022. Disponível em: <https://epocanegocios.globo.com/Mundo/noticia/2022/03/epoca-negocios-russia-criminaliza-cobertura-da-guerra-na-ucrania-e-bane-bbc-twitter-e-facebook.html>. Acesso em: 28 mar. 2022.

SADDY, André. *Regulação Estatal, Autorregulação Privada e Códigos de Condutas e Boas Práticas*. Rio de Janeiro: Lumen Juris, 2019.

SARLET, Ingo Wolfgang. Constituição e proporcionalidade: o direito penal e os direitos fundamentais entre proibição de excesso e proibição de insuficiência. *Revista Brasileira de Ciências Criminais*, [s.l.], n. 47, mar./abr. 2004.

SARMENTO, Daniel. A liberdade de expressão e o problema do hate speech. *Revista de Direito do Estado*, Rio de Janeiro: Renovar, n. 4, p. 53-105, out./dez. 2006. Mimeografado.

_____. *Dignidade da Pessoa Humana: conteúdo, trajetórias e metodologia*. [S.l.]: [s.n.], 2021. No prelo.

_____. *Direitos Fundamentais e Relações Privadas*. Rio de Janeiro: Lumen Juris, 2004.

_____. Liberdade de expressão, pluralismo e o papel promocional do Estado. *Revista Diálogo Jurídico*, [s.l.], n. 16 mai./ago. 2007.

_____; PONTES, João Gabriel Madeira. Democracia Militante e Imunidade Material dos Parlamentares: Limites Constitucionais aos Discursos de Deputados e Senadores. *Revista da AJURIS*, Porto Alegre, v. 47, n. 149, dez. 2020.

SCHAUER, Frederick. Fear, Risk and the First Amendment: Unraveling the Chilling Effect. *Boston University Law Review*, Boston, v. 58, 1978.

_____. Liberdade de Expressão em Decisões na Europa e nos Estados Unidos: Um Estudo de Caso na Arquitetura Constitucional Comparada In: SAMPAIO, José Adércio Leite (Coord.). *Liberdade de Expressão no Século XXI*. Belo Horizonte: Del Rey, 2016.

SCHREIBER, Anderson. Cyberbullying: Responsabilidade civil efeitos na família. *Carta Forense*, [s.l.], [2018]. Disponível em: www.cartaforense.com.br. Acesso em: 28 mar. 2022.

_____. De São Conrado para o STF: uma nota sobre a liberdade de expressão. *Jota*, [s.l.], 3 ago. 2021. Disponível em: <https://www.jota.info/opiniao-e-analise/colunas/coluna-do-anderson-schreiber/de-sao-conrado-para-o-stf-uma-nota-sobre-a-liberdade-de-expressao-03082021>. Acesso em 23 mai. 2022.

_____. Liberdade de expressão e Tecnologia. In: SCHREIBER, Anderson; MORAES, Bruno Terra de; TEFFÉ, Chiara Spadaccini (Coords.). *Direito e Mídia. Tecnologia e Liberdade de Expressão*. São Paulo: Foco, 2020.

SCHREIBER, Mariana. A controversa lei alemã que inspira o projeto de lei das Fake News. *O Globo*, [s.l.], 26 ago. 2020. Disponível em: <https://oglobo.globo.com/epoca/brasil/a-controversa-lei-alema-que-inspira-projeto-de-lei-das-fake-news-1-24606576>. Acesso em: 26 mai. 2022.

SCHULZ, Wolfgang. Roles and Responsibilities of Information Intermediaries: fighting misinformation as a test case for a human rights-respecting governance of social media platforms. *A Hoover Institute Essay*, [s.l.], n. 1904, 2019.

SECTION 230 of the Communications Decency Act. *EFF*, [s.l.], [2015]. Disponível em: <https://www.eff.org/issues/cda230>. Acesso em: 18 mar. 2022.

SILVA, Bruno Bomquimpani. *A autorregulação: delineamento de um modelo policêntrico de regulação jurídica*. 2010. Dissertação (Mestrado em Direito Público) – Faculdade de Direito, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2010.

SILVA, Bruno Lemos Morrison da. *O paradoxo do direito de manifestação: a regulação que liberta*. 2016. 252 f. Dissertação (Mestrado em Direito do Estado) – Faculdade de Direito, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2016.

SILVA, Virgílio Afonso da. *A constitucionalização do direito: os direitos fundamentais nas relações entre particulares*. São Paulo: Malheiros, 2005.

SILVERMAN, Craig. Facebook Banned Mask Ads. They're Still Running. *BuzzFeed.News*, [s.l.], 13 mai. 2020. Disponível em: <https://www.buzzfeednews.com/article/craigsilverman/facebook-mask-ads-ban-zestads-coronavirus>. Acesso em: 20 abr. 2022.

SIMON, Herbert. Designing Organizations for in Information Rich-World. In: GREENBERGER, M. (Ed.). *Computers, Communication, and the public interest*. Baltimore, MD: The Johns Hopkins Press, 1971. Disponível em: <https://digitalcollections.library.cmu.edu>. Acesso em: 13 abr. 2022.

SOPRANA, Paula. Ação do Facebook contra rede ligada aos Bolsonaros também atingiu esquerda latina. *Folha de S. Paulo*, [s.l.], 10 jul. 2020. Disponível em: <https://www1.folha.uol.com.br/poder/2020/07/acao-do-facebook-contra-rede-ligada-aos-bolsonaros-tambem-atingiu-esquerda-latina.shtml>. Acesso em: 17 dez. 2021.

SOUZA, Carlos Affonso de. Análise: inconstitucional, MP de Bolsonaro cria ‘Ministério da Mentira’. *Tilt Uol*, [s.l.], 7 set. 2021. Disponível em: <https://www.uol.com.br/tilt/colunas/carlos-affonso-de-souza/2021/09/07/medida-provisoria-de-bolsonaro-sobre-redes-sociais-e-inconstitucional.htm>. Acesso em: 28 mar. 2022.

SOUZA NETO, Cláudio Pereira. *A Democracia em Crise no Brasil: valores constitucionais, antagonismo político e dinâmica institucional*. São Paulo: Contracorrente, 2020.

STANLEY, Jason. *Como Funciona o Fascismo: a política do “Nós” e “Eles”*. 2. ed. Porto Alegre: L&PM, 2019.

SUMPTER, David. *Dominado pelos números. Do Facebook e Google às fake news. Os algoritmos que controlam nossas vidas*. São Paulo: Bertrand, 2019.

SUNDFELD, Carlos Ari. Princípio é preguiça? In: _____. *Direito Administrativo para Céticos*. 2. ed. São Paulo: SBDP, Malheiros, 2014.

SUNSTEIN, Cass. Governing By Algorithm? No Noise and (Potentially) Less Bias. Preliminary Draft. *Duke Law Journal*, [s.l.], set. 2021.

_____. *Liars: Falsehoods and Free Speech in Age of Deception*. Nova Iorque: Oxford University Press, 2021.

_____. *#Republic: Divided Democracy in the Age of Social Media*. Nova Jersey: Princeton Press, 2017.

_____. *Republic.com 2.0*. Princeton, Oxford: Princeton University Press, 2009.

_____. *Too Much Information: understanding what you don't want to know*. Cambridge, MA: The Mit Press, 2020.

_____. *Why nudges? The Politics of Libertarian Paternalism*. New Haven, Londres: Yale Press, 2015.

SUSSKIND, Jamie. Future Politics. *Living Together in a World Transformed by Tech*. Nova Iorque: Oxford University Press, 2018. pt. 4.

SUZOR, Nicolas. Digital Constitutionalism: Using the Rule of Law to Evaluate the Legitimacy of Governance by Platforms. *Social Media + Societ*, [s.l.], jul./set. 2018.

_____. *Lawless: the secret rules that govern our digital lives*. [S.l.]: Cambridge University Press, 2019.

TAVEIRA, Christiano de Oliveira. *Democracia e Pluralismo na esfera comunicativa: uma proposta de reformulação do papel do Estado na garantia da Liberdade de Expressão*. 2010. Tese (Doutorado em Direito) – Faculdade de Direito, Universidade do Rio de Janeiro, Rio de Janeiro, 2010.

TEFFÉ, Chiara Antonia Spadaccini. 2015. A Responsabilidade Civil do Provedor de Aplicações de Internet Pelos Danos Decorrentes do Conteúdo Gerado por Terceiros, de Acordo com o Marco Civil da Internet. *Revista Fórum De Direito Civil*, Belo Horizonte, ano 4, n. 10, set./dez. 2015. Disponível em: <https://www.editoraforum.com.br/wp-content/uploads/2015/12/A-responsabilidade-civil-do-provedor-de-aplicacoes-de-internet.pdf>. Acesso em: 26 abr. 2022.

THALER, Richard. Nudge, not sludge. *Science*, [s.l.], v. 361, n. 6401, ago. 2018. Disponível em: <https://science.sciencemag.org/content/361/6401/431>. Acesso em: 14 jun. 2022.

THE facebook files. *The Wall Street Journal*, [s.l.], [2021]. Disponível em: <https://www.wsj.com/articles/the-facebook-files-11631713039>. Acesso em: 21 abre. 2022.

TRUMP sugere atirar em manifestantes, e Twitter inclui aviso sobre glorificação da violência. *Folha de S. Paulo*, [s.l.], 29 mai. 2020. Disponível em: <https://www1.folha.uol.com.br/mundo/2020/05/trump-sugere-atirar-em-manifestantes-e-twitter-inclui-aviso-sobre-glorificacao-da-violencia.shtml>. Acesso em: 11 nov. 2022.

TUSHNET, Rebecca. Power without responsibility: intermediaries and the first amendment. *The George Washington Law Review*, [s.l.], v. 76, 2008. Disponível em: <https://www.gwlr.org/wp-content/uploads/2012/08/76-4-Tushnet.pdf>. Acesso em: 14 nov. 2022.

TUTT, Andrew. An FDA For Algorithms. *Admin. L. Rev.*, [s.l.], v. 83, 2017. Disponível em: <https://ssrn.com/abstract=2747994>. Acesso em: 06 dez. 2022.

UNIÃO EUROPEIA. *Comunicação da Comissão ao Parlamento Europeu, ao Conselho, ao Comitê Econômico e Social Europeu e a ao Comitê das Regiões. Inteligência Artificial para a Europa*. [S.l.]: UE, 2018, p.1. Disponível em:

<http://data.consilium.europa.eu/doc/documents/ST-8507-2018-INIT/pt/pdf>. Acesso em: 11 abr. 2022.

UN REFUGIADO sirio pierde su batalla judicial contra Facebook por un ‘selfie’ con Merkel. *20 minutos*, [s.l.], 7 mar. 2017. Disponível em: <https://www.20minutos.es/noticia/2978420/0/refugiado-sirio-batalla-facebook-alemania/foto-montaje-terrorista/>. Acesso em: 15 jun. 2022.

UNITED NATIONS. *Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*. [S.l.]: UN, 2017. Disponível em: <https://digitallibrary.un.org/record/1304394>. Acesso em: 15 jun. 2022.

VALFRÉ, Vinícius; GALZO, Wesley; FRAZÃO, Felipe. Live que distorce informações sobre urnas é assistida por 415 mil pessoas. *Uol Notícias*, [s.l.], 4 nov. 2022. Disponível em: <https://noticias.uol.com.br/ultimas-noticias/agencia-estado/2022/11/04/live-que-distorce-informacoes-sobre-urnas-e-assistida-por-415-mil-pessoas.htm>. Acesso em: 30 nov. 2022.

VAN MILL, David. Free speech. In: ZALTA, Edward (Ed.). *The Stanford Encyclopedia of Philosophy*. [S.l.]: [s.n.], 2002. Disponível em: <http://plato.stanford.edu/archives/win2002/entries/freedom-speech/>. Acesso em: 20 dez. 2022.

VESTING, Thomas. A Mudança da Esfera Pública pela Inteligência Artificial. Trad. Pedro Henrique Ribeiro. In: ABOUD, Georges; CAMPOS, Ricardo; NERY JR., Nelson (Orgs.). *Fake News e Regulação*. 2. ed. São Paulo: Revista dos Tribunais, 2020. p. 193-212.

VOLOKH, Eugene. The Mechanisms of Slippery Slope. *Harvard Law Review*. Vol.116. nº4. pp.1026-1137.2003.

WALDRON, Jeremy. *The harm of hate speech*. Cambridge: Harvard University Press, 2012.

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

WISCHMEYER, Thomas. “What is illegal offline is also illegal online”: The German Network Enforcement Act 2017. *Bilyana Petkova & Tuomas Ojanen: Fundamental Rights Protection Online: The Future Regulation of Intermediaries*, [s.l.], out. 2018. Mimeografado.

WU, Tim. Is Internet Exceptionalism Dead? In: SZOKA, Berim; MARCUS, Adam (Eds.). *The Next Digital Decade: Essays on the Future of Internet*. Washington, DC: TechFreedom, 2010.

_____. Is the First Amendment Obsolete? *Michigan Law Review*, [s.l.], v. 117, p. 547, 2018.

ZUBOFF, Shoshana. *A Era do Capitalismo de Vigilância: a luta por um futuro humano na nova fronteira do poder*. Rio de Janeiro: Intrínseca, 2020.

ZUCKERBERG, Mark. The Internet needs new rules. Let’s start in these four áreas. *The Washington Post*, [s.l.], 30 mar. 2019. Disponível em: https://www.washingtonpost.com/opinions/mark-zuckerberg-the-internet-needs-new-rules-lets-start-in-these-four-areas/2019/03/29/9e6f0504-521a-11e9-a3f7-78b7525a8d5f_story.html. Acesso em: 10 nov. 2022.