

**UNIVERSIDADE ESTADUAL PAULISTA “JÚLIO DE MESQUITA FILHO”
FACULDADE DE CIÊNCIAS HUMANAS E SOCIAIS
CÂMPUS DE FRANCA**

ARTUR MARCHIONI

**A RESPONSABILIZAÇÃO DAS PLATAFORMAS DIGITAIS SOBRE O
CONTEÚDO POSTADO POR TERCEIROS: possibilidades no enfrentamento à
disseminação de discurso de ódio e
fake news.**

Franca-SP
2023

PROGRAMA DE PÓS-GRADUAÇÃO EM DIREITO

ARTUR MARCHIONI

A RESPONSABILIZAÇÃO DAS PLATAFORMAS DIGITAIS SOBRE O CONTEÚDO POSTADO POR TERCEIROS: possibilidades no enfrentamento à disseminação de discurso de ódio e fake news.

Dissertação apresentada ao Programa de Pós-Graduação em Direito da Faculdade de Ciências Humanas e Sociais– Universidade Estadual Paulista Júlio Mesquita Filho-UNESP, como parte dos requisitos para obtenção do título de Mestre em Direito.

Área de Concentração: Cidadania Civil e Política e Sistemas Normativos

Linha de pesquisa: Democracia e arranjos jurídico-institucionais

Orientador: Professor Associado Murilo Gaspardo

M317r Marchioni, Artur

A RESPONSABILIZAÇÃO DAS PLATAFORMAS DIGITAIS SOBRE O
CONTEÚDO POSTADO POR TERCEIROS:

possibilidades no enfrentamento à disseminação de discurso de ódio e
fake news / Artur Marchioni. -- Franca, 2023

236 p.

1. Fake news. 2. Discurso de ódio. 3. Capitalismo de vigilância. 4.
Liberdade de expressão. 5. Plataformas digitais. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp.
Biblioteca da Faculdade de Ciências Humanas e Sociais, Franca. Dados
fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

Impacto potencial desta pesquisa

A presente pesquisa tem como objeto a responsabilização das plataformas digitais sobre o conteúdo postado por terceiros enquanto um mecanismo de combate à disseminação perniciosa de discurso de ódio e *fake news* nas plataformas digitais. o objetivo é ampliar a compreensão científica sobre a comunicação digital e avaliar o potencial regulatório de medidas como a responsabilização das plataformas.

Potential impact of this research

This research focuses on the accountability of digital platforms for content posted by third parties as a mechanism to combat the harmful spread of hate speech and fake news on digital platforms. The aim is to enhance scientific understanding of digital communication and assess the regulatory potential of measures such as platform accountability.

ARTUR MARCHIONI

**A RESPONSABILIZAÇÃO DAS PLATAFORMAS DIGITAIS SOBRE O CONTEÚDO
POSTADO POR TERCEIROS: possibilidades no enfrentamento à disseminação de
discurso de ódio e *fake news*.**

Dissertação apresentada à Faculdade de Ciências Humanas e Sociais da Unesp – Campus de Franca, como requisito parcial para obtenção do título de Mestre em Direito. Área de concentração: Sistemas Normativos e Fundamentos da Cidadania

BANCA EXAMINADORA

Presidente:

Prof. Dr. Murilo Gaspar do, UNESP

1º Examinador:

Prof. Dr. Eugênio Bucci, USP

2º Examinador:

Prof. Dr. Rubens Beçak, USP

Este trabalho é dedicado a todos aqueles que, ao longo de minha trajetória acadêmica, me inspiraram e tornaram possível que eu pudesse dedicar parte de meu tempo e esforço à investigação científica.

Dedico especialmente à minha família que sempre incentivou e celebrou os esforços despendidos.

AGRADECIMENTOS

Agradeço ao meu orientador, com quem tenho a felicidade de partilhar meu caminho na pesquisa científica desde a graduação, pela dedicação e ensinamentos, bem como pelas reflexões sobre a Democracia e a comunicação digital, proporcionadas por esta dissertação. Tenho a honra e privilégio de trabalhar ao lado de um pesquisador dedicado e comprometido com a Ciência Jurídica.

Agradeço aos colegas e professores da Pós-Graduação em Direito da Unesp-Campus de Franca, especialmente às minhas colegas de mestrado Letícia e Júlia, pelas diversas oportunidades em que discutimos e debatemos os objetos de nossas pesquisas. Esse trabalho deve muito às vossas contribuições.

Agradeço ainda àqueles com quem compartilhei experiências e leituras ao longo desta dissertação, especialmente por meio de seminários científicos em que comuniquei os andamentos e desafios deste trabalho, e onde tive a oportunidade de conhecer autores e ideias que mudaram o curso desta pesquisa.

Agradeço à Unesp, que me acolheu desde a graduação, por meio de seus servidores, que foram sempre solícitos ao longo desta caminhada.

Agradeço à minha família, pelo apoio e confiança no meu trabalho, aos meus amigos, companheiros de vida, que a tornam um pouco mais leve todos os dias e à Ágata, que me inspira a buscar sempre o melhor em mim.

Sem vocês, não apenas este trabalho seria inviável, como não haveria sentido.

“Caminante, son tus huellas
el camino y nada más;
Caminante, no hay camino,
se hace camino al andar.
Al andar se hace el camino,
y al volver la vista atrás
se ve la senda que nunca
se ha de volver a pisar.
Caminante no hay camino
sino estelas en la mar.”
(Machado, 1999)

RESUMO

A estabilidade das instituições democráticas vem sendo ameaçada, nos últimos anos, pelo acirramento das tensões nas relações sociopolíticas, econômicas e culturais que se institucionalizaram sob a égide do Estado Democrático de Direito. Nesse panorama, a comunicação pelas mídias digitais vem sendo apontada como um dos fatores que alimentam a polarização, por meio da disseminação de *fake news* e discurso de ódio, enviesando e fragmentando o debate público. A presente pesquisa guiou-se pela hipótese de que a isenção de responsabilidade das plataformas digitais, prevista no Marco Civil da Internet, teve como efeito não esperado ou pretendido a ampla disseminação de conteúdos perniciosos no âmbito das mídias digitais. Ao longo deste trabalho, buscamos compreender aspectos sobre a comunicação digital e sua interferência na formação da opinião pública e, especialmente, o papel das redes sociais enquanto um espaço que é, paradoxal e concomitantemente, público e privado, e as implicações do capitalismo de vigilância sobre a Democracia Liberal-Representativa. Em um segundo momento, recorreremos ao debate sobre o princípio constitucional da liberdade de expressão, enquanto institucionalização, pelo Direito, de um dos fundamentos basilares para a efetivação de uma esfera pública democrática. Verificamos, assim, que o uso dos dados dos usuários para influenciar a comunicação social, por meio de uma psicopolítica dado-pulsional, constitui uma nova forma de ameaça à autonomia e à autodeterminação democrática, atuando sobre a formação da vontade, de forma que a possibilidade de regulamentação dos conteúdos veiculados pelas plataformas digitais não constitui, por si só, afronta ao princípio da liberdade de expressão, mas que, ao contrário, mostra-se indispensável para a promoção de um debate público efetivamente democrático. Por fim, após a revisão bibliográfica que constitui o fundamento teórico deste trabalho, recorreremos à aplicação sistemática do método de mapeamento de arranjos jurídico-institucionais, para avaliar os impactos da isenção de responsabilidade das plataformas digitais sobre o conteúdo postado por terceiros, estabelecido pelo art. 19 do Marco Civil da Internet, na ampla disseminação de *fake news* e discurso de ódio. Conclui-se que, desde 2014, ainda sob a égide da isenção de responsabilidade do Marco Civil da Internet, há um movimento de diversos atores envolvidos na comunicação digital, como Estados, Organizações Internacionais, sociedade civil, usuários e as próprias plataformas, para ampliar a moderação de conteúdo nas mídias digitais, de maneira a combater a disseminação sistemática de conteúdos perniciosos, que tem se materializado tanto pela ação jurídico-institucional propriamente dita, no Brasil especialmente por meio do Programa de

Enfrentamento à Desinformação do Tribunal Superior Eleitoral, quanto pela atuação das plataformas, por uma espécie de governança privada dos discursos.

Palavras-chave: *Fake news*; discurso de ódio; capitalismo de vigilância; liberdade de expressão; plataformas digitais.

ABSTRACT

The stability of democratic institutions has been threatened in recent years by the intensification of tensions in sociopolitical, economic, and cultural relations that have become institutionalized under the aegis of the Democratic Rule of Law. In this context, communication through digital media has been identified as one of the factors fueling polarization through the spread of fake news and hate speech, skewing and fragmenting public discourse. This research is guided by the hypothesis that the liability exemption for digital platforms, as outlined in the Marco Civil da Internet (Brazilian Internet Civil Rights Framework), had the unintended consequence of widespread dissemination of harmful content in the digital media sphere. Throughout this work, we sought to understand how digital communication interferes with the formation of public opinion, especially the role of social networks as a space that is paradoxically both public and private, and the implications of surveillance capitalism on Liberal-Representative Democracy. In a second phase, we delved into the debate on the constitutional principle of freedom of expression as the institutionalization, by law, of one of the fundamental pillars for the realization of a democratic public sphere. We found that the use of user data to influence social communication through a psycho-politics of data-driven impulses constitutes a new form of threat to democratic autonomy and self-determination, influencing the formation of will. The possibility of regulating content on digital platforms is not inherently a violation of the principle of freedom of expression; instead, it proves essential for the promotion of a genuinely democratic public discourse. Finally, after the literature review that forms the theoretical foundation of this work, we applied a systematic mapping method of legal-institutional arrangements to assess the impacts of the liability exemption for digital platforms on third-party content, as established by Article 19 of the Marco Civil da Internet, in the widespread dissemination of fake news and hate speech. It is concluded that since 2014, still under the exemption provided by the Marco Civil da Internet, various actors in digital communication, including states, international organizations, civil society, users, and the platforms themselves, have been moving to enhance content moderation on digital media. This is aimed at combating the systematic spread of harmful content, manifested both through legal and institutional actions, in Brazil especially through the Superior Electoral Court's Program to Combat Disinformation, and through the platforms' efforts in a kind of private governance of discourses.

Keywords: Fake news; hate speech; surveillance capitalism, freedom of expression; digital platforms.

SUMÁRIO

INTRODUÇÃO.....	14
CAPÍTULO 1- COMUNICAÇÃO DIGITAL, CAPITALISMO DE VIGILÂNCIA E EPISTEMOLOGIA TRIBAL	21
1.1. Da sociedade em rede à infodemia: relações derivadas do capitalismo e informacionalismo	21
1.2. Auto-comunicação de massa, ascensão das redes sociais e a formação de autoidentidades mediadas.....	32
1.3. A sociedade da vigilância: da racionalidade discursiva à racionalidade digital	45
1.4. Os espaços públicos mediados e a epistemologia tribal.....	56
1.5. <i>Fake news</i> e discurso de ódio na infocracia: filtros-bolha e câmaras-de-eco.....	67
1.6. A mídia-hiperpartidária desvelada pelo discurso de ódio	83
1.7. <i>Fake news</i> : uma proposta de conceito.....	95
CAPÍTULO 2- DEMOCRACIA, LIBERDADE DE EXPRESSÃO E GOVERNANÇA DOS DISCURSOS NAS MÍDIAS DIGITAIS.....	99
2.1. A racionalidade digital e a fragmentação da esfera pública: impactos no princípio democrático da liberdade de expressão.....	101
2.1.1. Democracia em cheque: repensar a liberdade de expressão na era digital	109
2.1.2. A liberdade que limita: o princípio da liberdade de expressão tolhido pelo excesso de vigilância	113
2.1.3. Regulamentando a tecnologia: um desvio de rota pautado na justiça	117
2.2. Prática institucional: a governança digital no início dos anos 2020	121
2.3. <i>Fake news</i> e discurso de ódio: o que e como moderar? Um esboço	132
2.3.1. Reféns da governança privada: entre a cruz e a espada	141
2.3.2. Em busca do meio termo.....	146
2.4. <i>Fake news</i> , discurso de ódio e o caso brasileiro	151
CAPÍTULO 3- MAPEAMENTO DE ARRANJO-JURÍDICO INSTITUCIONAL: A REGULAMENTAÇÃO DA COMUNICAÇÃO DIGITAL NO BRASIL A PARTIR DO MARCO CIVIL DA INTERNET.....	163
3.1. Algumas palavras sobre o método	163
3.2. Periodização.....	168
3.2.1. O Caso Snowden e o Marco Civil da Internet	168
3.2.2. O escândalo da Cambridge Analytica e o Programa Permanente de Enfrentamento à Desinformação.....	170
3.3. Seleção e descrição de Arranjos Jurídico-Institucionais	172

3.4. Documentação jurídica.....	172
3.5. Contexto e atores políticos	175
3.5.1. O Marco Civil da Internet.....	175
3.5.2. O Programa de Enfrentamento à Disseminação de Desinformação com Foco nas eleições de 2020	179
3.5.3. O Programa Permanente de Enfrentamento à Desinformação no âmbito da Justiça Eleitoral	180
3.6. Análise das motivações, efeitos esperados e efeitos obtidos.....	182
3.6.1. O Marco Civil da Internet.....	182
3.6.2. O Programa de Enfrentamento à Disseminação de Desinformação com Foco nas eleições de 2020	189
3.6.3. O Programa Permanente de Enfrentamento à desinformação da Justiça Eleitoral: Plano Estratégico Eleições 2022	203
3.6.4. Moderação de conteúdo realizada pelo <i>Facebook</i>	208
3.7. Análise combinada dos tópicos.....	220
CONCLUSÃO	224
REFERÊNCIAS BIBLIOGRÁFICAS	229
FONTES DOCUMENTAIS.....	232

INTRODUÇÃO

Ao longo do desenvolvimento histórico da Democracia Liberal-Representativa e, portanto, de um processo localizado no espaço e tempo, que encontra suas raízes no estabelecimento dos direitos civis na Europa no Século XVII, com o crescimento e fortalecimento da burguesia enquanto classe sociopolítica e econômica dominante, foi sendo desenvolvida a estrutura jurídico-institucional que se consubstanciou no atual Estado Democrático de Direito. O modelo ocidental de democracia deriva, portanto, do estabelecimento da chamada esfera pública e da correlação entre Estado e sociedade civil, de maneira que a ideia-fantasia de debate público ocupa um lugar privilegiado na expectativa de um quadro jurídico-institucional vibrante e que, efetivamente, faça jus ao ideal democrático.

Desde o estabelecimento do Estado Burguês, entretanto, questiona-se a amplitude participativa da chamada sociedade civil que, na verdade, inicialmente, é restrita a um pequeno grupo de indivíduos dotados de poder econômico que, posteriormente, se consubstanciou em poder político. Assim, desde o surgimento da esfera pública pautada no individualismo burguês, que nasce enquanto contraposição ao modelo de estratificação social, é possível tecer críticas quanto ao caráter efetivamente democrático da democracia institucional.

A comunicação exerce um papel central nas democracias, tanto aquela realizada de maneira difusa nas praças públicas, quanto a comunicação que se dá por meio das mídias, inicialmente pela imprensa e o livro e, posteriormente, com o surgimento das chamadas mídias de massa, como o rádio, os jornais impressos diários, revistas e, principalmente, a televisão. Se é possível identificar, ao longo do desenvolvimento histórico das mídias, uma revolução entre a mídia impressa e as chamadas mídias de massa, a comunicação digital representa uma profunda transformação na maneira em que nos comunicamos enquanto sociedade, pautada em um novo modelo chamado comunicação de todos-para-todos.

Os temas que são pautados no mundo contemporâneo duram cada vez menos tempo nos palanques de debate vigentes, havendo um processo de mitigação de ideologias, ou seja, de afinidade ideológica e identificação partidária, que têm sido substituídas pelas pautas (mais atomizadas e circunstanciais) enquanto elemento central do debate público (Sponholz, 2021, p. 228-230). As mudanças são diárias e as plataformas digitais fazem com que a opinião pública e o imaginário coletivo sejam cada vez mais influenciados pela capacidade de profissionais do *marketing* digital de diagnosticar os anseios específicos de cada cidadão ou comunidade, mostrando para os consumidores de maneira personalizada matérias ou assuntos que acabam

por conduzir o debate público (Morozov, 2018). Porém, as bolhas ideológicas, ou os filtros-bolha, ou seja, a segmentação de públicos mediante interesses, dentro dos quais circulam informações pautadas em crenças pré-existentes do grupo (Dourado, 2020, p. 66-68), bem como pela segregação causada pela possibilidade de incluir/excluir pessoas e assuntos das plataformas digitais (Marichal, 2018), tendem a gerar análises defeituosas sobre os anseios efetivos e as necessidades reais da população de um modo geral, produzindo-se um processo autossustentável no qual a opinião pública é constantemente informada por debates fabricados com a intenção de favorecer determinados sujeitos, seja por interesses econômicos ou políticos.

O rápido desenvolvimento e popularização das tecnologias de comunicação digital, que partiu de computadores-robustos e fixos-, mídias físicas para transmissão de conteúdo entre terminais e bandas de *internet* bastante limitadas para *smartphones* com acesso a uma rede veloz de *internet* capaz de conectar todas as localidades do mundo em tempo real, em apenas alguns anos, resultou em uma reorganização da estrutura da comunicação social que, historicamente, fundamentou o quadro jurídico-institucional do Estado Democrático de Direito. Nesse sentido, a transformação da estrutura sociopolítica que fundamenta o rol de princípios e garantias constitucionais, que tem enquanto objetivo a promoção da democracia, precisa ser cientificamente observada, tendo em vista a necessidade de se reinterpretar os arranjos jurídico-institucionais à luz da nova realidade social, política, cultural, econômica e tecnológica.

Nos últimos anos, assim, em diversos países de todo o mundo, como na Itália com o Movimento 5 Estrelas, nos EUA com Donald Trump, na Inglaterra com Boris Johnson e no Brasil com Jair Bolsonaro surgiram, quase que concomitantemente, diversos líderes populistas que se apresentaram como *outsiders*, pessoas que vieram de fora da política e cuja inexperiência seria a marca da autenticidade, conectando-se com o povo. Esse movimento só foi possível, entretanto, graças ao trabalho de especialistas em *marketing* digital e análise de *big data*, que se utilizam dos algoritmos das mídias digitais, que, pela estrutura das mídias digitais, não se importam se os conteúdos são verdadeiros ou falsos, para disseminar conteúdos que geram engajamento nas redes, propiciando o alcance político destas personagens. Ou seja, por trás desses líderes, há equipes inteiras que atuam como engenheiros do caos, manipulando sentimentos e desejos para direcionar o debate político. Os líderes populistas utilizam as *fake news* e o absurdo como um elemento de narrativa, para arregimentar sentimentos de pessimismo, medo e ódio para descredibilizar a política burocrática, que vem tendo sua legitimidade questionada ao longo dos últimos anos. Essa onda, portanto, se produz tanto sobre a cólera derivada da insatisfação popular com a política tradicional, pautada em razões sociais

e econômicas plausíveis, quanto sobre a supermáquina de comunicação digital (Empoli, 2019, p. 5-17).

Assim, mediante as profundas mudanças e turbulências causadas pela comunicação digital no quadro jurídico-institucional em democracias de todo o mundo, há uma proliferação de pesquisadores e estudos científicos que têm observado por diferentes prismas e em diversas áreas os impactos e características da comunicação digital, bem como sua correlação com a democracia.

A eleição brasileira de 2018 foi marcada pela supressão dos meios tradicionais de debates políticos, exemplificada pela ausência do candidato vencedor das eleições da Presidência da República de debates televisivos no segundo turno (Azevedo; Trigueiro; Martins, 2018), cumulado com a crise institucional, o descontentamento da população brasileira com a política de um modo geral, integrando um fenômeno internacional de descontentamento com os sistemas de representação em âmbito nacional¹ e com a crise epistêmica derivada das estratégias de comunicação da pós-verdade e a disputa de narrativas concorrentes, ainda que pautadas em falsidades, no espaço público (Dourado, 2020, p. 97-100).

A tônica do momento atual, tanto no Brasil quanto em diversas outras localidades, em especial países de tradição republicana ocidental, é de ampla adoção das plataformas digitais enquanto principais meios de comunicação, pautando inclusive debates políticos. No Brasil, segundo dados de 2018, o *Facebook* contava com 127 milhões de usuários mensais ativos, o que implica que 60% da população acessa esta plataforma (Oliveira, 2018); dados da pesquisa do Instituto DataSenado apontam que 45% dos eleitores utilizaram informações veiculadas pelas plataformas digitais para votar nas eleições de 2018 e 83% consideram que as mídias digitais influenciam muito o voto (Baptista, 2019); pesquisa do Instituto Reuters após as eleições de 2018 indicam que a desconfiança da população sobre as notícias de fontes tradicionais de mídia, como o rádio e a televisão, aumentou de 37% para 48% em relação ao ano anterior; e grupos em plataformas digitais, ainda que não vinculados a questões políticas, são utilizados para veicular materiais sobre votações políticas: por exemplo, 90% dos grupos de família tiveram disseminação de material político nas eleições gerais brasileiras de 2018, sendo que 60% deste material era parcial ou totalmente falso (Dourado, 2020).

¹ Nesse sentido, Chevallier (2009, p. 190-192) descreve a ocorrência de uma crise do liame político, sendo a interconexão de três crises: econômica, social e moral; Castells (2018) analisa a crise como um momento de ruptura da relação entre os governantes e governados, decorrente do predomínio dos interesses das elites econômicas, em especial em relação à crise de 2008-2009 e a revelação de diversos casos de corrupção.

Tendo em vista o alcance global das plataformas digitais, como mecanismos de pesquisa, redes sociais e plataformas de aplicativos², o regime legal que envolve a utilização da rede é dotado de duas esferas: a esfera nacional e a esfera transnacional. De fato, a intercessão entre lei, tecnologia, normas e mercados constitui uma rede de governança das intermediárias (Gasser; Schultz, 2015), constituída por normas do país em que a intermediária é utilizada, bem como normas e princípios do país em que a sede da intermediária está situada e, cada vez mais, por normativas internacionais. Ademais, as próprias plataformas, conforme discutiremos no segundo capítulo, exercem governança privada do discurso, de maneira que em seus serviços digitais atuam no estabelecimento de normas, chamadas geralmente de termos e políticas, em uma relação contratual ou quase-contratual com os usuários, bem como na aplicação (*enforcement*) destas políticas autoestabelecidas e heterodirigidas. Desta maneira, torna-se necessário entender qual o papel exercido por cada um dos atores envolvidos no processo da comunicação digital e sua regulamentação.

A partir deste diagnóstico preliminar, constatou-se que a tecnologia, no caso as tecnologias de informação e comunicação digitais, representam um importante aspecto das relações socioculturais e econômicas, em um processo de sentido duplo, em que as tecnologias se desenvolvem de acordo com as características culturais, econômicas e políticas da sociedade que as sustentam, enquanto o próprio desenvolvimento de tecnologias moldam a realidade na medida em que abrem possibilidades que eram inviáveis até então. Uma vez que as mídias digitais, enquanto elemento central da comunicação social contemporânea, são frequentemente apontadas como corresponsáveis pela instabilidade ou crise no quadro jurídico-institucional das democracias liberais-representativas, possibilidades de regulamentação pelo Direito vêm sendo discutidas ao redor do globo. O problema de pesquisa desta dissertação, nesse sentido, pode ser sintetizado da seguinte maneira: a regulamentação da comunicação digital, por meio da responsabilização das plataformas digitais sobre o conteúdo postado por terceiros, pode ser um meio eficaz e democraticamente legítimo para reduzir a disseminação sistemática de conteúdos perniciosos, como *fake news* e discurso de ódio, na comunicação social contemporânea?

Esta dissertação tem como terreno epistemológico a Ciência Jurídica. Entretanto, uma perspectiva excessivamente dogmática, que pretenda compreender o Direito puro, sem sua correlação com as características políticas, sociais, econômicas, culturais e tecnológicas que

² As empresas que atuam nesta área são chamadas de “*online intermediaries*”, em tradução livre intermediárias online. Conforme conceitua a *Organization for Economic Co-operation and Development* (OECD), intermediárias online juntam ou facilitam a transação entre terceiros na *internet*, dando acesso, hospedando, transmitindo e elencando conteúdos, produtos e serviços originários de terceiros ou provendo serviços baseados na *internet* para terceiros (OECD, 2010).

são objeto de ciências correlatas, seria incapaz de apreender adequadamente os fenômenos dos quais derivam o quadro jurídico-institucional. Ou seja, em um tema relevante, atual e complexo, como a regulamentação da comunicação digital, a interação entre Direito e Ciências não-jurídicas se faz imprescindível, sob a pena de se construir um conhecimento descontextualizado.

Nesse sentido, no primeiro capítulo deste trabalho é elaborada uma revisão bibliográfica que busca compreender, a partir de autores de Ciências não-jurídicas, as características das mídias digitais e seus impactos na comunicação social. A partir de autores da Sociologia, Ciência Política, Comunicação Social, Filosofia, Ciência da Computação e Filosofia do Direito, buscou-se desenvolver um quadro sobre as relações entre o indivíduo e a *internet*, a estrutura econômica e tecnológica das plataformas digitais e da sociedade da vigilância e os impactos sociopolíticos da fragmentação da esfera pública pela comunicação digital. Trata-se de uma leitura de um autor da área do Direito sobre as reflexões e conclusões de outras áreas, especialmente das Ciências Humanas, sobre o fenômeno.

A partir do quadro construído ao longo do primeiro capítulo, passa-se à revisão bibliográfica de autores do Direito sobre a regulamentação das mídias digitais, especialmente da possibilidade de responsabilização das plataformas digitais sobre o conteúdo postado por terceiros. Observa-se que a discussão jurídica sobre o objeto desta pesquisa orbita em torno do princípio constitucional da liberdade de expressão, seu desenvolvimento histórico e sua reinterpretação frente à estrutura social e comunicacional contemporânea. Nesse capítulo, a partir de uma abordagem eminentemente jurídica, busca-se compreender de que maneira o fenômeno da comunicação digital e, especialmente, da disseminação de *fake news* e discurso de ódio pelas mídias digitais, vem sendo apreendido contemporaneamente, na dogmática jurídica, e a maneira como tem se consubstanciado em normatização.

O levantamento bibliográfico foi realizado por meio do uso de bases de dados de instituições científicas e plataformas de disponibilização de periódicos, tais como *Scielo*, CAPES Periódicos e *Scopus*, de forma que foi feito monitoramento das publicações recentes sobre o tema, fundamental para compreender os avanços recentes na área que está em constante transformação.

Um dos principais desafios que se apresenta ao estudo da área é a vasta quantidade de produção científica na área. Desta feita, foi elaborado um levantamento preliminar, com a seleção de alguns artigos para mapear a discussão quanto ao tema, seguido de leitura e fichamento deste material. A partir do recorte desta pesquisa, qual seja a responsabilização das plataformas digitais sobre o conteúdo postado por terceiros, foi feita a delimitação dos autores e instituições com profundo impacto na temática, tanto no campo acadêmico, quanto na

normatização do tema, notadamente no Brasil e na União Europeia, utilizando-se por critérios a data de publicação do material; a importância que o material adquiriu na recente produção legislativa sobre o tema, como no caso de relatórios mencionados no Programa Permanente de Enfrentamento à Desinformação do Tribunal Superior Eleitoral; o periódico e a instituição ao qual o(a) autor(a) é vinculado e o índice de citação. Ademais, houve a inclusão de materiais em razão da participação em eventos científicos, como a Conferência *Justice & Technology*, organizado pelo *Yale Global Justice Program*, da Universidade de Yale.

A partir desta panorama sobre a comunicação digital e suas implicações para a Democracia contemporânea, a hipótese de pesquisa que guiou esta dissertação é a de que a isenção de responsabilidade das plataformas digitais sobre o conteúdo postado por terceiros, no caso do Brasil instituída por força do artigo 19 do Marco Civil da Internet, que teve como objetivo declarado a proteção da liberdade de expressão, teve enquanto efeito não esperado ou não pretendido a ampla disseminação de conteúdos perniciosos, como discurso de ódio e *fake news* no âmbito das mídias digitais.

Para estudar esta hipótese, utilizou-se a aplicação sistemática do método de mapeamento de arranjos jurídico-institucionais em que, através de uma abordagem do plano mesoinstitucional, busca-se identificar os fatores históricos que caracterizam determinada política ou ação institucional. Parte-se do reconhecimento de que as instituições moldam comportamentos e regulam expectativas, sendo fontes de modelos morais e cognitivos, de maneira que o momento da tomada de decisão normativa representa um ponto de inflexão, uma vez que as alternativas que deixaram de ser tomadas perdem força frente à opção institucional tomada. Nesse sentido, as conjunturas críticas são caracterizadas por definir os contornos da evolução institucional subsequente de determinada política.

Um dos principais méritos do método de mapeamento de arranjo jurídico-institucional é o aclaramento das relações entre os atores e interesses envolvidos no processo decisório, com seus discursos, argumentos, dimensões legislativas e regulamentares e as transformações, inflexões e evoluções através de uma análise jurídica do quadro normativo desta política, de maneira a compreender a racionalidade do conjunto normativo a partir de uma abordagem jurídica (Bucci; Gaspardo, 2024).

O método aborda, assim, os conjuntos dinâmicos que resultam de um sentido compartilhado pelos atores envolvidos em determinada política, conjuntamente com as normas, processos e instituições ou problemas jurídicos daquele recorte. O problema é abordado a partir da perspectiva de multidisciplinaridade e multidimensionalidade, de maneira que os aspectos extrajurídicos são necessários para a compreensão da dinâmica estabelecida pela integração

entre os fatores do arranjo estudado. O arranjo não é estático, mas dotado de uma flexibilidade formal, que se institucionaliza especialmente por meio das diversas normas, fontes normativas e polos de competência que regulamentam o tema observado. No caso desta dissertação, nota-se que o sistema de responsabilização instituído pelo Marco Civil da Internet, enquanto norma que guia o sistema de moderação de conteúdo postado por terceiros nas mídias digitais, é objeto de modificações derivadas de normas infralegais, como as Resoluções do Tribunal Superior Eleitoral, como também pela atuação de entidades internacionais, com a aprovação de normativas do bloco europeu, por exemplo, a pressão por parte da sociedade civil ou, ainda, pelo próprio interesse das plataformas, que exerce governança privada dos discursos.

Assim, busca-se aferir a um só tempo tanto a dinâmica estabelecida entre atores e normas, quanto o quadro institucional (normativo) que sistematiza determinada política, conferindo um nexo de unidade e sentido que une os elementos que compõem o arranjo, que tem a tendência de se estabilizar no tempo. Esse quadro é dotado, ainda, de uma ideia diretriz, subjacente, que orienta todo o conjunto em suas transformações ao longo do tempo. Uma vez instituídas as normas e delimitadas as competências, o quadro como um todo, por ser originado do direito, promove a despersonalização do arranjo, no sentido de que este aplica-se para todos, independentemente dos instituidores da política (Bucci; Gaspardo, 2024).

O método de mapeamento de arranjos jurídico-institucionais neste trabalho se deu por meio da aplicação sistemática das seguintes etapas, que estão delimitadas nos tópicos do terceiro capítulo: 1) periodização; 2) seleção e descrição dos arranjos jurídico-institucionais; 3) documentação jurídica; 4) contexto e atores políticos; 5) análise das motivações, efeitos esperados e obtidos; 6) análise combinada dos tópicos.

Através de um estudo do desenvolvimento histórico do debate e da normatização da comunicação digital, partindo-se do processo de elaboração e aprovação do Marco Civil da Internet, passando pelas transformações na perspectiva sobre o funcionamento das plataformas digitais devido à divulgação de escândalos como o Caso Snowden, a eleição de Donald Trump e a votação do *Brexit*, com a atuação da empresa *Cambridge Analytica*, a atuação do Tribunal Superior Eleitoral nas eleições de 2018, 2020 e 2022 e a atuação do *Facebook* nas redes sociais *Facebook* e *Instagram*, busca-se desenvolver um quadro sobre a moderação de conteúdos na *internet* e a governança privada dos discursos no Brasil, em sua interação com o princípio constitucional da liberdade de expressão. Assim, sob uma perspectiva jurídica, torna-se viável analisar as estruturas, processos e valores de determinado arranjo institucional em sua relação com o contexto histórico em que se insere. Tem-se como resultado uma esquematização explicativa da evolução temporal e das forças políticas que incidem sobre aquele arranjo.

CAPÍTULO 1- COMUNICAÇÃO DIGITAL, CAPITALISMO DE VIGILÂNCIA E EPISTEMOLOGIA TRIBAL

Neste capítulo, delimitaremos o problema de pesquisa, identificando os elementos que constituem os conceitos de *fake news* e discurso de ódio, enquanto fenômenos da tecnologia de mídia digital e categorias jurídicas, e os impactos dos meios de comunicação digital nas democracias contemporâneas, em especial no debate político e na formação da opinião pública. No primeiro momento, nos debruçaremos sobre a forma como os indivíduos se comunicam por meio das plataformas digitais, os incentivos criados pela chamada arquitetura da divulgação e o consequente estabelecimento de uma visão privada sobre a vida pública. No segundo momento, faremos o esforço para identificar os elementos constitutivos dos conceitos de *fake news* e discurso de ódio, observando os aspectos sociais e tecnológicos da disseminação em massa destes conteúdos por meio da comunicação digital, entendida enquanto superação da estrutura comunicacional estabelecida pelas mídias tradicionais. Desta feita, há dois recortes explicativos complementares nesta seção: o recorte individual cívico, que explica os impactos da comunicação digital sobre a formação das identidades políticas mediatizadas; e o recorte tecnológico comunicacional digital, que afirma o papel das plataformas digitais enquanto mídias e evidencia o impacto das dimensões social e tecnológica no estabelecimento (e fragmentação) do debate público contemporâneo.

1.1. Da sociedade em rede à infodemia: relações derivadas do capitalismo e informacionalismo

A Democracia Liberal-Representativa, consubstanciada no Estado Democrático de Direito, é a face institucional que confere fundamento, por via constitucional, a todo o sistema jurídico-político que pauta as relações sociopolíticas no Estado contemporâneo. Seu desenvolvimento histórico tem raiz na consolidação de uma forma de organização econômica, o capitalismo, que tem implicações em todas as relações, de produção e interpessoais, estabelecidas entre os cidadãos, organizações e o Estado. A institucionalidade democrática, assim, a despeito de por vezes se fazer crer obra divina, natural, sobre-humana ou de ordem superior, é eivada de limites e características que derivam das condições socioculturais em que

se desenvolve, demarcado no espaço-tempo, que dão forma efetiva ao jogo de relações e poder na sociedade civilizada.

No final do Século XX, as condições sociais, tecnológicas, econômicas e culturais foram modificadas de maneira tão profunda, em relação àquelas que precediam, e especialmente daquela existente no processo histórico de consolidação do Estado Democrático de Direito, que pode-se dizer que se deu origem a uma nova forma de sociedade, carregada pelas modificações na estrutura comunicacional, caracterizada por um sistema de redes horizontais de comunicação organizadas na *internet* e por meio de comunicação sem fio, criando um hiato geracional entre aqueles que nasceram antes da Era da *Internet* (1969) e os que cresceram em um mundo conectado. Essas mudanças impactaram de tal maneira a forma como nos comunicamos enquanto sociedade, que resulta na existência de uma espécie de sociedade em crise, uma vez abaladas suas estruturas, ou seja, uma sociedade em que a mudança é uma constante (Castells, 2011, p. I-III).

Os efeitos das novas tecnologias de informação e comunicação impactam os sujeitos de maneiras distintas, a depender do contexto social específico. Nesse sentido, todas as grandes mudanças sociais (vivenciadas socialmente e na maneira como nos relacionamos) são alterações promovidas na maneira como entendemos e vivenciamos o espaço e o tempo. A organização histórica dos seres humanos em grandes cidades, nesse sentido, é pautada em um sistema de comunicação lastreado na contiguidade física, ou seja, por meio da aproximação no espaço desenvolve-se a sociedade em termos de coordenação, troca de bens e serviços e afins, o que pode ser chamado de “espaço da contiguidade” (Castells, 2011, p. VI-IX; XVI-XXI).

A revolução das tecnologias de comunicação digital possibilitou, entretanto, de maneira inédita, fluxos de comunicação e informação que trouxeram o aspecto de simultaneidade, antes restrito ao espaço físico contíguo, para além deste, de forma que o espaço passa a abranger também a própria existência da rede, ao que, por sua vez, dá-se o nome de espaço dos fluxos, constituído pela produção, transmissão e processamento de fluxos de informação. As redes que se estruturam globalmente ressignificam e reclassificam espaços físicos, constituindo as regiões metropolitanas, ou seja, espaços físicos não necessariamente contíguos que exercem função para as redes globais, constituindo, assim, nós que conferem materialidade ao sistema. Desta forma, as redes globais possibilitadas pelas tecnologias de comunicação e informação conferem funções e características para as regiões metropolitanas, que representam os nós das redes globais, os quais demandam estrutura para cumprirem sua função, como a existência de transporte multimodal, redes de comunicação, redes de computadores, sistemas avançados de informação e serviços de infraestrutura (Castells, 2011, p. VI-IX; XVI-XXI).

O motivo pelo qual a estrutura depende desses nós é a necessidade de tomada de decisão de alto nível que, pela especialidade e necessidade de ser feita com discrição, depende da proximidade física. Assim, os “ambientes” que se constituem nas diversas redes, como ambientes financeiros, que ocorrem em Nova York, Londres e Tóquio, ou ambientes tecnológicos, como o Vale do Silício, constituem os centros em que ocorrem os processos decisórios, uma micro-rede operada por comunicação pessoal, e que se articulam com as demais localidades por meio do espaço de fluxos, comunicando para a macro-rede global eletrônica as decisões tomadas. A principal contradição que resulta da coexistência entre o espaço dos lugares e o espaço dos fluxos é que, enquanto as pessoas vivem e ganham sua vida no espaço físico, portanto dos lugares, os processos de tomada de decisão do mundo globalizado e organizado em redes está pautado na lógica do espaço dos fluxos (Castells, 2011, p. XXI-XXV).

A maneira como o tempo é experimentado decorre da organização social, tecnologia, cultura e condição biológica da população. Na Idade Média, o tempo era medido de maneira pouco precisa, sendo marcado pelos ciclos solar e agrícola, e as feiras coincidiam em um mesmo espaço a produção agrícola, o comércio, a sociabilidade e a festividade. Durante a revolução industrial, a invenção do relógio possibilitou que o tempo do trabalho passasse a imperar, determinando o cotidiano da população. Com o capitalismo, entretanto, o tempo passa a ser entendido enquanto fonte de lucro, uma vez que quanto mais rapidamente o sujeito for capaz de reinvestir o capital, maior a rentabilidade. Ao tempo foi conferido valor de mercado, constituindo uma espécie de “tempo atemporal”, em que há um descompasso na ordem sequencial das práticas sociais, característico da sociedade em rede. Se o mercado financeiro foi o primeiro ponto em que se percebeu a ocorrência do tempo atemporal, cada vez mais as tecnologias de comunicação e transporte tem impulsionado os indivíduos a uma forma de vida pautada na atemporalidade, uma vez que se gera a percepção de que é possível realizar várias atividades de maneira simultânea, tendo em vista as tecnologias de comunicação e informação, confundindo-se a jornada de trabalho, o tempo pessoal e o tempo familiar e consubstanciando-se o tempo enquanto ferramenta do poder, impelindo a uma ideia de como é viver bem (Castells, 2011, p. XXV-XVII).

Mediante o esgotamento do modelo keynesiano de crescimento capitalista, que se iniciou na década de 1970, atingindo os limites do crescimento da capacidade de consumo e gastos públicos, pautado no Estado de Bem-Estar Social, resultando na espiral inflacionária irrefreável, o capitalismo passou por uma reestruturação ao longo dos anos 1980 e 1990, passando a haver um esforço pela desregulamentação, privatização, desmantelamento do contrato social entre capital e trabalho, tanto por parte do Estado quanto das empresas privadas,

operacionalizado por políticas econômicas similares entre os países da Organização para Cooperação e Desenvolvimento Econômico (OCDE) e imposta pelo Bundesbank, o Federal Reserve Board e o FMI, que passaram a apostar em um modelo de negócio com quatro objetivos principais: busca do lucro na relação entre capital e trabalho; aumento da produtividade do trabalho e do capital; globalização da produção, circulação e mercados tendo em vista o lucro; e apoio estatal para aumentar a produtividade e competitividade de economias nacionais, em detrimento da proteção social (Castells, 2011, p. 53-56).

O cenário da sociedade de redes, assim, é de incertezas e desmantelamento de estruturas, pautado pela desconexão entre os interesses da Rede e do Ser. Esses momentos de incerteza privilegiam a busca por uma identidade e sentido, de forma que ao longo dos anos 1990, concomitante com a expansão do informacionalismo, a construção da ação social e das políticas passou a tender para a formação de identidades primárias. O resultado é a busca de sentido em uma identidade partilhada, em uma contraposição ao sentido tradicional de individualidade. A sociedade informacional, desta maneira, tem em suas raízes a busca de identidade enquanto princípio organizacional. É nesse sentido que, paradoxalmente à existência de mais conexão, ocorre também um recrudescimento das relações sociais, com a ascensão de movimentos nacionalistas, fundamentalistas religiosos e racistas (Castells, 2011, p. 57-60).

A digitalização maximiza os efeitos dos processos de globalização e hiperculturalização que dissolvem o mundo da vida comum, ou seja, produzem uma desfaturização e descontextualização crescente que implicam em que os sujeitos deixam de se entender enquanto viventes de uma mesma comunidade, sob um mesmo pano de fundo holístico, onde residem questões em comum. O ser humano, entretanto, tem a necessidade de pertencer e encontra na formação de uma rede por meio da comunicação digital a satisfação deste anseio, tendendo à tribalização. A informação veiculada pelas plataformas digitais, na ausência de elementos de pertencimento e identificação, assume a natureza de identidade, ao invés de ser fonte de conhecimento. Identidade e opinião, desta forma, dentro destas redes, coincidem, deixando de lado a racionalidade e os fatos, de forma que não há qualquer espaço para a reflexão ou à renúncia de preconceitos, uma vez que a assimilação do conteúdo se dá pela crença. Favorece-se uma guerra de identidades, em que há a compreensão de que aqueles que não fazem parte do grupo, identificados enquanto “os outros”, devem ser combatidos, gerando uma sociedade de identidades inconciliáveis (Han, 2022, p. 53-62).

Essas características da sociedade em rede abalaram as estruturas do Estado Moderno, instituindo novas relações entre o mercado financeiro, o Estado e a sociedade, tornando fluidas as relações sociais e incertas as trajetórias pessoais. A busca por uma identidade em tempos de

insegurança é o resultado inevitável de um sistema em que as decisões estratégicas são tomadas de acordo com o papel que determinado indivíduo, grupo, região ou país exercem para a rede globalmente instituída. Assim, ocorre o reagrupamento em torno de identidades religiosas, étnicas, territoriais e nacionais, de forma que, socialmente, sobressai uma oposição entre a rede, que reduz tudo à função, e o ser, que deseja ter significado. Desse processo decorre uma fragmentação social, em que a comunicação, mesmo que conflituosa, por meio de lutas sociais ou oposição política, é suprimida, e, grupos e indivíduos passam a entender uns aos outros enquanto estranhos e, até mesmo, uma ameaça (Castells, 2011, p. 39-41).

Na racionalidade discursiva característica do estabelecimento de uma esfera pública, o discurso, dotado do pressuposto do diálogo com os outros perante um mesmo pano de fundo que abarca a todos, provê a possibilidade de readequação e reflexão. Na racionalidade digital, entretanto, o discurso é substituído pelos algoritmos e o aprendizado de máquina, ou *machine learning*, em que a correção de rota se dá sem interação com o outro. O fundamento da racionalidade discursiva é o reconhecimento da limitada capacidade humana de compreensão, restrita em razão de aspectos como espaço e tempo, enquanto na racionalidade digital há uma falsa crença de que todo o conhecimento e informação pode ser processado por meio de máquinas que superam tais limitações, sugerindo uma espécie de infocracia enquanto pós-democracia digital, em que os resultados obtidos pelo processamento computadorizado de dados seria capaz de maximizar todos os bens necessários à satisfação das necessidades humanas, tornando obsoletos os princípios e o sistema democrático (Han, 2022, p. 63-71).

Na lógica da racionalidade digital do dataísmo, a vontade geral da sociedade pode ser aferida pela aritmética, somando-se a vontade de cada um dos indivíduos e obtendo-se a média. Deixa-se de levar em consideração, assim, o papel da comunicação enquanto elemento que transforma interesses e opiniões e que constitui um dos pontos fundamentais sobre o qual se sustenta a vida em comunidade, de forma que entende, ao revés, que a comunicação em verdade deturparia a busca pela vontade geral, ao influir na maneira como se expressam os interesses individuais. Essa racionalidade estaria pautada na ideia de ser possível prever as ações dos indivíduos em função do levantamento de dados, ignorando a autonomia e liberdade dos indivíduos (Han, 2022, p. 71-79).

Na verdade, entretanto, a comunicação molda a realidade na medida em que não vemos o mundo como ele é, mas sim como nossas linguagens nos permitem apreendê-lo. Ou seja, nossas linguagens, e através dos meios de comunicação, constituem as metáforas que dão sentido ao mundo que, por sua vez, constituem o que é chamado de cultura. No campo da comunicação, desta feita, há duas grandes revoluções que demarcaram a história humana: 1) o

surgimento do alfabeto, que possibilitou a comunicação cumulativa baseada em conhecimento e 2) a cultura audiovisual, que ao longo do século XX, primeiramente por meio do filme e do rádio, e posteriormente por meio da televisão, passa a constituir o centro da comunicação de massa. A terceira grande revolução, vivida hodiernamente, é a introdução da comunicação digital, pautada na metalinguagem, e que une de maneira inédita as formas de expressão escrita, oral e audiovisual em um mesmo sistema, modificando a maneira como nos comunicamos (Castells, 2011, p. 413-415).

A introdução da televisão transformou os meios de comunicação na medida em que rádio, jornais, livros e filmes passaram a ser pautados pela lógica dominante desta mídia. O motivo pelo qual a televisão atrai esta atenção foi objeto de debate nos estudiosos da comunicação em massa, não havendo consenso, entretanto uma hipótese interessante recorrentemente aventada é a de que o ser humano busca, seja por natureza, seja pela organização social pautada em longas jornadas de trabalho e faltas de alternativa de envolvimento pessoal/cultural, o caminho menos custoso para a obtenção de informação e entretenimento, de forma que a televisão ocupou esse espaço de maneira primorosa. A lógica da comunicação por meio da televisão é da produção de conteúdo que é transmitido para milhões de telespectadores, de forma que o conteúdo e o formato das mensagens são pautados na busca de um denominador comum mais simples, capaz de alcançar toda uma massa homogênea ou homogeneizável, que é o coração da sociedade de massa (Castells, 2011, p. 415-417).

A esfera pública, enquanto uma construção comunicacional que teve início no comércio renascentista e na revolução iluminista do século XVII, surge enquanto troca de informações e mercadorias no seio de um grupo social específico com interesses privados, no início do capitalismo, e apenas posteriormente adquire intenções políticas ou, mesmo, sociais, tendo, portanto, primeiro, caráter eminentemente econômico. Nesse sentido, as plataformas sociais, ao serem pautadas primariamente em interesses comerciais, para utilizar dados e olhar como valor econômico, apenas continuam a implementar prioritariamente os interesses do capital e, em segundo plano, o interesse cultural e político. A estrutura comunicacional, assim, adquire forma a partir dos interesses de mercado, desde o estabelecimento da instância da palavra impressa, donde, portanto, se localiza sua origem material (Bucci, 2021, p. 41-45).

Se ao menos a mídia impressa é caracterizada por produzir conteúdo, primeiro comercial, e depois político e cultural, a mídia de massas se caracteriza pelo entretenimento, substituindo a produção de conteúdo pela fabricação de audiência, centralizada no alcance em larga escala, prescindindo de argumento racional. A comunicação digital intensifica a produção

industrial de um material acrítico, calcado em um individualismo narcisístico, estimulando a agir irrefletidamente em favor dos interesses do mercado e do capital (Bucci, 2021, p. 45-49).

A esfera pública, que na instância da palavra impressa era caracterizada pelo uso da palavra enquanto instrumento da racionalidade, passa a veicular novas formas de comunicação voltadas à estética e às emoções, fagocitando os ramos comunicacionais em oligopólios cuja finalidade é gerar o entretenimento, ocupando o tempo ocioso do consumidor, para extrair seu olhar e comercializá-lo. As funções que eram exercidas pela religião, ideologia, divulgadores científicos, educadores, jornalistas e o discurso político passam a ser imperadas pela instância da imagem ao vivo, pela lógica do entretenimento, sob pena de não alcançarem as massas, tornando-se, portanto, o centro exclusivo do qual irradiam as possibilidades de identificação e juízo de valor dos indivíduos (Bucci, 2021, p. 50-56).

A indústria do entretenimento serve, desta maneira, à criação dos mitos modernos, definindo o que é virtuoso ou vicioso, e preferindo o mito em relação aos fatos. Nesse sistema comunicacional, a desinformação atua em duas frentes, a primeira em um simulacro jornalístico, fazendo-se passar por informação confiável, e a segunda, através das técnicas do entretenimento, pela criação e exploração dos mitos, utilizando-se de sentimentos primários, como o ódio e o preconceito, para explorar os medos e desejos, manipulando fatos e afetos (Bucci, 2021, p. 56-59).

De acordo com a sociologia sistêmica de Luhmann, há uma correlação entre os tipos de sociedade e as tecnologias comunicacionais, ou meios de disseminação, que os sustentam. Assim, o surgimento da imprensa, ao possibilitar a crítica pública dos atos do poder estatal, desestabiliza a sociedade estratificada, para estabelecer a chamada opinião pública, enquanto um espelho anônimo que julga as ações estatais. A sociedade civil passa a ser considerada a arena de cidadãos reflexivos que se informam e disputam sentido pela imprensa. Essa opinião pública, por sua vez, reclama meios democráticos para influenciar as decisões públicas, resultando na criação de uma esfera pública pluralista de grupos e em uma disputa organizada, por meio de organizações partidárias, sindicais e outras, pautada sobretudo sobre os meios de comunicação de massa tradicional. As redes sociais, nesse sentido, por meio da fragmentação e indeterminação na opinião pública, desestabilizam a estrutura sociopolítica estabelecida ao longo dos séculos XIX e XX, ao modificar a tecnologia comunicacional que sustenta o sistema (Amato; Saba; Loschiavo Leme de Barros, 2022, p. 545-546).

Tecnologia e sociedade são aspectos que se correlacionam de maneiras bastante diversas, a depender da realidade local e da organização social de uma maneira geral. Assim, o desenvolvimento de tecnologias tem relação com características da sociedade em que se

desenvolvem e na maneira como a sociedade utiliza-as, da mesma forma em que a própria tecnologia implica em modificações sociais. A rede mundial de computadores, principal revolução tecnológica de informação e comunicação, teve seu primeiro desenvolvimento no seio da Agência de Projetos de Pesquisa Avançada do Departamento de Defesa dos Estados Unidos na década de 1960, porém o avanço da tecnologia e a ampla disseminação de dispositivos tecnológicos, desde os anos 1960 e 1970, se deu de acordo com a cultura dos *campi* norte-americanos, pautada na liberdade, inovação individual e iniciativa empreendedora. Nessa época, o Vale do Silício, por meio de um segmento específico da sociedade estadunidense, conectado à economia global e à geopolítica mundial, passa a ser o centro de um estilo de produção, comunicação, gerenciamento e vida que pautaram o desenvolvimento das tecnologias de informação (Castells, 2011, p. 43-45).

O desenvolvimento de tecnologias, portanto, é localizado no tempo e no espaço em que ocorrem, especialmente dando-se destaque às correlações da inovação tecnológica com a organização social e as ações do Estado. Quanto às tecnologias de informação e comunicação que revolucionaram a sociedade pós-Moderna, o momento histórico em que ocorreu é concomitante com a reestruturação global do capitalismo, sendo um instrumento fundamental deste e moldando as características desta nova sociedade que é, ao mesmo tempo, capitalista e informacional (Castells, 2011, p. 45-50).

Nesse sentido, há três processos que organizam as sociedades por meio de relações historicamente determinadas: a **produção**, que é a transformação da natureza em objetos de consumo de acordo com os objetivos socialmente determinados, pautando a divisão dos bens em relação ao consumo e ao excedente; a **experiência**, que é a relação entre as identidades biológicas e naturais dos seres humanos em relação aos ambientes, sociais e naturais; e o **poder**, que é a relação de sujeição estabelecida por meio da violência potencial ou real que dá origem às instituições enquanto meio de coerção das relações de poder, especialmente por meio do Estado. A intercomunicação entre esses três processos se cristaliza ao longo do tempo em territórios específicos, estruturando o que é denominado cultura e identidades coletivas (Castells, 2011, p. 50-52).

A produção coletiva abrange tanto o papel do trabalhador quanto dos organizadores da produção, se consubstanciando no uso dos meios de produção para agir sobre a matéria, por meio da tecnologia. O produto do processo produtivo se subdivide em consumo e excedente e as estruturas sociais, por sua vez, pautam a apropriação, distribuição e uso do excedente, o que é denominado modo de produção, definindo as relações sociais e a existência de classes sociais de acordo com a posição que ocupam neste sistema. Durante o século XX, ocorreram dois

modos de produção distintos: o capitalismo, pautado na transformação do excedente em *commodity* e a posse privada dos meios de produção, que pautam as regras de apropriação e distribuição com vistas à maximização dos lucros; e o estatismo, no qual o Estado exerce o controle sobre o excedente, com vistas à maximização do poder (Castells, 2011, p. 52-53).

O modo de desenvolvimento, por sua vez, é caracterizado pelas relações técnicas de produção, ou seja, os instrumentos utilizados para ampliar o produto da produção, e consequentemente o excedente, com a mesma unidade de insumo. Assim, no modo agrário de desenvolvimento, o excedente foi ampliado pelo aumento quantitativo da mão-de-obra e dos recursos naturais, especialmente as terras. No modo industrial de desenvolvimento a principal revolução é na aplicação de novas fontes de energia e na descentralização destas fontes ao longo do processo produtivo e de circulação. Já no denominado modo de desenvolvimento informacional, a fonte de produtividade reside na tecnologia de geração de conhecimentos, processamento de informação e comunicação de símbolos em um processo que se repete.

Cada modo de desenvolvimento tem um princípio de desempenho que pauta os processos tecnológicos: no modo informacional, a busca por acumulação de conhecimentos e maiores níveis de complexidade do processamento de informação é central. O modo de desenvolvimento impacta diretamente as relações e estruturas sociais, influenciando as estruturas de poder e experiência, modelando o comportamento social, inclusive a comunicação simbólica. Na sociedade pós-Moderna, nesse sentido, novas formas de interação, controle e transformação social derivam da interligação entre cultura e forças produtivas, baseadas nas tecnologias de informação e comunicação (Castells, 2011, p. 53-54). Portanto, esses são dois aspectos fundamentais que pautam a sociedade pós-Moderna: a correlação entre o modo de produção, o capitalismo, e o modo de desenvolvimento, o informacionalismo

Durante a pandemia de Covid-19, no ano de 2020, as maiores empresas de tecnologia do mundo seguiram um processo de valorização que havia se iniciado anteriormente, de maneira que as cinco maiores empresas do setor no mundo, *Apple*, *Google*, *Microsoft*, *Facebook* e *Amazon* atingiram, juntas, o valor de cinco trilhões de dólares, valor somente inferior ao PIB dos EUA, de cerca de vinte trilhões, e da China, cerca de quinze trilhões, demonstrando, pela magnitude das empresas, a dificuldade de se regulamentar as chamadas *big techs* a nível nacional. O motivo pelo qual esse mercado é tão valioso deriva da capacidade das empresas de praticar uma espécie de extrativismo digital, pautado em uma economia da atenção, em que através da coleta de dados dos usuários realiza o chamado *database marketing*, direcionando o olhar e a atenção dos usuários das plataformas para os interesses das empresas (Bucci, 2021, p. 8-12).

É importante destacar o papel da comunicação digital, em que os dados não são apenas o novo petróleo, ou um novo produto ou matéria-prima, mas atua, enquanto corolário do sistema capitalista, para a substituição dos produtos corpóreos por um mercado de signos e discursos que conferem sentido, enquanto centro da produção capitalista. O capital é, nesse sentido, um “narrador, um contador de histórias, e se faz um produtor de significações”. O olhar, neste processo, não é um mero “receptor de imagens prontas”, mas uma “força constitutiva de sentido social”, de forma que olhar para uma imagem exerce a função de trabalhar para que ela adquira sentido. É por meio do olhar social, portanto, que se constroem grifes e marcas, reputações de políticos ou empresas, em uma verdadeira superindústria do imaginário, caracterizada por centralizar nas relações econômicas, em detrimento de outros mecanismos cuja finalidade é produzir significados, ou seja, criar representações simbólicas, como ritos religiosos e celebrações autônomas em relação ao capital, a fabricação de representações imaginárias. Por meio deste processo, a produção pouco se relaciona com a necessidade (aspecto material), mas sim com o desejo (inconsciente), produzido pela superindústria do imaginário e, posteriormente, materializado na imagem da mercadoria. A superindústria do imaginário é, portanto, a apropriação do capital sobre a linguagem e os significantes associados a significados (signos) (Bucci, 2021, p. 12-21).

Nesse sentido, o papel da mídia no debate público revela-se importante aspecto da degeneração da democracia em uma infocracia: a democracia liberal-representativa foi possibilitada pelo surgimento de uma cultura em torno da mídia impressa, dos livros, pautada em desenvolvimentos lógicos complexos e em um “tempo” maior de debate e reflexão. Com a inserção das mídias de massa, a televisão tornou a atividade do público menos reflexiva e interativa, de forma a privilegiar a posição de meros consumidores em detrimento da condição de sujeitos, fragilizando a possibilidade de emancipação da esfera pública, estabelecendo uma espécie de midiocracia em que o debate público é capturado pelo entretenimento oferecido pela mídia de massa (Han, 2022, p. 25-33).

Desde o Século XVIII e até o início do Século XX, a mídia impressa exerceu papel central na apreensão da realidade, no que pode ser denominado “instância da palavra impressa”, período no qual os jornais impressos pautavam o debate público, atuando como uma espécie de pano de fundo que conferia sentido e confiabilidade ao olhar social. O desenvolvimento de tecnologias de comunicação audiovisuais, entretanto, transmitida por meio de telas, substituiu a mídia impressa, passando a constituir o centro da construção do imaginário, da busca social por referenciais e sentidos comuns, instituindo-se uma nova era, a da “instância da imagem ao vivo”. Por instância deve-se entender uma espécie de pano de fundo que abarca, atrai, confere

sentido e confiabilidade a todo o olhar social, que inclui a todos e que, entretanto, não está à vista, nas palavras de Bucci (2021, p. 34) “instância é a casca imantada e inapreensível que recobre o olhar social” (Bucci, 2021, p. 28-35).

A instância da palavra impressa e a instância da imagem ao vivo se diferenciam pela espacialidade, uma vez que a palavra impressa encontra limitações de alcance em razão do volume de tiragens, dos canais de distribuição e da linguagem, tendo em vista que somente o leitor bem alfabetizado era capaz de acessá-la, e pela temporalidade, já que foi possível superar o ciclo pautado nas edições diárias por meio de imagens transmitidas ao vivo. A transmissão de imagens popularizou o acesso às mídias sociais, ao mesmo tempo em que escanteou o jornalismo para tornar o entretenimento no carro-chefe da comunicação midiática, tornando o espaço público em um “parque virtual de diversões”. Por meio da instância da imagem ao vivo, o pensamento é suplantado pelo olhar, e o estímulo visual supera a atividade reflexiva de tal forma que mesmo a palavra escrita é capturada pela lógica da imagem. Nesse sentido, a comunicação digital é a continuidade e exacerbação da instância da imagem ao vivo, ao intensificar suas características de superação do raciocínio abstrato, da precipitação de bolhas e de identidades primárias (Bucci, 2021, p. 35-41).

Portanto, deste sistema dominado pela comunicação de massa, especialmente a televisão, que estrutura a forma como os demais meios de comunicação se organizam, por exemplo o rádio e a mídia impressa, passamos a um sistema cujo elemento central é a comunicação digital. Se durante a predominância da mídia escrita a atividade reflexiva era incentivada, forjando o que se chama esfera pública, na era da mídia de massa esta foi substituída pelo entretenimento. É nesse momento, portanto, que o incentivo à introspecção deixa de ser presente nos meios de comunicação. Na mídia digital, essa falta de reflexão derivada da comunicação de massa ainda é sentida, porém agora os sujeitos deixam de ser telespectadores passivos para tornarem-se emissores ativos que produzem e consomem (e são consumidos) conteúdo permanentemente, o que caracteriza uma verdadeira infodemia (Han, 2022, p. 33-35).

A infodemia é o excesso de informação e comunicação que, por sua vez, resulta no desvio da atenção da esfera pública de temas socialmente relevantes. A produção e o compartilhamento incessantes de conteúdo produz alterações na forma como compreendemos o mundo, uma vez que a noção de atualidade passa a abarcar um período muito restrito de tempo, dificultando ou mesmo impedindo a construção de narrativas coesas, fruto do raciocínio lógico, necessárias ao processo democrático. Em detrimento da racionalidade, prevalece nesta comunicação acelerada a busca por benefícios de curto prazo e o apelo à emoção, ou ao afeto,

que são instrumentos mais adequados para lidar com a temporalidade da sociedade da informação. Ou seja, em lugar de raciocinar sobre as informações socialmente veiculadas, há um estímulo a privilegiar a emoção como instrumento de assimilação (Han, 2022, p. 35-37).

1.2. Auto-comunicação de massa, ascensão das redes sociais e a formação de autoidentidades mediadas

O rápido avanço do acesso à *internet*, ou seja, sua popularização, deu-se ao longo dos anos 1990 e 2000, quando o número de usuários e de serviços oferecidos aos particulares foi objeto de larga expansão. A tecnologia da *internet*, na verdade, como referimos anteriormente, remonta o ano de 1969, utilizada inicialmente para fins militares. Entretanto, na década de 1990, desenvolveram-se tecnologias para o uso privado da *internet*, dando-se destaque ao desenvolvimento do navegador *World Wide Web*, projetado por Tim Berners-Lee em 1990, que possibilitou a conexão entre os computadores mundo afora. A *World Wide Web*, ou *www*, ditou a forma como a comunicação pela *internet* se deu nos anos 1990, ao concentrar em um mesmo espaço, por meio da criação de *homepages*, a produção de conteúdo por indivíduos, instituições, empresas, associações e pessoas físicas, estabelecendo o sistema de comunicação mediada por computadores como conhecemos hoje (Castells, 2011, p. IX-XI; p. 441).

O número de usuários da *internet* passou de 40 milhões em 1995 para 1,5 bilhão em 2009, ao mesmo tempo em que a tecnologia de comunicação sem fio por meio de telefones celulares passava por uma revolução semelhante: em 1991, havia cerca de 16 milhões de contratos de serviço telefônico no mundo, enquanto em 2008 esse número chegou a 3,4 bilhões. Ao longo dos anos 2000, a união entre a tecnologia de comunicação sem fio e de acesso à *internet* estruturou a nova forma de comunicação predominante na sociedade contemporânea (Castells, 2011, p. IX-XI).

A *internet* se popularizou muito rapidamente, com sua expansão no número de usuários sendo alavancada permanentemente em razão da acumulação de mais conexão. Se as projeções de usuários na rede de computadores eram de uma expansão rápida, prevendo-se algo em torno de dois bilhões de usuários para antes de 2010 (Castells, 2011, p. 433), a evolução constatada de fato confirma o quanto projetado: segundo dados da Data Reportal (2023), em janeiro de 2023 o número de usuários de *internet* no mundo era 5,16 bilhões, o equivalente a 68% da população mundial. A *internet* se expandiu de maneira inédita, em comparação a outras formas de mídia, atingindo sessenta milhões de usuários nos EUA em apenas três anos, por exemplo,

enquanto o rádio levou 30 anos e a TV 15 anos para alcançar o mesmo patamar (Castells, 2011, p. 439).

A comunicação mediada por computadores foi a revolução que efetivamente modificou as características da comunicação de massa. O desenvolvimento da rede de computadores se deu, primeiramente, pela iniciativa estatal, sendo especialmente bem-sucedidas as experiências estadunidense, por meio da ARPANET, que se tornou a *Internet*, e a francesa, por meio da MINITEL. Enquanto o sistema estadunidense evoluiu para o uso de computadores como pontos de conexão da rede e pautou-se no investimento em empresas de tecnologia, bem como no objetivo de expandir-se mundialmente, a experiência francesa apostou na construção de servidores conectados e ficou restrita ao território francês, porém apostou em tarifas homogêneas e participação nos lucros transparentes, enquanto a outra experiência pautou-se em um sistema tarifário anárquico e serviços incontroláveis (Castells, 2011, p. 428-429). Não é de se surpreender, portanto, que a ARPANET tenha se sobreposto enquanto modelo nativo do capitalismo.

A comunicação por meio da *internet* constitui uma nova forma de mídia, distinta da mídia tradicional constituída pela televisão, rádio e jornal, uma vez que abrange diversos aspectos da vida cotidiana, como trabalho, conexões pessoais, informações, entretenimento, serviços públicos, política e religião. De fato, a introdução desta tecnologia de comunicação modifica até mesmo a atuação das formas de mídia tradicional, uma vez que os usuários podem acessar os mesmos conteúdos veiculados pela televisão, por exemplo, no horário e local que melhor lhe aprouver por meio da *internet*, de forma que mesmo a mídia em sentido tradicional é, agora, comunicação baseada na *internet*, tendo em vista a necessidade de adaptação mediante esta nova forma de comunicação (Castells, 2011, p. XI-XII).

A possibilidade de comunicação entre os próprios usuários surgiu com o desenvolvimento da Web 2.0 e Web 3.0, que se constitui das tecnologias, dispositivos e aplicações que possibilitaram o desenvolvimento de redes horizontais de comunicação. Dentro do sistema de comunicação digital, assim, há a comunicação entre os próprios usuários, por meio de compartilhamento de arquivos e redes *peer-to-peer* (p2p), constituindo-se verdadeiras redes de comunicação de massa entre usuários por meio de SMS, *blogs*, *podcasts*, *wikis* e afins. Essas plataformas digitais são, portanto, meios de comunicação de massa, entretanto possuem características distintas dos meios de comunicação de massa tradicionais. Essas redes são multimodais, ou seja, podem ser constituídas de diversas formas e veicular conteúdos distintos, como fotos ou projetos cooperativos de grande escala como o *Wikipedia*, músicas, filmes, redes de ativismo social/político/religioso e afins (Castells, 2011, p. XII-XIII).

É importante ressaltar um aspecto inserido pela comunicação digital, qual seja a concentração de atividades distintas, como trabalho, lazer, família, negócios e personalidade em uma só mídia, digital, o que relativizou a percepção de separação entre os diversos aspectos da vida entre os usuários e a existência de diversas formas de comunicação dentro do mundo digital. Dentro desse mesmo sistema, estão incluídas as inovações tecnológicas do audiovisual que haviam sido inseridas por meio da televisão, a veiculação de conteúdos escritos, o oferecimento de filmes, músicas, jogos eletrônicos, enfim, uma variedade inviável nos meios de comunicação anteriormente existentes. O desenvolvimento tecnológico necessário para oferecer esses canais, entretanto, foi altamente custoso, de forma que a elaboração do sistema, assim como outros aspectos da organização social, se concentrou nos interesses empresariais (Castells, 2011, p. 448-453), ou seja, passaram a ser guiados pelas escolhas feitas pelo que vieram a ser as *big techs*, grandes empresas do setor de tecnologias de comunicação e informação digitais, em consonância com o capitalismo informacional.

O desenvolvimento de tecnologias de comunicação digital privilegiou dois processos que alteraram a configuração das relações sociais: a centralidade da casa, uma vez que os indivíduos conseguem se relacionar com diversos aspectos da vida por meio da comunicação à distância e o individualismo, uma vez que cada pessoa passou a ter melhores instrumentos para decidir sobre o seu próprio tempo e espaço, mediante as oportunidades oferecidas pela conexão digital (Castells, 2011, p. 456).

As características do padrão cultural/ social estabelecido a partir da comunicação multimídia são: 1) a diferenciação cultural e social e segmentação entre os usuários, de forma que não apenas a produção de conteúdo pelo emissor, como já havia ocorrido na televisão, mas o próprio conteúdo que é acessado difere de acordo com o interesse do usuário; 2) o aumento da estratificação social no acesso à rede, de forma que indivíduos com maior capacidade econômica podem tirar melhor proveito das conexões digitais, enquanto a massa recebe conteúdos pré-programados, intensificando recortes de classe, raça, sexo e país no acesso aos bens, em contraposição ao poder unificador da televisão que apresentava a mesma mensagem/conteúdo para todos; 3) através do acesso a diversos tipos de mensagens pela mesma mídia, de forma que há a integração de todas as mensagens em um padrão cognitivo comum, havendo uma convergência entre “as várias fontes de envolvimento cognitivo e sensorial”. Ou seja, por meio da comunicação digital, interação social, videogames, novelas, noticiários, enfim, qualquer conteúdo que seja acessado, passa a se aproximar, tanto do ponto de vista do receptor, quanto do emissor, através de um mesmo processo de comunicação simbólica; 4) a multimídia une, em um mesmo domínio, praticamente todas as expressões culturais: desde a

mídia impressa à audiovisual, da cultura popular à erudita, entretenimento e informação, educação e persuasão, criando um verdadeiro universo digital que se constitui enquanto um ambiente simbólico que, apesar de virtual, constitui cada vez mais parte da realidade cotidiana (Castells, 2011, p. 456-458).

A globalização iniciou um processo de trocas incessantes, vivido hodiernamente, em que pessoas, moedas, bens, imagens e palavras são constantemente, e por todo o mundo, objeto de troca, expansão e comunicação, de forma que a circulação é um elemento estruturante da modernidade. O espaço social, que conferia elementos de identificação, foi convertido a mero cenário em que se dá uma espécie de cultura mundial: a cultura do consumo e do entretenimento, em que os espaços se tornam rotas de passagem. O resultado é uma sociedade do consumo, em que todos os aspectos da vida são abarcados pela lógica do entretenimento, e até mesmo os elementos de conexão e de cidadania são entendidos a partir de bens de consumo e dos meios de comunicação de massa, em detrimento das regras abstratas da cidadania ou da participação em espaços públicos (Bucci, 2021, p. 116-122).

A supermodernidade inaugura o não lugar, caracterizado por excessos, ou superabundâncias: a superabundância factual, que produz acontecimentos tão históricos quanto efêmeros, reproduzindo incessantemente ocorrências que serão reproduzidas para consumo e, rapidamente, substituídas por outras; a superabundância espacial, que congrega tanto as localidades físicas por meio tanto das tecnologias de transporte, quanto do virtual, pela rápida circulação da imagem; e a individualização de referências, em que todos os elementos da vida são comercializados na forma de consumo, homogeneizando elementos tão distintos quanto a informação, a publicidade e a ficção em um mesmo padrão comunicacional. O sentido é personalizado e vendido às pessoas individualmente, sem interação com o outro. A produção industrial da individualização serve como matéria-prima para a estruturação da coletividade por meio da chave do consumo, no que pode ser chamado de capitalismo superindustrial, no qual o lugar, que exercia a função identitária, relacional e histórica, perde espaço para a transitoriedade permanente do não lugar. A imagem surge enquanto uma necessidade para a superação da barreira linguística e, através da lógica do consumo, constitui a substância real (circulante) da mercadoria no capitalismo superindustrial, sobrepondo a própria materialidade. Se a prevalência da transitoriedade e da imagem se estabelece pela televisão, as plataformas digitais ampliam o escopo desse sistema, em que não só o real é transformado em virtual, como a virtualidade determina e transforma o real (Bucci, 2021, p. 122-129).

O tempo, conforme socialmente sentido no capitalismo superindustrial, é superacelerado em um ritmo alucinante que gera insatisfação e busca permanente pela

realização pessoal, em suas facetas produtiva, do consumo e do entretenimento, anuviando passado e futuro, condensados em um presente instantâneo que é o presente do capital. Tempo e espaço se fundem na instância da imagem ao vivo e, transcorrendo na velocidade da luz, gera um “presente estendido” caracterizado por, apesar de estar ocorrendo, não poder ser objeto de interferência do sujeito, a quem somente é permitido (ou exigido, sob pena de não participar do presente-contínuo veiculado pelas plataformas digitais) observar, atônito. O tempo humano é suprimido pelo tempo do capital, e a comunicação, que era pautada na mediação da palavra, necessária ao exercício da razão e reflexão, é substituída pela assimilação das imagens de maneira irrefletida. Nesse panorama, “a opinião pública se reduz a uma “sensação pública”, a uma “emoção pública”” (Bucci, 2021, p. 151). Esse processo não se dá de maneira incidental ou mesmo colateral, mas sim enquanto instrumento do capitalismo para melhor extrair olhar e dados para concentrar capital, poder e tecnologia (Bucci, 2021, p. 130-148).

Conforme estudamos anteriormente, as culturas se instituem por meio da comunicação que, por sua vez, age através de símbolos para constituir a forma como entendemos a realidade. Nesse sentido, de certa forma, toda a concepção de realidade da sociedade pós-moderna é pautada em uma espécie de virtualidade. A novidade do novo sistema instituído pela comunicação digital, ou da virtualidade real, se distingue justamente por concentrar em um mesmo ambiente que é também virtual, ou seja, o próprio digital, a experiência simbólica/material das pessoas, de forma que o digital não comunica a experiência, mas é, de fato, a própria experiência (Castells, 2011, p. 459-460).

Em um espaço de apenas uma ou duas gerações, a cultura que era expressa por meio da palavra impressa passou a ser cultivada pelas imagens eletrônicas, reconfigurando a ideologia em videologia, em que os signos que ganhavam materialidade pela palavra na instância da palavra impressa dão lugar aos signos derivados na imagem, na instância da imagem ao vivo, que corresponde “à mudança de forma das relações sociais, no embalo de um capitalismo que se especializou na fabricação da mercadoria como imagem” (Bucci, 2021, p. 181). A imagem se difere da linguagem, tanto pelo seu potencial de instigar desejos e instintos ocultos, quanto por não necessariamente utilizar do recurso à razão. Porém, a imagem tem servido cada vez mais à função de constituir signos, por meio de signos visuais compartilhados por uma comunidade global integrada, por exemplo por meio da logomarca de uma empresa famosa ou de um astro do esporte, de maneira que, apesar de não ser linguagem propriamente, se expressa enquanto linguagem na instância da imagem ao vivo (Bucci, 2021, p. 180-184).

Assim como a linguagem depende dos falantes para existir, pautada nos sentidos verbais comunitariamente constituídos a partir da fala, a imagem depende dos “olhantes”, que

reconheçam os signos compartilhados, de maneira que o olhar exerce “uma força constitutiva para a fixação do significado de cada figura” (Bucci, 2021, p. 184). Olhar e imagem se complementam para agir enquanto linguagem, em que a imagem exerce o papel de representar um objeto e ocultar uma subjetividade, em uma espécie de transparência pela qual o sujeito não vê a imagem propriamente, mas aquilo que está representado, que, por sua vez, é instrumentalizado pela indústria da imagem que atua para extrair o olhar, utilizando-o para confeccionar signos visuais, de forma que as imagens ganham força de significação, elemento característico da linguagem (Bucci, 2021, p. 184-188).

A comunicação digital utiliza-se das imagens, que independem da razão, para, por meio de técnicas de *machine learning* e inteligência artificial, que são os meios de produção, coletar os olhos e dados dos usuários para produzir *marketing* político, religioso e comercial. De fato, a lógica com que a imagem é entendida pode ser descrita em três momentos distintos: a lógica formal, característica da arquitetura, pintura e das artes, predominante até o Século XVIII; a lógica dialética, da fotografia e do cinema, predominante nos séculos XIX e XX; e a lógica paradoxal, que surge com a videografia, holografia e infografia que gera uma espécie de telepresença, suplantando a própria existência real, possibilitando a presença em múltiplos espaços em um mesmo momento, ao mesmo tempo em que múltiplos espaços subjetivos podem ocorrer simultaneamente dentro de uma pessoa, processo que se iniciou com a televisão e se potencializou nas mídias digitais (Bucci, 2021, p. 94-96).

Assim, tendo em vista as características da comunicação digital, estamos diante de uma nova forma de comunicação social, chamada auto-comunicação de massa. É uma forma de comunicação de massa vez que atinge, potencialmente, uma audiência global; é multimodal porque praticamente qualquer forma de conteúdo pode ser veiculada por meio das redes digitais; é auto-gerado, tem emissão auto-dirigida e recepção auto-selecionada pela forma como os indivíduos acessam essas redes. O que torna a comunicação digital tão particular, assim, não é a mobilidade em si, mas sim a conectividade perpétua, criando uma nova cultura de virtualidade real, na qual a virtualidade passa a ser uma dimensão fundamental da realidade (Castells, 2011, p. XIV-XVI).

O uso das plataformas digitais enquanto meio de comunicação é um fenômeno crescente, especialmente desde os anos 1990 e, no início dos anos 2020, a tendência ainda é de expansão do uso destas plataformas. Nesse sentido, um estudo realizado pela Kaiser Family Foundation constatou que os adolescentes nos Estados Unidos passam em média sete horas por

dia em tecnologias da comunicação e informação³ (Marichal, 2016, p. 1). Esse processo de conexão dos indivíduos foi ampliado com processadores avançados e tecnologia de conexão ponto-a-ponto, levando a uma aproximação no distanciamento no espaço/tempo, de forma que a maioria dos indivíduos, no início dos anos 2020, vive em uma *network society* (sociedade conectada), no termo utilizado por Castells, em que sua identidade é determinada por meio de suas relações. De fato, tal rede de conexões é tão intrincada que subsiste a questão quanto à existência de um “eu” distinto do “eu da *internet*” (Marichal, 2016, p. 1).

Nessa sociedade interconectada, com a ampla popularidade de redes sociais⁴, os indivíduos têm cada vez mais se retraído em suas confortáveis conexões pessoais, tornando-se cada vez menos capazes de aceitar e expressar dúvidas, flexibilidade e incertezas. As redes sociais, por meio de uma arquitetura da divulgação⁵ que leva os indivíduos a divulgarem mais e mais informações sobre si mesmos, conectando-se melhor com suas redes de conexões, porém tornando-se mais rígidos e intransigentes, favorece o florescimento de características que têm sido prejudiciais para as sociedades democráticas (Marichal, 2016, p. 1-2).

O surgimento das redes sociais, com o *Friendster*, *MySpace* e outros, possibilitou uma nova era dos meios de comunicação de massa, permitindo o compartilhamento de conteúdo em uma ou mais comunidades de interesse. Por meio dos interesses em comum, há a possibilidade de grupos de indivíduos criarem uma espécie de conteúdo co-produzido, ou *crowdsourcing*, união das palavras *crowd* (multidão) e *outsourcing* (terceirização), conferindo aos indivíduos uma capacidade de produção e divulgação de conteúdos jamais vista anteriormente (Marichal, 2016, p. 2).

As redes sociais funcionaram, inicialmente, como uma extensão da vida *offline* de um indivíduo, conectando indivíduos que se conheciam fora das redes sociais por meio de perfis com descrições de si próprios. A utilização de perfis, a possibilidade de elaborar uma rede de conexões e de aumentar essa rede, adicionando indivíduos que não eram conhecidos fora das redes sociais, levou a *internet* de um espaço anônimo para um espaço com “sujeitos conhecidos”, em que as relações são mediadas por meio de perfis com descrições dos usuários. O crescimento das redes sociais foi explosivo, de forma que em 2010, 43% dos usuários de *internet* nos Estados Unidos alegavam utilizar as redes sociais várias vezes ao longo do dia (Marichal, 2016, p. 2-3).

³ Information communication technology (ICTs).

⁴ Social networking sites- SNS

⁵ *Architerture of disclosure*, no original.

A primeira rede social foi criada em 1997 por meio de um site denominado *sixdegrees.com*, iniciando um processo de expansão deste tipo de plataforma no final dos anos 1990, especialmente por meio de sites que contemplavam determinada comunidade étnica ou racial, como as redes *MiGente*, *Black Planet* e *Asian Avenue* e tiveram seu alcance aumentado, no início dos anos 2000, por meio do site *Friendster* e com a explosão em meados dos anos 2000 de diversos sites com alcance maior, especialmente o *MySpace*, predecessor do *Facebook*. A plataforma *MySpace* se popularizou especialmente entre estudantes de ensino médio norte-americanos no final dos anos 2000 e, em 2004, foi lançado o *Facebook*, inicialmente restrito para usuários que utilizassem um e-mail da Universidade de Harvard. Entretanto, ainda no mesmo ano, a rede foi ampliada para usuários de outras universidades dos EUA e, em 2005, para estudantes de ensino médio dos EUA (Marichal, 2016, p. 2-3).

Em setembro de 2006 a plataforma foi disponibilizada para qualquer pessoa com mais de 13 anos que tivesse um endereço de e-mail, quando o *Facebook* começou a se popularizar. No início houve uma certa divisão de classe e raça entre o *Facebook* e o *MySpace*, com o primeiro sendo mais popular entre a população branca e mais escolarizada e o segundo entre a população mais pobre e minorias. Entretanto, o crescimento explosivo do *Facebook* suplantou o *MySpace*, superando-o em número de usuários em 2009, atingindo a marca de 500 milhões de usuários em julho de 2010, com um aumento de 300 milhões de usuários em apenas 15 meses. A empresa teve abertura de capital em 2005 e um grande investimento da *Microsoft* ainda em 2007, quanto esta investiu 240 milhões de dólares por 1,5% de participação (Marichal, 2016, p. 3-4). No Brasil, a estimativa de 2018 é de que havia 127 milhões de usuários mensais do *Facebook*, o que implica em que 60% da população utiliza a referida plataforma digital (Oliveira, 2018).

O sucesso das redes sociais, e especialmente do exponencial crescimento do *Facebook*, está intimamente relacionado com a necessidade humana de se conectar socialmente. Uma vez que o ser humano é um animal social, este tem a tendência de se identificar enquanto membro de um determinado grupo e a rede social facilita a criação e manutenção de conexões sociais desta natureza. De fato, manter relações sociais demanda o empreendimento de tempo e esforço, de forma que há limitações de natureza temporal e espacial para a construção de redes de conexão. Desta forma, as redes sociais permitem a criação e manutenção de relações com os indivíduos de maneira a exigir menos tempo e esforço. Nesse sentido, o *Facebook* facilita a comunicação entre os indivíduos e suas redes de amigos por meio de um mural de notificações, em que novidades como fotos, mudanças de *status* ou a participação em grupos na plataforma

ficam visíveis para os outros usuários, promovendo as relações sociais de maneira a superar algumas das barreiras das relações *offline* (Marichal, 2016, p. 4-6).

A utilização de plataformas como o *Facebook* corresponde à satisfação da necessidade humana de participar socialmente e da construção do chamado capital social, compreendido, segundo a concepção de Bourdieu, enquanto o conjunto de recursos, existentes ou potenciais, derivados de uma rede de relações mais ou menos institucionalizadas, de conhecimento e reconhecimento mútuos (1986, p. 248, *apud* Marichal, 2016, p. 36). Possibilidades como oportunidades de emprego, assistência técnica nos negócios ou oportunidades imobiliárias são alguns exemplos de benefícios derivados do capital social, e a estrutura de redes sociais, como o *Facebook*, promove um ambiente no qual são realizadas interações que criam e mantêm redes de relações interpessoais desta natureza. Os maiores níveis de capital social na plataforma foram encontrados em usuários de alta intensidade, identificados pelo nível de intimidade, reciprocidade e abertura emocional nas interações na plataforma (Marichal, 2016, p. 35-36).

Quanto à natureza das relações mantidas por meio das plataformas digitais, os usuários das redes sociais não apenas reproduzem sua rede de conexões *offline* nas plataformas, mas utilizam esta rede preexistente para criar uma nova *online*. Pesquisa de Parks (*apud* Marichal, 2016, p. 6) mostra que mais da metade dos amigos do *Facebook* são relações distantes, como conhecidos, amigos distantes ou amigos de amigos e que apenas 10% são declarados como amigos próximos, de forma que as redes de conexões no *Facebook* são versões expandidas, tanto espacial quanto temporalmente, das redes *offline*. É importante reconhecer, todavia, que a conexão pelo *Facebook* é uma forma de comunicação mediada por computador, de forma que se trata da intermediação de um computador para transcender limitações no espaço e no tempo, estabelecendo relações distintas daquelas chamadas *offline* (Marichal, 2016, p. 6).

Outro aspecto importante para caracterizar a natureza das relações *online* é que os usuários têm o poder de determinar quais são os sujeitos que estão em suas listas de amigos e que, conseqüente, podem ver as atualizações publicadas mutuamente. Assim, para compreender os impactos sociais dessa forma de comunicação mediada por meio das redes sociais, é preciso entender como as plataformas digitais modificam os processos de auto apresentação das pessoas por meio da *internet*.

As redes sociais, em especial o *Facebook*, estão pautadas sobre uma forma de arquitetura de divulgação de todos os conteúdos quanto possíveis relativos aos seus usuários, os quais serão transformados em *comodity*, utilizados para produzir propagandas individualmente direcionadas. Assim, o processo de construção da autoidentidade *online* está intimamente relacionado com duas características arquiteturais das redes sociais: o fato de que

as redes de amigos *online*, a despeito de geralmente serem mais amplas que a de amigos *offline*, também inclui esses últimos; e a possibilidade de socializar por meio de uma plataforma na qual o usuário determina quem pode acessar os conteúdos publicados e quais conteúdos gostaria de compartilhar. Essas duas características implicam na possibilidade de o usuário testar diversas formas de expressão de si mesmos, como versões diferentes de si mesmos, de acordo com a aceitação do público a que permitiu acesso, ao mesmo tempo em que tal perfil deve estar necessariamente pautado em características reais, uma vez que parte dos amigos *online* conhecem o usuário pessoalmente, o que impossibilita a criação de um perfil inautêntico (Marichal, 2016, p. 6-9). Desta forma, o comportamento dos usuários de redes sociais é marcado pela criação de um espaço pessoal dedicado a expressar uma imagem de si a uma rede de amigos selecionada, em busca de validação. Essa ênfase no pessoal, por sua vez, influencia de maneira negativa a forma como os usuários compreendem a política (Marichal, 2016, p.11-12).

As redes sociais constituem um espaço paradoxal no sentido de que, ao mesmo tempo em que são espaços de livre participação, com liberdade de expressão e direito a busca de informação, constituindo-se enquanto um espaço social público gerado pela comunicação, também é um espaço cuja participação é filtrada e os interesses privados influem no debate. Assim, as plataformas digitais são tanto privadas, uma vez que são geridas por empresas que se utilizam de *softwares*, algoritmos e gestão imperscrutáveis para regular e administrar esses espaços, quanto públicas, uma vez que há a participação de bilhões de usuários que podem se expressar por este meio. Desta forma, são espaços paradoxais que pleiteiam, concomitantemente, garantias individuais, como a liberdade de expressão e privacidade, e do interesse público, por meio das normas de países democráticos (Bucci, 2021, p. 96-100).

Se a esfera pública no conceito habermesiano exerce o papel de intermediação entre o mundo da vida e o aparato estatal e do mercado, um e outro se aproximaram, de forma a não mais constituírem dois polos interconectados pela esfera pública. Nesse panorama, é a própria esfera pública e a mediação racional que esta exercia que perde relevância na sociedade das mídias digitais, da qual decorre a crise dos intermediários, ou seja, instituições como o parlamento, a imprensa e os mecanismos de delegação da democracia representativa são suplantados (Bucci, 2021, p. 107-111).

Assim, constitui-se um verdadeiro teleespaço público, que se caracteriza: 1) pela materialidade virtual, em que os espaços públicos deixam de ser necessários, transcendidos pela imagem eletrônica que, apesar de virtual, constitui-se enquanto materialidade e não abstração, ou seja, expansão digital do espaço social; 2) a prevalência do inconsciente em detrimento do

sujeito racional, uma vez que o telespaço público se ordena em função de tensões e desejos inconscientes, através da imagem, e não da palavra escrita; 3) o telespaço público não se ordena pela busca pelo consenso, através do diálogo, mas sim pela ordem anárquica dos conflitos. As agregações, nesse sistema, não dependem de métodos racionais, mas sim de identificações; 4) a dominância do significante sobre o significado, em que o primeiro adquire maior relevância que o segundo, de forma que a visualidade e a discursividade imagética exercem papel central no telespaço público, em detrimento da palavra escrita; 5) a fragmentação, caracterizada pela coexistência de diversos espaços públicos simultâneos que se relacionam e recombinaem, em que o telespaço público constitui uma teia que abarca, confere unidade e interconecta todos os seguimentos possíveis; e 6) a supressão das mediações e dos intermediários, estabelecendo-se uma relação direta entre sociedade civil e Estado que privilegia as identificações e o culto das personalidades, em detrimento da razão crítica, terreno fértil para a disseminação de afirmações sem bases factuais e teorias conspiratórias (Bucci, 2021, p. 111-114).

As redes sociais são, assim, tecnologias do eu pautadas na ideia neoliberal de liberdade, uma vez que é voltado ao pessoal por meio de apresentações de si mesmo. O resultado dessas tecnologias é o desenvolvimento de uma sociedade autocentrada no indivíduo e menos conectada com estruturas maiores que forjam a sociedade, reduzindo os indivíduos de sujeitos políticos a meros consumidores. A arquitetura da divulgação do *Facebook* cria a necessidade de os indivíduos se exporem de forma positiva, em detrimento de apresentar-se de maneira negativa ou mesmo crítica, e, uma vez que se trata de uma comunicação mediada, retira a espontaneidade da vida *offline*. Ao criar um ambiente controlado para relações mediadas com sujeitos escolhidos pelos usuários, a comunicação resultante produz sujeitos menos confortáveis com a espontaneidade e imprevisibilidade, de forma a prejudicar o desenvolvimento de elementos essenciais para o conhecimento cívico: humildade, senso de imprevisibilidade, desejo de ouvir e a habilidade de perguntar-se: “e se eu estiver errado?” (Marichal, 2016, p. 81-84).

A superindústria do imaginário atua sobre o que separa o inconsciente do consciente, para ofertar os signos que acalmam a inquietude derivada da complexidade humana, de maneira que ao mesmo tempo em que apela (e lucra) para e sobre os desejos do inconsciente, oferta aos indivíduos uma impressão de hiperconsciência, que afasta a inconveniência do sujeito dividido, criando uma imagem de um “eu uno”, livre e senhor de si. A superindústria não o faz, entretanto, de acordo com um plano ou sob uma organização de corporações, mas sim por meio de uma “multiplicidade sem fim de imagens que são representações de outras imagens (...) em uma

espiral interminável (...) de modo a embaralhar as referências que se perdem em meio aos reflexos dos reflexos dos reflexos” (Bucci, 2021, p. 230).

Através desse processo que se dá no Imaginário, o sujeito se constitui enquanto um eu coeso. Ao longo da história, as religiões, associações do trabalho, o Estado e diversas outras instituições elaboravam de maneira difusa os discursos e signos que conferiam sentido aos sujeitos. Entretanto, na Superindústria do Imaginário, a função que era exercida por todas essas instituições foi capturada pela produção industrial de signos visuais (imagens) e objetos (sígnicos) que são tornados mercadoria e comercializados. Na Superindústria, portanto, não é apenas o entretenimento que exerce a função de fabricação de signos, mas a religião, a política, a ciência etc., são arregimentados pelo monopólio da máquina superindustrial sobre o Imaginário (Bucci, 2021, p. 240-244).

O capitalismo se especializou em produzir substitutos do objeto *a-* que é a falta derivada do desejo inconsciente insaciável, ao mesmo tempo em que é aquilo que se põe a frente deste, preenchendo-o, para torná-lo invisível, e, portanto, suportável- por meio dos signos, ou seja, da comunicação, para preencher o vazio do que falta ao sujeito dividido, pacificando o desejo (inconsciente) e conferindo sentido. O cerne da produção superindustrial é, portanto, a satisfação desta necessidade, no campo do desejo, e não a utilidade prática/material do objeto (Bucci, 2021, p. 242-245).

O sujeito do mundo moderno, especialmente do mundo Ocidental, neoliberal em sua construção, tem como elemento essencial a concepção de busca por ser autêntico, autônomo e individual. Porém, há uma estrutura rígida que determina os limites da autonomia dos indivíduos, forjando, na verdade, identidades conformadas. Mesmo as tentativas de escapar da rotina do sujeito neoliberal seguem uma espécie de rotina e estão constringidas pelos limites da liberdade de um sujeito conformado às estruturas do neoliberalismo. Ou seja, as formas de entretenimento existentes no mundo ocidental contemporâneo também estão vinculadas ao neoliberalismo e tem como objetivo promover, ao mesmo tempo, a sensação de espontaneidade e novidade, porém sem desviar-se severamente da rotina do sujeito neoliberal, como, por exemplo, ao assistir programas de TV familiares, ou fazer uma viagem à Disney (Marichal, 2016, p. 84-85).

A arquitetura do *Facebook* ajudou a plataforma a suplantear outras mídias sociais como o *MySpace* porque permite uma espécie de escape da rotina ao mesmo tempo em que reproduz estruturas rígidas. Enquanto o *MySpace* permitia maior liberdade de personalização ao usuário, de forma que as páginas dos usuários eram mais autênticas, as páginas do *Facebook* são padronizadas e são permitidas aos usuários poucas modificações. Esse ambiente controlado

desincentiva os indivíduos a se exporem ao desconhecido e, uma vez que a todo momento o sujeito está a apenas um acesso nas redes sociais para conectar-se com seu ambiente protegido, este passa a buscar o conforto social promovido pela plataforma digital mais frequentemente (Marichal, 2016, p. 84-85).

Já no final dos anos 1990, como Castells assevera com fulcro nos dados de uso da *internet* à época, havia importantes marcadores sociais, raciais, sexuais, de nível educacional e de faixa etária na quantidade e qualidade do acesso à *internet*, que se concentrava em classes sociais mais ricas, brancos, homens e mais jovens (Castells, 2011, p. 431-434). Apesar da rápida expansão no número de usuários, esses marcadores, conforme discutimos alhures, seguem presentes no acesso às redes ainda hoje.

Os marcadores da desigualdade no acesso à *internet* presentes no surgimento desta não foram eliminados ao longo do processo de popularização do acesso, existindo até mesmo formas de acesso distintos mediante esses marcadores. Já no final dos anos 1990, Castells identifica que, mediante a capacidade dos usuários das redes de computadores serem também produtores de conteúdo, esse acesso desigual produz impactos significativos nesse meio de comunicação (Castells, 2011, p. 439).

Essa divisão digital se dá pela rede de conexões estabelecida pelos usuários das redes sociais, as quais têm a tendência à homogeneidade, em reprodução do dia a dia dos usuários, baseado em raça ou *status* socioeconômico. Ademais, é preciso destacar as características do uso das plataformas digitais por diferentes grupos de usuários, pautada em dificuldades de acesso à *internet*, nível de autonomia do usuário, quantos pontos de acesso à *internet* o usuário tem disponível e o nível de experiência que o usuário tem. Tais características demonstram um uso muito distinto das redes sociais entre os diversos grupos, implicando em significativas diferenças na formação cívica por meio da *internet*. Por exemplo, dados de uma pesquisa com jovens adultos com idades de 18 a 26 anos com alto nível educacional demonstrou o uso da *internet* para atividades de “incremento de capital”, como visitar sites relacionados a emprego, adquirindo informação financeira e de saúde ou notícias, uso bastante distinto de outros grupos (Marichal, 2016, p. 25-26).

Os estudos conduzidos sobre o uso da *internet* têm apontado que as relações *online* refletem características das relações *offline*, privilegiando a participação de indivíduos de acordo com aspectos socioeconômicos ou de raça, por exemplo. Nesse sentido, pesquisa de Schlozman, Verba e Brady (*apud* Marichal, 2016, p. 85-87) sobre doações de campanha nas eleições presidenciais dos Estados Unidos de 2008 indicou que o índice de doadores de pequenas quantias *online* apenas refletia o índice de doadores *offline*, de forma que os autores

concluíram que as relações *online* são fortemente influenciadas por aspectos socioeconômicos *offline*, ou seja, amortecendo eventuais benefícios da criação de canais de expressão de discursos marginalizados. Ao mesmo tempo, estudo de Himelboim (*apud* Marichal, 2016, p. 85-87) apontou que mais de metade das discussões iniciadas *online* são feitas por meio do compartilhamento de materiais criados por terceiros, e não opiniões ou comentários do usuário que está compartilhando e que, deste material compartilhado, 60% são de fontes de mídia tradicional. Ou seja, grande parte das discussões *online* são apenas o eco de opiniões produzidas por meios tradicionais de comunicação e produtores de conteúdo (*marketing*, publicidade, influenciadores digitais etc.), com capacidade econômica e técnica para produzir conteúdo que será amplamente veiculado, conforme discutiremos posteriormente.

Ao serem atraídos para os ambientes controlados das redes sociais, evitando contatos com desconhecidos, e tais ambientes incentivarem a divulgação de um “eu” construído a partir da perspectiva da aprovação de uma rede de amigos, a percepção de conhecer as opiniões daquele grupo que parece ser o público, mas na verdade é um ambiente controlado pelo próprio usuário, pautado em recortes sociais, onde participam apenas indivíduos mais íntimos, cria a falsa sensação de “conhecimento sobre o mundo”. Assim, ao falar sobre política nos moldes das plataformas digitais, os indivíduos partem de um pressuposto de conhecimento do mundo construído em um ambiente familiar e confortável, evitando o desconhecido. Todavia, uma espécie de desconhecimento quanto ao posicionamento político dos indivíduos, de ideias privadas, faz-se necessário para a conexão entre sujeitos que têm ideias distintas, motivo pelo qual há uma certa escassez de conversas políticas na vida pública, que pode ser evidenciada com o ditado “sobre política e religião não se discute”. Ao conectar-se majoritariamente com indivíduos que “conhecem o mundo” da mesma forma que si, prevendo a aprovação pela performance do “eu” construído no ambiente digital, os indivíduos tornam-se menos propensos a um debate político propriamente dito (Marichal, 2016, p. 85-87).

1.3. A sociedade da vigilância: da racionalidade discursiva à racionalidade digital

A maneira como a comunicação digital se aparelhou e estruturou, desde seus estágios iniciais e de ampla popularização, foi pautada em interesses de mercado, especialmente do entretenimento, de maneira que as características assumidas pela estrutura da rede têm essa origem. A despeito do uso com outros objetivos, como educação, política, saúde, entre outros, o coração do desenvolvimento das mídias digitais é o entretenimento (Castells, 2011, p. 453-

455), mercado com ampla participação de grandes empresas de tecnologia, detentoras de plataformas digitais com milhões de usuários em todo o mundo.

A estrutura econômica deste mercado pode ser compreendida através da maneira como o *Facebook*, enquanto uma das plataformas com mais usuários no mundo, é monetizado. Tal estrutura deriva da escolha da empresa (realizada historicamente, enquanto decorrência da estrutura do capitalismo e, também, determinante para a lógica subsequente deste mercado), no momento de seu maior crescimento em meados de 2008, de se tornar lucrativa por meio de anúncios veiculados pela plataforma, em detrimento da cobrança direta de valores pagos pelos usuários. A atividade econômica está, assim, intimamente relacionada com a capacidade de gerar dados sobre os usuários, de forma a possibilitar a veiculação de anúncios direcionados aos usuários, conforme discutiremos a seguir. Desta forma, a estrutura do *Facebook* está pautada em incentivar os usuários a divulgar informações sobre si mesmos, possibilitando o direcionamento dos conteúdos de seus anunciantes para os públicos-alvo adequados (Marichal, 2016, p. 33).

Como vimos anteriormente, as tecnologias não se desenvolvem a despeito das relações sociais que, hoje, são demarcadas pelo capitalismo e o informacionalismo, mas sim em função destas. A escolha do *Facebook*, pioneira plataforma em alcance e participação ampla de usuários, quanto à sua estrutura de negócio, não é uma mera coincidência, mas uma decorrência desta estrutura. Ao mesmo tempo, como se impõe reconhecer, a estrutura de captação de dados para a comercialização de sentidos, deriva em muito das estratégias empresariais que foram realizadas pela empresa, como um ponto de inflexão do que se convencionou chamar mercado de tecnologia. A empresa, que reinou inabalável enquanto rede social predominante por algo em torno de uma década, inclusive adquirindo plataformas e serviços de concorrentes que foram acoplados aos seus serviços, vem sendo objeto de concorrência, nos últimos anos, por novas empresas e serviços mundialmente populares e, ineditamente, sediados fora dos Estados Unidos, como o *TikTok*, por exemplo, plataforma chinesa. Se por um lado essa concorrência e diversidade impacta o “mercado tecnológico” e as relações entre Estado e tecnologia, como se verifica da atuação da União Europeia e do Congresso dos EUA, interpelando às plataformas por transparência e responsabilidade, é imperioso reconhecer que a lógica que subsiste a todo o sistema é, essencialmente, a mesma.

Morozov chama atenção para o fato de que, na sociedade contemporânea, há a percepção de que a função das plataformas digitais está relacionada à ideia de que as empresas oferecem serviços tecnológicos que promovem bens tradicionalmente difíceis de serem conquistados, públicos e políticos, como a liberdade de expressão, a segurança e o transporte,

tornando-os mais eficientes. Entretanto, o interesse das grandes empresas de tecnologia, chamadas *Big Techs*, reside na aferição de lucros e estes, por sua vez, proveem da maior participação dos indivíduos na *internet*, não necessariamente da realização de valores sociais. Assim, para as *Big Techs*, não há distinção entre a veiculação de *fake news* ou notícias verdadeiras, uma vez que o lucro que percebem deriva da quantidade de interações nas redes sociais, de forma que não há interesse direto dessas em coibir atividades consideradas socialmente perniciosas. Desta forma, tem-se mostrado cada vez mais evidente a necessidade de alguma espécie de regulação das atividades das *Big Techs*, uma vez reconhecida a importância do fluxo de dados e os impactos sociopolíticos derivado destas (Morozov, 2018, p. 7-12).

Assim, a estrutura da plataforma do *Facebook* é voltada para a criação do que a economia comportamental define enquanto arquitetura da escolha, ou, especificamente na estrutura de comunicação digital, de arquitetura da divulgação. Enquanto uma perspectiva pautada nos indivíduos enquanto tomadores de decisões racionais levaria em consideração a liberdade dos indivíduos para publicar conteúdo ou não nas plataformas digitais, baseado em uma contraposição racional dos ganhos de determinadas ações em contraposição aos seus custos, a economia comportamental traz ao centro do debate a capacidade de instituições influenciarem a tomada de decisão dos indivíduos por meio de escolhas estruturadas, de forma que os custos associados ao convencimento para o comportamento desejado pela instituição sejam menores que a maneira como os indivíduos agiriam *ceteris paribus*. Ou seja, o *design* da plataforma digital é voltado para, de maneira inconsciente para os usuários, incentivar a divulgação de informações com a impressão de que se trata de uma escolha livre dos usuários (Marichal, 2016, p. 36-39).

O esforço para transformar divulgação e conexão em informações a serem comercializadas por parte das plataformas depende da relativização entre o que é considerado público e o que é considerado privado, criando um ambiente que é supostamente público, porém pautado por interesses de mercado da empresa. Assim, para criar a ilusão de que os usuários estão em um ambiente democrático e do qual a própria comunidade é proprietária, as plataformas permitem que existam, em seus ambientes, por exemplo, grupos e *posts* que critiquem as próprias companhias, ou responde a críticas às suas políticas afirmando estar ouvindo aos usuários e desenvolvendo aprimoramentos. O objetivo dessas ações é fazer com que pareça se tratar de um bem público, não de uma empresa privada. Por diversas vezes, por exemplo, o *Facebook* utilizou-se da seguinte estratégia para ampliar o acesso e a publicação de dados dos usuários: elabora mudanças que podem ser impopulares, diminuindo a proteção

quanto a informação sensível, necessária para o modelo de negócio (ou seja, para poder captar e tratar mais dados); explica para o público os supostos benefícios da alteração; detalha os passos realizados pela empresa para *supostamente* garantir privacidade e alegam ouvir os usuários; admitem a responsabilidade por não divulgarem adequadamente as alterações; e não fazem mais comentários (Marichal, 2016, p. 44-47).

Um exemplo de como a arquitetura do *Facebook* é desenvolvida para incentivar a divulgação é o *feed* de notícias. No início, caso o usuário quisesse saber as atualizações dos seus amigos, deveria entrar na página de cada um individualmente. Porém a inclusão do *feed* de notícias tornou possível acompanhar a atualização de todos os amigos em apenas alguns cliques, diminuindo muito os custos de transação, criando um “ambiente de intimidade” em que as pessoas se relacionam com uma regularidade que seria impossível de outra maneira, por limitações de tempo e espaço. Assim, a despeito de cada pequena interação na plataforma, como a publicação de uma foto, ser considerada insignificante, quando considerado o conjunto de interações é possível determinar com precisão inédita características da vida dos usuários, bem como de seus amigos e familiares, atingindo-se os objetivos da intermediária de geração de dados (Marichal, 2016, p. 39-41).

Outra maneira pela qual a arquitetura do *Facebook* facilita a divulgação é por meio das alterações nas suas políticas de privacidade. Se em 2005 os usuários tinham um vasto controle sobre as informações que eram compartilhadas, somente podendo ser acessadas por um grupo especificamente escolhido pelos usuários, em 2006 a configuração foi alterada para permitir que as informações quanto à escolha, ao local e outras informações relevantes passassem a ser compartilhadas e, no ano seguinte, que todos os usuários pudessem acessar o nome do usuário, sua escola e sua foto de perfil. Em 2009 foi permitido acesso a terceiros, inclusive não logados ao *Facebook*, a informações como nome, foto de perfil, listas de amigos, páginas de que era fã, gênero e região geográfica (Marichal, 2016, p. 41-43).

Mais importante que a divulgação aberta para os demais usuários ou quaisquer interessados, tais dados foram disponibilizados para aplicativos de terceiros (possibilitando a análise de metadados, ou seja, dados sobre dados) e, no ano seguinte, todas as curtidas e *posts* passaram a ser tratadas como informação pública, bem como informações sobre educação, cidade de nascimento, cidade atual, histórico de trabalho, *hobbies* e afins. Todas essas alterações, com vistas a ampliar a quantidade de informações geradas pelo aplicativo, foram feitas de maneira súbita e sem necessidade de concordância dos usuários, de maneira que, para proteger sua privacidade, os indivíduos deveriam acessar as configurações da plataforma e alterá-las manualmente, o que torna menos provável que o façam, especialmente tendo em vista

que, por diversas vezes, os usuários desconhecem a possibilidade de fazê-lo. Entretanto, sob a alegação de que os usuários aderem ao *Facebook* de maneira deliberada, considerando-os consumidores racionais, a empresa omite as estratégias de incentivo à produção e publicação de informações dos usuários, o que configura o principal ativo da empresa (Marichal, 2016, p. 41-43).

Uma das principais alterações realizadas pela plataforma para ampliar o alcance das informações coletadas dos usuários e as possibilidades de utilização comercial destas ocorreu em 2010, por meio de uma plataforma chamada *Facebook Connect*, um *application programming interface* (API), o qual permitiu compartilhar os dados dos usuários com aplicativos de terceiros, que permite estender o conhecimento sobre as atividades digitais dos usuários para além das fronteiras do aplicativo, abrangendo, idealmente (conforme a própria empresa expunha, para mostrar o alcance que tinham sobre o consumidor!), toda a atividade digital dos indivíduos, na construção do chamado gráfico social (*social graph*), ou seja, um mapeamento de todas as relações de determinado sujeito na *internet* (Marichal, 2016, p. 49-50).

Enfim, toda a lógica de redes sociais, como o *Facebook*, está baseada na ideia de que os usuários são consumidores racionais que tem a possibilidade de livremente aderir aos produtos oferecidos pela empresa e de deixar de utilizá-los. Porém, isso nem sempre é verdadeiro, tendo como exemplo a dificuldade para excluir uma conta na plataforma digital: para o processo seria necessário procurar o botão para deletar que é deliberadamente difícil de ser encontrado; ao clicar no botão, o usuário inicia um processo de duas semanas de espera para a exclusão e ainda é oferecida a opção de manter a conta desativa, porém ainda existente; após decidir realmente deletar, o usuário recebe uma mensagem que diz “seu amigo sentirá sua falta”, juntamente com as fotos em que ambos foram marcados; por fim, ainda que após o processo sua conta tenha sido excluída, há a possibilidade de seus dados já compartilhados serem utilizados pela empresa mesmo assim. Ainda que sob a ótica de consumidores racionais, é difícil compreender que há liberdade quanto à entrada e saída dos usuários da rede social. Porém, quando levada em consideração a necessidade humana de se conectar e pertencer, fica ainda mais evidente a dificuldade de um indivíduo optar por não utilizar uma plataforma que facilita a interação com seus amigos, sob pena de ser excluído de participar desse meio de interação, de forma que a liberdade de não participar está diretamente vinculada a um alto custo emocional e social (Marichal, 2016, p. 52-54).

Ademais, a despeito de os usuários compartilharem suas informações de certa forma deliberadamente por meio das plataformas digitais, o fazem com a finalidade de conectar-se com seus amigos e, geralmente, cuidam para transmitir uma imagem positiva de si mesmos e

preocupada com aspectos referentes à sua privacidade. Ao compartilhar suas informações preferencialmente e na maior parte do tempo com conhecidos e, ao mesmo tempo, ter interesse pelas publicações desse mesmo grupo, os usuários não estão utilizando seu tempo em um espaço público propriamente dito, mas ampliando as oportunidades de conexão com sua rede de amigos já estabelecida. Assim, apesar de a era das redes sociais ser a era da conectividade exacerbada, também é um momento de declínio da empatia: segundo estudo de Konrath e outros (*apud* Marichal, 2016, p. 55), os estudantes nos EUA eram 40% menos empáticos em 2010 do que eram quarenta anos antes, nos anos 1970, tendo resultados bem piores em índices de tomada de perspectiva e preocupação empática. O impacto social da intensificação das relações com sua rede de amigos por meio das redes sociais não ocorre às custas da vida pública, mas sim por incentivar que a vida pública passe a ser vista pelas lentes da vida privada (Marichal, 2016, p. 54-56).

As plataformas atuam, nesse sentido, para incentivar a conexão e o uso dos aplicativos que produzem, utilizando estratégias do chamado condicionamento operante, que por meio de punições e recompensas agem sobre o subconsciente dos usuários para conferir benefícios psicológicos pelo uso da plataforma, identificando-se por exemplo que o uso de redes sociais por apenas dez minutos aumenta o nível de ocitocina em até 13%, e prejuízos pela desconexão, gerando sentimentos de inquietação, ansiedade e impaciência. Assim, as redes sociais promovem um tipo de vício nos usuários, de forma que a escolha de conectar-se é (biológica, psicológica e socialmente) condicionada. Há, desta feita, uma espécie de consentimento dos usuários na troca entre conexão e compartilhamento dos dados, entretanto este é pouco informado e pautado em estratégias de condicionamento comportamental cuja correção é questionável (Deibert, 2019, p. 5-7).

Assim, o modelo de negócio das grandes empresas de tecnologia, como *Google* e *Facebook*, se caracteriza pelo oferecimento de um serviço, que geralmente é gratuito, em troca de informações sobre o comportamento dos usuários, o que vem sendo chamado de capitalismo de vigilância. As plataformas, entretanto, esforçam-se para não deixar claro esse *trade-off*, apontando como sua missão, por exemplo, “organizar toda a informação do mundo e torná-la universalmente acessível e útil”, ou conectar as pessoas, por meio de uma “comunidade”, como alegam *Google* e *Facebook*, respectivamente. Todavia, há um esforço constante para aprimorar o mapeamento de dados dos usuários, a fim de direcionar publicidade, levantando dados como “hábitos, relações sociais, gostos, pensamentos, opiniões, emoções, consumo de energia, batimentos cardíacos”, entre outros (Deibert, 2019, p. 4). Conjuntamente com a captação e tratamento dos dados por parte dessas empresas, há a atuação pouco conhecida de uma série de

empresas de análises de dados, que atuam nos bastidores, ficando em evidência apenas em escândalos, como no caso da *Cambridge Analytica* (Deibert, 2019, p. 3-4).

Fato é que a natureza do lucro das plataformas digitais é pautada no uso pelo maior tempo possível de seus produtos, munindo as empresas dos dados dos usuários necessários para a produção e divulgação de conteúdos direcionados, especialmente propagandas. Diante de debates públicos de grande importância sociopolítica, como períodos eleitorais e votações legislativas de grande repercussão, há um aumento na divulgação de materiais com desinformação. A estratégia retórica adotada é de natureza populista, pautada sobre uma suposta oposição às elites, a velha política e a mídia tradicional, integrando um projeto maior de desinstitucionalização, o qual, por sua vez, resulta em reações das instituições, por exemplo por meio da CPMI das *Fake News* ou do inquérito sobre o tema no Superior Tribunal Federal, conforme estudaremos adiante (Amato; Saba; Loschiavo Leme de Barros, 2022, p. 544-545).

A estrutura do sistema de vigilância, em contraposição à transparência socialmente estimulada, omite-se, ao fundir-se ao cotidiano por meio do uso de tecnologias que fornecem os dados necessários para a dominação. Ou seja, pelo uso cotidiano de *smartphones*, relógios inteligentes, aparelhos domésticos conectados à rede, redes sociais e afins, a aparência oferecida aos usuários é de um sistema que preza pela funcionalidade, serviços e liberdade, ocultando-se o interesse na coleta de dados. Todas essas tecnologias são pautadas, portanto, não na coação, mas no incentivo à transparência dos usuários aliado à sensação de liberdade, influenciando o comportamento dos indivíduos em um nível subconsciente, de forma a afetar toda forma de sociabilidade, por meio de uma “psicopolítica dado-pulsional”, ou seja, por meio de instrumentos que atuam por meio dos dados para gerar desejos e comportamentos (Han, 2022, p. 15-24).

As plataformas digitais possuem diversos acordos de compartilhamentos de dados com outras empresas e fabricantes de dispositivos que garantem lucros exorbitantes para o ecossistema da economia de vigilância. Até mesmo empresas de outras áreas passaram também a coletar os dados e atuar em parceria para oferecer produtos aos seus consumidores, como por exemplo empresas de transporte aéreo que atuam conjuntamente com redes hoteleiras, companhias de táxi e agências de viagem. Desta feita, os dados coletados são da mais variada gama, e as plataformas digitais, ao oferecerem serviços, têm como contrapartida o acesso aos dados de seus usuários (Deibert, 2019, p. 4-5).

Esse *trade-off* entre o serviço oferecido pela plataforma e o compartilhamento dos dados pelos usuários é feito de maneira consentida, uma vez que a maior parte das redes sociais exigem a aceitação de termos de serviço para acessá-las. Entretanto, seja pela linguagem

jurídica pouco acessível ou pela facilidade de aceitar-se os termos sem lê-los, a vasta maioria dos usuários não se atenta às autorizações que conferem aos aplicativos e, mesmo quando tomam conhecimento, ainda assim podem entender que é uma troca vantajosa. Além disso, pela complexidade do funcionamento das plataformas, que se constitui por camadas de algoritmos e propriedades intelectuais protegidas por lei e acordos de confidencialidade, é difícil que os usuários compreendam de fato ao que estão aderindo, ainda mais levando-se em consideração o objetivo que as plataformas alegam ter: apenas gerar mais conexão (Deibert, 2019, p. 5-7).

As relações nas redes sociais estabelecem uma espécie de controle social que, apesar de ser dotado de características específicas, uma vez que as redes de amigos são controladas pelos usuários, constroem o conteúdo e a forma do que é postado pelos usuários. Essa espécie de panóptico, compreendido, no conceito de Foucault, enquanto uma forma de interiorizar comportamentos em função da permanente visibilidade perante os demais, adquire contornos específicos por meio das redes sociais uma vez que o conteúdo é deliberadamente postado pelos usuários, de forma a constituir uma espécie de “panóptico participativo”. Tendo em vista que a natureza do controle exercido sobre a maneira como os indivíduos se manifestam por meio das redes sociais é mais associada ao coletivo que à uma autoridade, sendo mais horizontal que hierárquica, pautada sobre a interatividade e não a disciplina, seria melhor caracterizar por “vigilância participativa”, substituindo a ideia de panóptico (Marichal, 2016, p. 75-77).

Essa necessidade de vigilância na sociedade moderna é compreendida enquanto resultado paradoxal da exacerbação da privacidade em detrimento da vida pública que forjou, por sua vez, uma sociedade de estranhos, em que os indivíduos são tratados como desconhecidos no meio social, resultado do individualismo e isolamento social. Em um primeiro momento, as redes sociais como o *Facebook* podem parecer estabelecer conexões entre os indivíduos, tornando-os mais próximos e portanto diminuindo a necessidade de vigilância, entretanto a maneira como as pessoas se relacionam pelas redes sociais, criando identidades de si mesmos pautadas na aceitabilidade de terceiros, gera um receio nos usuários quanto à autenticidade das identidades de terceiros, além da segmentação entre os usuários, que tendem à homogeneidade, de forma a afastá-los ainda mais no sentido de um hiper ceticismo que incentiva, por sua vez, a vigilância social na *internet*, uma espécie de vigilância lateral, realizada pelos próprios usuários entre si (Marichal, 2016, p. 75-77).

Autores da Teoria Crítica como Adorno e Habermas identificam a dificuldade na sociedade moderna de se construir uma verdadeira individualidade, perante a atração da sociedade de consumo, a crescente sofisticação da propaganda e a profissionalização da comunicação política, que rendeu os indivíduos a identidades conformadas que promovem o

consumo baseado nos interesses do mercado. O processo de conformação do indivíduo moderno com o poder político e a sociedade do consumo deriva do esforço de instituições sociais e políticas, desde a Primeira Guerra-Mundial, de trazer para o centro do conceito de sucesso pessoal e profissional o cuidado com si mesmo, em detrimento de uma vida virtuosa e ética. Desta feita, as ferramentas utilizadas para determinar o que é uma boa vida, o que é denominado por Foucault de tecnologias do eu (ou de si), foram voltadas à promoção da satisfação pessoal e da autoestima (Marichal, 2016, p.77-81).

O que se considera autocuidado e autodesenvolvimento é resultado de normas públicas arregimentadas pelo poder político e a sociedade de consumo, de maneira a forjar sujeitos dóceis com aspirações que coadunam com os interesses do próprio poder político e do mercado, de forma que o aparato regulatório do Estado Moderno não é imposto por um agente externo, mas sim internalizado pelos sujeitos. Por esse motivo, o indivíduo autocentrado é ideal para a sociedade neoliberal, uma vez que o indivíduo que busca a autossatisfação raramente questiona a eficácia do liberalismo e do sistema capitalista, vendo-se enquanto um sujeito ativo e dotado de liberdade e escolha, no controle para escolher os diferentes tipos de consumo. As redes sociais são um palco ideal para a propagação de identidades conformadas baseadas na legitimação neoliberal e do acúmulo do capital, uma vez que representam uma espécie de “refeudalização” da esfera pública, transformando-a em propriedade de algumas poucas companhias que são pautadas por interesses de mercado (Marichal, 2016, p.77-81).

O ambiente das plataformas digitais não se constitui, nesse sentido, de uma aldeia global pós-capitalista, conferindo capacidade semelhante entre os indivíduos para manifestarem-se, mas sim em uma espécie de domínio feudal de poucas empresas, privadas, em sua maioria sediadas nos Estados Unidos da América e que são dotadas de interesses particulares. Há pouca resistência perante a hegemonia de plataformas estadunidenses, exceto algumas poucas empresas chinesas ou russas (e mesmo essa resistência é dominada e funciona sob a lógica do capital).

Compreendemos, portanto, que a construção da concepção das plataformas digitais enquanto espaço para livre participação dos indivíduos e de liberdade de expressão tem seu fundamento na contracultura tecnológica dos EUA, capitaneada pelas empresas de tecnologia do Vale do Silício, com fundamento na perspectiva neoliberal de desenvolvimento interno e crescimento pessoal (aspecto individual) e da necessidade de mais capitalismo (aspecto social ou coletivo), coadunando com a organização econômica e social que tece todas as relações interpessoais, o capitalismo e o informacionalismo.

Assim, se mediante o capitalismo industrial desenvolveu-se o chamado regime disciplinar biopolítico, em que por meio da disciplina sobre os corpos e as energias, os indivíduos submetiam-se às estruturas do capital, tornando-se dóceis para o sistema, diante do capitalismo da informação a forma de dominação econômica, política e social, não é mais pautada na repressão do regime disciplinar, mas sim na captação e processamento de dados dos indivíduos por meio das tecnologias de comunicação e informação. Se no primeiro sistema o isolamento espacial e a regulamentação rigorosa do trabalho eram os instrumentos utilizados, na estrutura do panóptico social, submetendo os indivíduos a uma sensação de vigilância latente constante, na sociedade da informação o domínio reside sobre a percepção de liberdade, autenticidade e criatividade dos indivíduos, possibilitando a geração da matéria-prima do poder na sociedade contemporânea: a conexão que gera dados (Han, 2022, p. 7-12).

A mercadoria, na lógica vigente, é destinada ao desejo (inconsciente) do sujeito dividido, mas é justificada para o “eu” por meio de uma ideia de utilidade, para que, ao final, o sujeito possa encontrar o gozo pela comunicação. A mercadoria perde seu nexos com a utilidade objetiva (necessidade), da qual independe, passando a exercer apenas o papel de prover o gozo imaginário (valor do gozo) e, ao se libertar do aspecto material, permite a acumulação irrestrita do capital (Bucci, 2021, p. 292-297). Ao final, o gozo não se apresenta como opção, mas, no “capitalismo superindustrial, a oferta do gozo imaginário se apresenta como imperativo moral, que manda gozar.” (Bucci, 2021, p. 297).

Enquanto o sujeito moderno recebia uma ordem do imperativo moral que lhe tolhia o gozo, o reprimindo **para** que não saísse do lugar, o sujeito supermoderno é interpelado o tempo todo a gozar, e o faz exatamente **por** não sair do lugar, ou seja, por ser imobilizado pela oferta constante do gozo imaginário que, entretanto, jamais poderá ser efetivamente consumado. “A busca pelo objeto do desejo”, assim, “nunca cessa”, de maneira que depende de cargas sempre maiores. A sobreposição do valor do gozo sobre a lógica vigente no capitalismo superindustrial foi possível pela extração não apenas do “trabalho socialmente necessário empregado na fabricação da coisa corpórea”, mas principalmente pela extração do olhar social para produzir a imagem da mercadoria (Bucci, 2021, p. 297-299).

Ou seja, tanto no regime disciplinar, quanto no regime da informação, a vigilância é elemento essencial para a dominação. O que distingue um sistema do outro é que, no primeiro, o isolamento dos corpos era o instrumento utilizado, enquanto, no segundo, a conexão, possibilitada pelos instrumentos tecnológicos, é que se alia à vigilância. Enquanto no primeiro é a repressão dos corpos que exerce papel central, no segundo, liberdade, conexão e vigilância convergem enquanto elementos de dominação, de forma que as pessoas não se sentem vigiadas,

mas, sim, livres, o que, por sua vez, é pressuposto para a estrutura de compartilhamento e captação de dados. Na sociedade pós-Moderna o sujeito não precisa ser coagido à vigilância, mas oferece-se e põe-se sob vigilância de maneira “deliberada”, mediante a existência de um imperativo social de transparência. Nessa sociedade, não são as pessoas que são livres, mas sim a informação (Han, 2022, p. 12-15).

O papel dos agentes que se propõem a tentar produzir significação através de determinados interesses, “por detrás dos balcões de atendimento da publicidade, do entretenimento, das relações públicas, das telerreligiões, da comunicação corporativa e das grandes celebrações esportivas” (Bucci, 2021, p. 322), é de um pré-trabalho para a elaboração de uma proposta de signo a ser exposta ao olhar social que, só então, poderá se constituir em significado. A trama desses agentes, desta forma, não manda no olhar e, tampouco, o olhar manda na trama (Bucci, 2021, p. 322-324).

“Nas *big techs*, o grau de exploração da Superindústria do Imaginário chegou a um patamar de tapeações e ocultamentos tão requintado que nem os mais sovinas, sagazes e impiedosos barões da Revolução Industrial ousariam supor. Numa rede social ou num grande site de busca, o “usuário”, que imagina usufruir de um serviço que lhe é ofertado em generosa cortesia, é a mão de obra (gratuita), a matéria-prima (também gratuita) e, por fim, a mercadoria (que será vendida, no todo ou em partes, em esquitejamentos virtuais, e nem desconfia da gravidade disso). Nunca o capitalismo desenhou um modelo de negócio tão perverso, tão acumulador e tão desumano.” (Bucci, 2021, p. 326).

Essa estrutura econômica da comunicação digital possibilita o uso de instrumentos de psicometria, também chamado de psicografia, que através da captação de dados de determinado sujeito é capaz de prever suas reações, melhor até mesmo que um amigo ou familiar ou, a depender do volume de dados, a própria pessoa. Se no regime disciplinar característico do capitalismo industrial os dados demográficos permitiam um maior controle dos corpos, enquanto instrumento da chamada biopolítica, no regime da informação a obtenção de dados psicográficos permite o uso de ferramentas que possibilitam a chamada psicopolítica. É por meio destes processos que empresas que atuam em parceria ou com a obtenção dos dados de plataformas digitais exercem o chamado *microtargeting*, ou seja, o envio e produção de conteúdo personalizado de acordo com dados psicográficos capturados do usuário individualmente. O comportamento, através destas estratégias, passa a ser influenciado no nível do subconsciente, como o comportamento de consumo, prejudicando dois elementos fundamentais da democracia: a autonomia e a liberdade de vontade (Han, 2022, p. 37-39).

A regulamentação da comunicação digital, por meio da legislação, precisa partir da compreensão de que os meios de comunicação não são meros distribuidores de conteúdo, mas

sim extrativistas do olhar para a fabricação da imagem da mercadoria, ou seja, são meios de produção de valor de gozo, explorando o trabalho do olhar. Ao mesmo tempo, os dados coletados dos usuários são analisados pelas plataformas para perscrutar o desejo inconsciente, sem que até mesmo o próprio usuário tenha melhor consciência deste. O modelo de negócio das *big techs* gera assimetrias de informação e exerce um controle não transparente sobre o olhar, as ideias e as imagens e, desta forma, distorce o debate público e os processos decisórios que envolvem participação popular. Limitar o alcance do voraz extrativismo do olhar da superindústria do imaginário, assim, não se afigura um confronto à liberdade de expressão, mas, na verdade, uma necessidade democrática que ainda não tem sido adequadamente reconhecida. Não se trata, é fundamental dizer, de regular a tecnologia, mas sim o modo de produção que a alimenta, que permite que “a formação da subjetividade, a integridade psíquica e os circuitos personalíssimos do desejo de cada um e cada uma” (Bucci, 2021, p. 330-331) sejam comercializados (Bucci, 2021, p. 326-330).

Existe, portanto, uma tensão entre a racionalidade digital que é pautada no entendimento de que o comportamento humano pode ser plenamente ou, ao menos, satisfatoriamente e preferencialmente, entendido e influenciado por meio da capacidade de captação e processamento de dados e que, portanto, busca proporcionar ainda mais situações em que os indivíduos compartilham os seus dados e melhoria no processamento desses, por meio do aprendizado de máquina, e a racionalidade discursiva, um dos fundamentos da democracia, que se sustenta sobre a capacidade de argumentar e dialogar, tendo enquanto pressuposto um pano de fundo de existência social comum, enquanto forma de realizar bens socialmente desejáveis e promover qualidade de vida.

1.4. Os espaços públicos mediados e a epistemologia tribal

A visão sobre o impacto social que seria causado pela adoção massiva da comunicação digital se modificou substancialmente ao longo dos anos. No final dos anos 1990 e início dos anos 2000, havia duas grandes correntes de pensamento sobre os impactos da *internet* na vida pública que se contrapunham: a visão utópica da *internet* e a visão distópica. A visão utópica se subdivide em duas correntes, que são uma utopia liberal (privada) e uma comunitária (pública), a primeira centrada na liberdade e no individualismo e a segunda nos benefícios comunitários das redes sociais. É a comunhão de características das duas correntes que explica o apelo e o sucesso da narrativa das plataformas digitais, compreendidas enquanto meio de

promoção da vida sociopolítica, pautado nas ideias de união da liberdade individual, proveniente da visão liberal, com a percepção de benefícios para a coletividade, derivada da visão comunitária (Marichal, 2016, p. 17-19).

Por outro lado, já no final dos anos 2000 e início dos anos 2010, propagou-se uma terceira corrente sobre o uso da *internet*, pautada em certo pessimismo em contraposição à perspectiva utópica, gerando uma visão distópica do uso das referidas plataformas. Essa corrente aponta algumas questões que são ignoradas pelas correntes utópicas, das quais destacam-se o excesso de informações; o viés de seletividade; o *slacktivism*; a divisão digital; e o declínio da empatia e amizade.

Se a perspectiva utópica previa a possibilidade de facilitação no acesso às informações, o uso efetivo da *internet* e especialmente das mídias digitais apontam que há produção de um excesso de informações, o qual, por sua vez, gera interações mais curtas e menos significativas, resultando na ausência de busca por confirmação da veracidade do conteúdo. O volume de informações a que os indivíduos estão expostos geram a tendência de resumir as informações, resultando no consumo irrefletido dos conteúdos a que os indivíduos estão expostos (Marichal, 2016, p. 19-21). É fundamental levar em consideração, ainda, que o acesso a redes sociais está vinculado, principalmente, à busca por socialização, e não na procura por informações, de forma que ocorre a exposição a conteúdos segmentados de acordo com a rede de amigos e interesses do usuário.

Assim, um dos maiores prejuízos à vida pública derivada da comunicação digital deriva do uso das plataformas para socialização e não em busca de informação, uma vez que há a tendência de formar redes de amigos baseadas em características compartilhadas, o que é denominado viés de seletividade. Ou seja, o que ocorre na plataforma não é a desinformação neutra, mas a exacerbação da socialização com pares ideologicamente semelhantes, o que tem levado à crescente polarização característica das democracias contemporâneas. De fato, participar de discussões políticas nas redes sociais tem demonstrado um aumento na participação política em ações como votar, doar para campanhas ou tentar persuadir outras pessoas, porém, ao mesmo tempo, leva à polarização. Esse fenômeno é ainda mais intenso quando observado o uso de *blogs* ou outras mídias digitais em que se discute diretamente sobre política, uma vez que as redes de amigos com ideologias distintas são mais amplas nas redes sociais propriamente ditas. De fato, parte da rede de amigos das redes sociais é composta por familiares e amigos conhecidos no dia a dia, os quais tem opiniões distintas, o que, por sua vez, leva a um abrandamento no diálogo político por meio das redes sociais em contraposição a plataformas especificamente voltadas para a discussão política (Marichal, 2016, p. 21-23).

Evidencia-se, portanto, que a formação de identidades políticas em um mundo cada vez mais conectado pelas redes sociais passa por modificações importantes. Alguns estudiosos têm apontado que a facilidade de comunicação por meio das redes sociais diminuiu os custos de transação envolvidos no ativismo político, de forma que mais pessoas conseguem participar politicamente por meio das plataformas digitais. Segundo pesquisa realizada por Park e outros (*apud* Marichal, 2016, p. 9), há quatro motivos principais pelos quais os usuários do *Facebook* participam de grupos: socialização, entretenimento, busca por *status* social e informação. Os autores identificaram que estudantes em busca de informações ou socialização no *Facebook* tinham maior inclinação a participar ativamente da política. Outro estudo, de Feezell e outros (*apud* Marichal, 2016, p. 10), identificou que indivíduos que participam de grupos políticos no *Facebook* tem maior nível de atuação política, demonstrando maior comprometimento com votação e expressando maior interesse em política (Marichal, 2016, p. 9-10).

De fato, segundo pesquisa realizada pelo Instituto DataSenado no ano de 2019, cerca de 45% dos entrevistados alegaram ter decidido voto nas eleições levando em consideração informações vistas em alguma rede social e, quando perguntados sobre se o conteúdo postado em redes sociais influencia a opinião das pessoas, 83% responderam que as redes sociais influenciam muito (Baptista, 2019), de forma a evidenciar a importância das redes sociais sobre o posicionamento político e suas formas de expressão contemporâneas.

Entretanto, o maior engajamento político em redes sociais não tem se associado a níveis maiores de informação política, uma vez que os debates em grupos de redes sociais são caracterizados pela incivilidade, afirmações falsas e meias-verdades. Ademais, há a possibilidade de a participação política por meio das redes sociais, por ser facilitada, ter o efeito negativo de inibir a atuação política efetiva, uma vez satisfeita a necessidade de participação dos sujeitos por meio da manifestação em redes sociais. Tal comportamento foi descrito por Morozov como *slacktivism* e por White como *clicktivism*, indicando a falta de comprometimento característica da atuação política pela *internet* (Marichal, 2016, p. 10-11).

Assim, sob a perspectiva utópica privada do uso da *internet*, de natureza liberal, sugere-se que a *internet* devolve o poder ao indivíduo e ao mercado, suplantando o poder do Estado, resultando em uma sociedade cibernética baseada na concepção neoliberal de cidadania, tendo uma visão otimista de uma sociedade mais tolerante, menos nacionalista e mais adepta aos valores do livre mercado. Posteriormente, a ideia de que os indivíduos poderiam tornar-se criadores de conteúdo gerou a visão otimista de que a divulgação de ideias seria mais democrática e, no *Facebook*, a liberdade para criar e divulgar conteúdos, bem como de escolher quais são os demais usuários que compõem sua rede de amigos que estão habilitados a acessar

tais conteúdos, está em consonância com a ideia de liberdade de escolha, para além das fronteiras do Estado nacional e das autoridades públicas. A *internet* livre seria um ambiente em que as noções utilitárias do bem seriam realizadas, por meio da liberdade de comércio e da incapacidade dos Estados de restringir o conteúdo postado (Marichal, 2016, p. 26-28).

A utopia derivada da visão comunitária, complementa esta perspectiva, amplamente difundida pelas big techs, de que a *internet* e as redes sociais são possibilidades de transformação da comunidade para o bem, por meio de mais conexão. As tecnologias de comunicação resultariam em uma cultura livre, participativa, em uma sociedade de pessoa para pessoa (Marichal, 2016, p. 28-29).

Todavia, a construção de uma esfera pública (no sentido político, de convivência sociopolítica) deriva, evidentemente, do fato de que o ambiente de participação não é privado, de forma que se faz necessária a existência de um projeto público comum a todos. As redes sociais, como o *Facebook*, a despeito de parecerem constituir uma esfera pública, são, na verdade, compostas essencialmente por esferas privadas, uma vez que os usuários têm a capacidade de determinar quem são os demais usuários que comporão suas redes de conexões e acessar os conteúdos postados (Marichal, 2016, p. 28-31).

As tecnologias de informação e comunicação, em especial as redes sociais, forjam uma espécie de espaço público mediado, em que o público é filtrado por meio, principalmente, do interesse em determinado tema ou pela proximidade entre os usuários (amizades, familiares, colegas de trabalho etc.). O material publicado nas redes é de temas muito variados, porém os usuários tendem a postar temas de acordo com suas inclinações pessoais e predisposições e, tendo em vista a natureza das interações por meio das plataformas, geralmente mais efêmeras em face da vasta quantidade de conteúdo publicado no *feed* de notícias diariamente, os usuários têm a tendência de postar conteúdos com forte apelo emocional para angariar atenção (dos demais usuários e, conseqüentemente, do algoritmo) (Marichal, 2016, p. 60).

Nesse sentido, a maior parte dos conteúdos sobre política no *Facebook* está relacionada a problemas localizados, específicos, que se preocupam com questões pessoais a despeito de uma reflexão sobre problemas maiores, refletindo a estrutura das interações na plataforma. Quando temas de interesse público de maior proporção são pautados nas redes sociais, a tendência é que sejam publicados materiais de grande apelo emocional, geralmente vinculados a sentimentos negativos, e que não debatem efetivamente as questões pertinentes ao tema, mas que apenas demonstram o descontentamento de um grupo com determinado fato, sem engajar a reflexão (Marichal, 2016, p. 61-65).

Portanto, temas de interesse público, quando debatidos por meio das redes sociais, refletem a arquitetura desenvolvida pelas plataformas para expressão de sentimentos e emoções pessoais, em busca de autoafirmação e aceitação social, ressalte-se, apenas da rede de amizades selecionadas, não do público geral. Expor-se ao debate público propriamente, entretanto, enquanto construção de uma aptidão para o público, ou uma espécie de conhecimento cívico, implica em expor-se ao desconhecido, por meio da conversa, e lidar com o medo, a incerteza, o desconforto, a solidão, o sentimento de desconexão, enfim, sentimentos que fazem com que os sujeitos se conectem com o mundo impessoal (público) e reflitam sobre suas narrativas preconcebidas. Trazer para o centro do que se supõe debate público, e para a forma como os indivíduos veem a esfera pública, esse eu mediado, pautado no que se espera de si e em uma auto apresentação, em detrimento da reflexão introspectiva, representa a transposição da visão de uma sociedade compartilhada, ou uma comunidade impessoal, para uma discussão quanto aos sentimentos dos indivíduos como elemento central da vida em sociedade, ou seja, para uma perspectiva privada da vida pública (Marichal, 2016, p. 65-71).

O que é denominado arquitetura de divulgação, portanto, incentiva a personalização, enquanto na política conceitos como bem comum privilegiam uma espécie de consenso social em detrimento dos interesses particulares, de forma que o primeiro incentiva a manifestação de experiências individuais, enquanto o segundo as submetem ao coletivo. Essa abordagem mais vinculada a experiências individuais pode produzir um debate público que privilegia determinados discursos marginalizados que seriam deixados de lado na busca de um consenso. O resultado é que a cidadania, na sociedade moderna, não está vinculada à esfera pública formal e impessoal, mas é exercida, ao contrário, em uma espécie de “cidadania íntima”, pautada nas relações e experiências individuais. A comunicação mediada por computadores favorece o estabelecimento e ampliação desta perspectiva privada da vida pública, entendida enquanto política. As plataformas digitais, em função da forma como são estruturadas, não criam fóruns de deliberação, resultado do debate, mas sim espaços de “proclamações pessoais”, ou seja, é uma forma de promover a expressão de compreensões individuais, ainda que compartilhadas com determinado grupo que corrobora tal perspectiva. (Marichal, 2016, p. 89-92).

A atividade de reflexão em colocar-se no lugar do outro é ausente na comunicação digital, impedindo ou dificultando o exercício dialético que constrói a opinião, de forma que a comunicação passa a se pautar na autopromoção ou autopropaganda, que encontra reverberação nas câmaras-de-eco derivadas dos filtros-bolha. Ao discurso, elementar à qualidade da democracia, é inexorável a escuta atenta, sem a qual não há ação comunicativa, não há o reconhecimento do outro. Os filtros-bolha, entretanto, mantém o sujeito imerso em conteúdos

sugeridos de acordo com as curtidas ou compartilhamentos do próprio usuário, criando uma espécie de *looping*-do-eu, reduzindo o horizonte de assuntos que ganham relevância por meio das redes aos limites do que já interessa ao próprio sujeito. Por não haver um reconhecimento de problemas comuns à comunidade, não há sequer a formação de um espaço público (Han, 2022, p. 49-53).

Assim, ter um meio de expressão é apenas um dos elementos de uma democracia. Outro pressuposto de um debate político é que os participantes sejam, em alguma medida, ouvidos, sendo necessário, portanto, algum grau de deliberação que, por sua vez, depende das capacidades de ouvir ao outro, estar aberto à mudança de opinião e a vontade de conviver com a diferença, capacidades que, conforme exposto, são prejudicadas pela estrutura das plataformas digitais (Marichal, 2016, p. 92-101).

O uso de técnicas para tratar e manipular os dados provenientes das mídias digitais constitui um novo aspecto da comunicação durante períodos eleitorais, mediante a vasta quantidade de dados que são capturados em tempo real pelas mídias digitais, possibilitando que os conteúdos sejam expostos a testes, com o objetivo de influenciar a maneira como os cidadãos votam, através da avaliação dos hábitos, preferências, opiniões e emoções dos usuários. Por meio desta técnica, um material pode ser testado quanto ao seu alcance, impacto e reações em tempo real, tornando possível que a rota seja ligeiramente corrigida a todo o momento, testando-se os conteúdos em um ciclo contínuo até que gerem o resultado desejado. Apesar de não poder influenciar diretamente a tomada de decisão de cada indivíduo, tais técnicas atuam direcionando o debate para determinado sentido, atuando sobre o aglomerado como um todo. O processo passa, portanto, pela identificação dos perfis que podem ser influenciáveis, ou que têm dúvidas, em determinada votação; a sua categorização de acordo com interesses e desejos; a produção de centenas ou milhares de conteúdos que serão testados e aprimorados em ciclo contínuo através dos dados coletados em tempo real sobre a aceitação do conteúdo; e o reforço dos conteúdos bem-sucedidos, que é constantemente aprimorado (Empoli, 2019, p. 116-124).

Os conteúdos produzidos, assim, não necessariamente contém a mesma mensagem, de forma que para um grupo pode ser elaborada mensagem que exalte um aspecto e para outro uma mensagem diametralmente oposta, sutilmente elaborada e testada, direcionada individualmente e, por posteriormente serem veiculados pelos próprios usuários, dá-se ainda um véu de veracidade, ou uma impressão de que o conteúdo é orgânico, e não fruto de uma produção industrial sobre os dados dos usuários, em uma estrutura que inicialmente foi concebida para alcançar os indivíduos enquanto consumidores, mas que posteriormente passou a ser utilizada para fins políticos. Se antes das mídias digitais uma campanha política bem-

sucedida deveria tentar produzir mensagens que alcançassem a todos ou, ao menos, não os indispussem com determinados setores, hoje a individualização da disseminação de conteúdo torna possível que as mensagens sejam produzidas de acordo com os anseios individuais, sem que o todo seja exposto à estratégia. Ao desmobilizar um espaço público abrangente, possibilita-se que grupos se unam em torno de interesses em comum, desde interesses inofensivos ou até mesmo cívicos, até apoiadores da Ku-Klux-Klan, oferecendo-se um espaço para que se encontrem, ainda que virtualmente (Empoli, 124-130).

Desta feita, na midiocracia, pautada nos meios de comunicação de massa tradicionais, o debate político era capturado pela performance pautada na indústria do entretenimento. Na infocracia, o próprio palco em que se realizavam os debates performáticos é pulverizado. O debate é fragmentado de tal maneira que, não somente não ocorre a construção de narrativas, como também a própria informação passa a ser ponto de disputa, gerando uma guerra de informação. As armas utilizadas nessa guerra são de variadas naturezas, como o uso de *trolls*, *bots* sociais, contas *fakes*, entre outros, os quais são utilizados para disseminar *fake news* e discurso de ódio, influenciando inconscientemente a opinião da população e o próprio debate público (Han, 2022, p. 40-46).

Nesse sistema, não há espaço para o discurso e argumentos, que são substituídos pelo uso mais eficiente dos algoritmos. A comunicação social fica fragilizada de tal maneira que materiais vazios e sem conteúdo, produzidos com intenção cômica, chamados *memes*, exercem papel fundamental na guerra de comunicação. Esses materiais têm como objetivo primordial provocar emoções, através da simplicidade condizente com a estrutura da comunicação digital, se propagando de maneira irrefletida e viral, dando o tom do debate público. Na infocracia, portanto, a verdade é incompatível com a velocidade da informação e comunicação (Han, 2022, p. 40-46).

Assim, ao invés de derrubar as barreiras de distância da representação na democracia e possibilitar uma espécie de “democracia em tempo real”, com *feedbacks* contínuos da comunidade para as instituições políticas, o que se percebe é que a comunicação por meio das mídias digitais não é livre ou democrática, mas, tendo como pano de fundo as comunicações efêmeras e privadas, é ditada por influenciadores digitais e algoritmos. A própria informação torna-se produto e a comunicação voltada apenas para o privado, em detrimento da organização de um espaço público, sem que temas sociais relevantes sejam pautados (Han, 2022, p. 47-49).

A sociedade da informação inaugura, desta feita, um processo em que a desvalorização da verdade não corresponde à valorização da mentira. De fato, dentro de um sistema lógico pautado na verdade, a mentira tem o condão de confirmar esta, ao ser confrontada com os fatos

e o conhecimento socialmente validado. O que ocorre na sociedade da informação, diferentemente, é um sistema que é pautado na desfactualização da realidade, ou seja, em uma afronta direta à própria facticidade e não à verdade. As *fake news*, nesse sentido, não são mentiras propriamente ditas, mas sim elementos que desconstituem a própria facticidade, não negando a realidade, mas criando uma nova realidade que se adequa. Por meio da comunicação digital, ocorre um processo em que a identificação da verdade é desvinculada dos fatos e passa a corresponder com a forma como o indivíduo se sente em relação àquela afirmativa, resultando na prevalência de uma espécie de “verdade sentida”, em inglês *truthiness* (Han, 2022, p. 83-88).

A perda de conexão com a realidade gera a existência de um espaço hiper-real, derivado da sociedade da informação, em que há um processo de desfactualização, ou seja, uma desvalorização da verdade factual que gera, por sua vez, a impossibilidade da estruturação de um “mundo comum”, necessário às sociedades democráticas. A verdade (ou a realidade) é, portanto, um lugar comum (construído socialmente e pela comunicação) que, fragilizado pela tribalização característica da comunicação digital, perde a sua capacidade vinculativa (Han, 2022, p. 81-83).

A perda da racionalidade discursiva e a prevalência da racionalidade digital, resulta em que as informações não são sequer confrontadas, mediante o abundante fluxo de conteúdo. Não há, assim, um processo de questionamento, de exposição aos contra-argumentos, característico do discurso, em que a participação do outro confere certa solidez à verdade. Dentro da racionalidade digital, a informação é sempre ambivalente, nunca posta à prova, de forma que a própria concepção de verdade fica prejudicada (Han, 2022, p. 93-97).

Desassociar a realidade fática da perspectiva de mundo e sociedade em determinados grupos, incentivando a descrença nas instituições, é uma estratégia utilizada por autocratas contemporâneos. Buscam, assim, que o discurso público se torne apenas um espaço de narrativas concorrentes, menos relacionado aos fatos, e mais às estratégias de comunicação da pós-verdade. A elevação dos discursos ao patamar de fatos alternativos, relativizando o valor da verdade e da mentira, colocando tudo em termos narrativos, atribuindo valor equânime às diversas narrativas, mesmo que apartadas da realidade, está intimamente relacionada com a crise de confiança nas instituições, sejam elas o Governo, a imprensa ou, mesmo, a ciência. A esse fenômeno dá-se o nome de crise epistêmica (Dourado, 2020, p. 97-100).

A epistemologia tribal passa a imperar nos meios de comunicação e informação digital, de forma que os indivíduos necessitam cada vez menos de mediadores e especialistas e passa a optar pelas informações que acessa e compartilha, de acordo com a conformidade com a

perspectiva pessoal ou de seu grupo. O processo, portanto, se constitui de duas etapas: a primeira destrutiva, desvinculante, que implica em gerar desconfiança das instituições sociais relacionadas à verdade, e a desvalorização do conhecimento socialmente validado; a segunda é construtiva, ou vinculante, criando uma rede de informação e comunicação exclusiva de determinado grupo, fora da qual todo material é entendido falso ou tendencioso, dificultando ou até mesmo impedindo o contraditório, a deliberação pública e a pluralidade democrática (Dourado, 2020, p. 100-102).

Ocorre, portanto, uma crise de narrativa em correlação direta com uma crise da identidade e sentido. As narrativas, entendidas enquanto resultado de um processo que demanda tempo e diálogo, são mitigadas em informações que carecem de sentido. Neste terreno, teorias da conspiração, *fake news* e discurso de ódio são amplamente disseminados uma vez que são elementos que carregam micronarrativas e conferem recursos de identidade e sentido, carentes na sociedade da informação. Predomina, assim, uma racionalidade digital desconectada de elementos duradouros da verdade, caracterizada pela fugacidade e vazio da informação (Han, 2022, p. 98-107).

O grande volume de conteúdo que é veiculado pelas mídias digitais, bem como a tendência de utilizar atalhos cognitivos e expor-se somente a conteúdos com os quais o usuário concorda, alimentado pela maior exposição de conteúdo com alta valência emocional por meio dos algoritmos, são características da comunicação digital que podem ser arregimentadas por líderes autoritários para semear a desconfiança nas instituições. Desde o escândalo das eleições estadunidenses de 2016, as plataformas digitais têm sido pressionadas a atuar contra a disseminação de conteúdos falsos em seus serviços, entretanto, em estudo de 2018, constatou-se que mais de 80% dos perfis que difundiram conteúdos falsos durante as eleições estadunidenses de 2016 mantinham-se ativos e publicando milhões de *posts* por dia dois anos depois, enquanto outro estudo aponta que cerca de 9% a 15% dos usuários do *Twitter* são robôs (Deibert, 2019, p. 8-10).

Há, assim, incentivos econômicos para que as plataformas sejam reticentes quanto a atuar contra tais perfis e conteúdo, tendo em vista o modelo de negócio das plataformas. Nesse sentido, quando, em 2018, o *Twitter* divulgou a exclusão de contas que compartilhavam notícias falsas, o preço de suas ações caiu severamente. As medidas tomadas pelas próprias empresas, portanto, tendem a não ser tão incisivas, tendo em vista que dependem do maior número de usuários e maior participação para gerar lucro. Pesquisadores conseguiram, por exemplo, no ano de 2018, durante período de intensa investigação por parte do Congresso dos EUA sobre o uso de *trolls* nas redes sociais naquele país, comprar anúncios políticos passando-se por *trolls*

russos, utilizando endereços e moeda russas, sem que o *Google* percebesse a manobra (Deibert, 2019, p. 10-12).

Em resumo, a economia que pauta o ecossistema de comunicação digital, baseada na vigilância dos dados dos usuários e concentrada na mão de algumas poucas empresas privadas, resulta na insuficiência de medidas tomadas voluntariamente pelas plataformas ou mesmo pequenas regulamentações por parte do Estado para enfrentar os graves prejuízos que podem ser causados à democracia. Será necessário ressignificar esse ecossistema comunicacional para proteger o complexo emaranhado de informação e comunicação proporcionado pelas plataformas digitais, sujeitando as plataformas a regulamentação que tenha o condão de torná-las responsáveis por seus atos, de maneira transparente, bem como reduzindo a participação relativa das grandes empresas de tecnologia por meio da aplicação de leis antitruste, tendo em vista reduzir o poder de algumas poucas empresas sobre as democracias (Deibert, 2019, p. 14-16).

Assim, quando as redes sociais surgiram e se popularizaram, a perspectiva sobre as possibilidades de ampliar as plataformas democráticas e empoderar a sociedade civil imperava. Nos últimos anos, entretanto, a manipulação dessas mídias com fins políticos, como no caso do Brexit ou das eleições estadunidenses de 2016, tem deixado cada vez mais evidente o uso das plataformas para atender interesses escusos, decorrendo na necessidade de regulação destas mídias. Assim, as estratégias de regulamentação devem partir de três pressupostos: 1) que as redes sociais lucram com a vigilância profunda e incansável dos dados pessoais para direcionar publicidade; 2) que voluntária e conscientemente aceitamos essa vigilância; e 3) que as redes sociais não são incompatíveis com o autoritarismo, sendo, na verdade, uma de suas ferramentas mais efetivas (Deibert, 2019, p. 1-3).

Desta forma, há uma correlação entre os tipos de sociedade e as tecnologias comunicacionais, ou meios de disseminação, que os sustentam. Assim, o surgimento da imprensa, ao possibilitar a crítica pública dos atos do poder estatal, desestabiliza a sociedade estratificada, para estabelecer a chamada opinião pública, enquanto um espelho anônimo que julga as ações estatais. A sociedade civil passa a ser considerada a arena de cidadãos reflexivos que se informam e disputam sentido pela imprensa. Essa opinião pública, por sua vez, reclama meios democráticos para influenciar as decisões públicas, resultando na criação de uma esfera pública pluralista de grupos e em uma disputa organizada, por meio de organizações partidárias, sindicais e outras, pautada sobretudo sobre os meios de comunicação de massa tradicional. As redes sociais, nesse sentido, por meio da fragmentação e indeterminação na opinião pública, desestabilizam a estrutura sociopolítica estabelecida ao longo dos séculos XIX e XX, ao

modificar a tecnologia comunicacional que sustenta o sistema. No sistema das redes sociais, os agentes que eram dotados de credibilidade, a imprensa, são desacreditados e passam a ser vistos como parciais e nenhum sujeito é capaz de garantir a veracidade dos conteúdos, uma vez até mesmo as agências de *fact-checking* são colocadas sob suspeição (Amato; Saba; Loschiavo Leme de Barros, 2022, p. 545-546).

O sistema jurídico, dentro da teoria dos sistemas sociais, é responsável pelo problema da licitude ou ilicitude, por meio da chamada comunicação jurídica- critério de validade-, substituindo a moral e a religião, que passam a ser identificadas enquanto escolhas individuais particulares na sociedade moderna. No centro do sistema jurídico estão os tribunais, que tomam decisões programadas, e na periferia estão as decisões programantes, constituídas por leis, normas, regulamentos e contratos. O que diferencia o centro do sistema jurídico de sua periferia é a proibição da denegação da justiça, sendo esse obrigado a decidir sobre o direito válido, portanto instituindo as expectativas normativas, mas a comunicação jurídica ocorre sempre que se estiver discutindo a licitude, validade ou constitucionalidade, ainda que na periferia do sistema (Amato; Saba; Loschiavo Leme de Barros, 2022, p. 546-547).

Os poderes políticos, assim, estão na periferia do sistema jurídico, mas constituem o centro do sistema político e são responsáveis pela tomada de decisão política que se aplica inclusive ao próprio Estado. Na periferia do sistema político estão organizações como partidos políticos, movimentos sociais e grupos de pressão, que organizam demandas sociais genéricas segmentadas, por meio de programas partidários, manifestos e declarações, os quais não se impõem perante toda a comunidade política, mas que se organizam sob o código governo/oposição para produzir decisões generalizadas por meio da comunicação política, denominada poder (Amato; Saba; Loschiavo Leme de Barros, 2022, p. 547). O sistema dos meios de comunicação de massa, nessa estrutura, é responsável pelo problema do que é ou não é informativo, que tem valor de informação, ou seja, se é notícia, de forma a cumprir a função de memória da sociedade, determinando o que deve ser lembrado e o que pode ser esquecido (Amato; Saba; Loschiavo Leme de Barros, 2022, p. 547).

As três esferas se relacionam entre si, de forma que os sistemas jurídico e político são impactados pelo sistema dos meios de comunicação. A substituição do sistema de mídia tradicional pelo sistema digital (que se iniciou com a mídia de massa, especialmente a televisão) gera implicações a questões de legitimidade política e de justiça. É nessa correlação entre os sistemas se localiza o problema das *fake news* nas democracias contemporâneas (Amato; Saba; Loschiavo Leme de Barros, 2022, p. 539-548), uma vez que através das novas tecnologias de comunicação alterou-se a configuração relacional entre as esferas.

Estabelecidas, para esta pesquisa, as premissas sobre a comunicação social, ou da interação entre tecnologia e sociedade, através da estrutura socioeconômica pautada no capitalismo e no informacionalismo, como uma espécie de pano de fundo que delimita e rege as relações sociais na sociedade contemporânea, podemos passar a discutir qual é o papel que a disseminação de conteúdos perniciosos à democracia, as chamadas *fake news* e discurso de ódio, exercem na estrutura das mídias digitais.

1.5. *Fake news* e discurso de ódio na infocracia: filtros-bolha e câmaras-de-eco

Na função de consolidar expectativas normativas por meio do processo decisório dos tribunais, as Cortes brasileiras não exerceram seu papel, no sentido de que inexistiu, até o momento, uniformização nas decisões sobre o tema das *fake news* no sistema jurídico brasileiro. Somente em relação aos casos julgados pelo Tribunal Superior Eleitoral sobre as eleições de 2018, por exemplo, além dos vários casos que não tiveram julgamento de mérito em razão da perda do objeto, nas decisões que mencionavam critérios para a identificação da regularidade do conteúdo divulgado foram mencionados diversos fundamentos divergentes:

“(i) a necessidade de a notícia ter caráter patentemente ou evidentemente inverídico ou mesmo de a falsidade ser verificável de plano; (ii) a necessidade de haver, na própria rede social, meio de contestação das ideias expostas, de modo a permitir ao próprio eleitor que julgue a veracidade daquilo que é divulgado; (iii) a consideração da capacidade do eleitor para avaliar os conteúdos que lhe são apresentados, sendo indevida a intervenção paternalista da Justiça Eleitoral; (iv) o potencial lesivo da divulgação, diferenciando-se posts realizados em páginas de pouco ou muito acesso; (v) postagens jocosas ou humorísticas (aqui incluídos os *memes*) têm como pressuposto a manipulação da realidade e até mesmo sua distorção, porém não podem ser consideradas irregulares; (vi) não é possível aferir se as notícias foram produzidas intencionalmente para desvirtuar ou falsear a realidade, ou se decorreram de equívocos ou de negligência dos veículos de mídias sociais; (vii) uma informação imprecisa mas não sabidamente inverídica não configura *fake news*; e (viii) o controle sobre o conteúdo ou nível das informações veiculadas deve ser realizado pela própria sociedade civil, não devendo a Justiça Eleitoral atrair para si a função de *fact-checking*.” (Amato; Saba; Loschiavo Leme de Barros, 2022, p. 552).

Vê-se, portanto, que uma das principais dificuldades sobre o tema é a definição de critérios para a identificação do que constitui *fake news* ou desinformação. Ademais, os Tribunais Superiores brasileiros produziram jurisprudência no sentido de, à luz da liberdade de expressão, somente conferir legitimidade a censura judicial *a posteriori*, ou seja, após decisão judicial sobre o caso específico, ainda que o conteúdo seja manifestamente inverdade, de forma a evitar a uniformização do tratamento do tema.

O termo *fake news* se popularizou durante as eleições presidenciais estadunidenses de 2016 em que disputaram o cargo Hillary Clinton e Donald Trump e surgiu para denominar os sites de notícias falsas criados pela campanha de Donald Trump para prejudicar a candidata adversária. Pela forma como termo se popularizou e seu uso indiscriminado, há na literatura autores que criticam seu uso, considerando-o impreciso, enquanto outros autores defendem o uso do termo. Em relatório oficial, a plataforma *Facebook* critica o uso do termo, afirmando que este é demasiadamente genérico, abrangendo conteúdos distintos que, em comum, contém fatos inverídicos, como artigos de opinião, paródias, sarcasmo, boatos, rumores, *memes* e afins. Por isso, preferem o termo *false news*, definindo-o por conteúdo que contém distorções factuais com o objetivo de enganar ou atrair audiência.

No mesmo sentido, a Organização das Nações Unidas para a Educação, Ciência e a Cultura (Unesco), indica que o uso do termo notícias falsas, ou, em inglês, *fake news*, é incorreto, uma vez que o termo “notícia” implicaria em informações verificáveis de interesse público, o que não é o caso da produção de materiais veiculados pelas mídias digitais cuja intenção é confundir ou manipular os usuários por meio de dados falsos. Assim, a Unesco indica o uso do termo desinformação (Queiroz, 2022).

A discussão quanto à definição do que constitui *fake news* e desinformação na *internet* é bastante recente e comporta diversas correntes divergentes. Uma primeira corrente defende que o termo *fake news* é demasiadamente vago e, por motivos diversos, prefere o uso de outros termos, especialmente desinformação. Nesse sentido, Margareth Sullivan (*apud* Dourado, 2020, p. 41-42) sustenta que o termo *fake news* é impreciso, apesar de identificá-lo com a produção de materiais em formato de noticiário com o objetivo de enganar. Ethan Zuckerman acrescenta que o uso do termo colabora com o discurso político anti-imprensa, ao dar enfoque às dificuldades na produção jornalística no mundo digital (Dourado, 2020, p. 41-42).

Já Alcott e Gentzkow (*apud* Ribeiro; Ortellado, 2018, p. 72), em estudo mais citado no meio acadêmico, definem *fake news* enquanto notícias que são falsas, produzidas de maneira intencional que pode enganar os leitores, incluindo conteúdos que não tenham sido criados com essa finalidade, mas que, fora de contexto, especialmente veiculados pelas mídias digitais, podem ser tomados enquanto verídicos. Excetuam, entretanto, erros de reportagem não intencionais; rumores que não sejam derivados de um conteúdo específico; teorias da conspiração; sátira; declarações falsas de políticos; conteúdos tendenciosos ou enganosos, mas não completamente falsos.

Lazer et al (*apud* Ribeiro; Ortellado, 2018, p. 73), por outro lado, vinculam o termo *fake news* àquele conteúdo forjado sem processo editorial, diferenciando dos conceitos de

misinformation, no sentido de um engano em relação a informação que se acredita verídica, mas que não é, do conceito de *disinformation*, que configuram enquanto conteúdo falso difundido com a intenção de enganar. A especificidade do conceito de *fake news*, para os autores, residiria sobre a intenção de fazer o conteúdo parecer notícia na forma como é apresentado, mas sem ter sido realizado trabalho jornalístico.

Os conceitos propostos pelo Facebook, Alcott e Gentzkow e Lazer *et al*, centralizam o debate na elaboração de conteúdo que pretende, especialmente pela forma, parecer verídico e ser resultado de um processo jornalístico, quando, de fato, não é. Conforme exposto neste tópico, entretanto, o contexto de disputa da realidade fática característico do fenômeno da pós-verdade, em que está inserida a produção e disseminação em massa de *fake news*, deve constituir um dos fundamentos da elaboração do conceito. Ou seja, é preciso questionar se a forma do conteúdo, que pode deliberadamente querer passar por material jornalístico, é elemento essencial, ou meramente acessório, do conceito de *fake news*, ao que a identificação de algumas das características da comunicação digital elaborada nos tópicos anteriores nos permitirá lançar luz.

O Relatório do Conselho da Europa também é refratário ao uso do termo *fake news*, explicitamente por dois motivos: primeiro, por acreditar que o termo é reducionista e não possibilita a adequada apreensão do fenômeno da desordem de informação *online* e, segundo, porque o termo tem sido utilizado por políticos para suprimir a liberdade de expressão. De fato, a ausência de boas definições no campo da comunicação digital é um desafio, não apenas no que diz respeito ao termo *fake news*, uma vez que palavras como *disinformation* e *misinformation* também assumem significados distintos de acordo com o contexto em que são utilizadas, levando a pressuposições quanto a que tipo de conteúdo se refere, quem são os sujeitos que o compartilha e quais medidas são desejáveis, apropriadas ou mesmo possíveis de serem tomadas em relação à disseminação (Wardle; Derakhshan, 2017, p. 5-18).

Nesse sentido, Clare Wardle (*apud* Dourado, 2020, p. 42-44) sustenta que a questão fundamental da produção de mentiras na *internet* é saber quais os tipos, motivações e formas pelas quais os conteúdos estão sendo disseminados, defendendo a necessidade de diferenciar os conceitos de *misinformation* (compartilhamento inadvertido de informação falsa), *disinformation* (criação e partilha deliberada de informação reconhecidamente falsa) e *propaganda*. A autora sustenta a existência de sete tipos de mentiras (dentre as quais nem todas são reprováveis, ou seja, devem ser objeto de regulamentação, bem como entre as reprováveis a resposta institucional deve divergir de acordo com as características específicas, conforme discutiremos adiante) disseminadas pela *internet*, quais sejam: sátira ou paródia, conteúdos

enganosos, conteúdos impostores, conteúdos fabricados, falsas conexões, falso contexto e conteúdo manipulado (Dourado, 2020, p. 42-44).

Wardle sustenta, ainda, que há oito motivos pelos quais conteúdo desta natureza está sendo criado e disseminado, sendo eles: jornalismo de baixa qualidade, paródia, provocação, paixão, partidarismo, para obter lucro e propaganda. Note-se a ausência do conceito de *fake news* no trabalho da autora, que prefere o uso do termo desordem informativa, enquanto conceito mais abrangente, subdividindo-a em desinformação, enquanto aquela criada deliberadamente para prejudicar pessoa, grupo social, organização ou país; informação errada, enquanto aquela que é falsa, mas não há intenção de enganar; e informação maliciosa, que é baseada na realidade, buscando prejudicar pessoa, organização ou país. Nesse sentido, desinformação é a intersecção entre conteúdo falso e prejudicial (Wardle; Derakhshan, 2017).

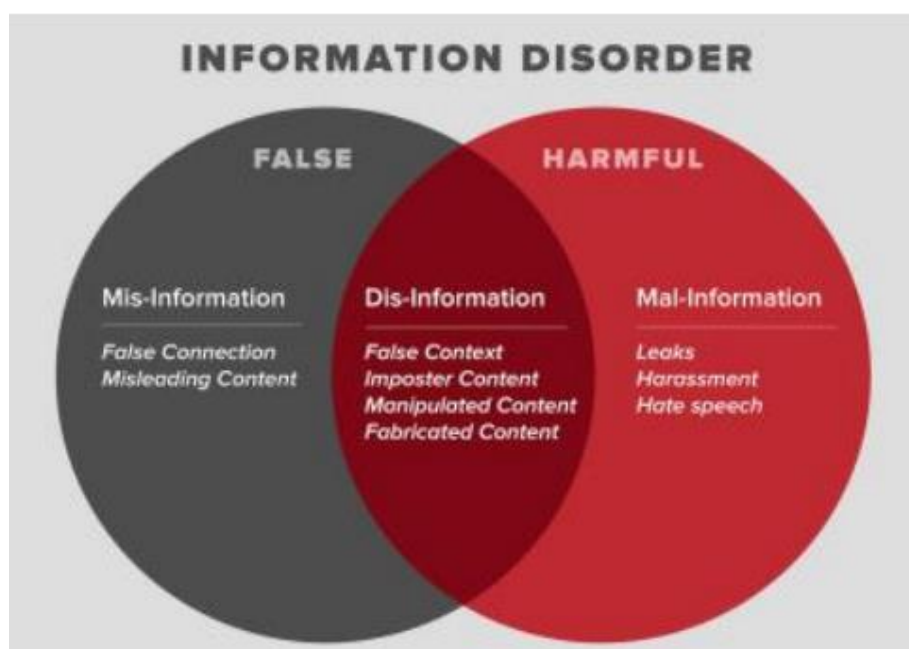
Ao utilizar o conceito de “desordem de informação” - *information disorder*- na comunicação digital, é possível analisar de maneira aprofundada os diferentes tipos de conteúdo veiculados digitalmente, bem como os desafios e panorama das mídias digitais e principais impactos derivados do fenômeno da desinformação. Um Relatório elaborado pelo Conselho da Europa salienta, nesse sentido, que, apesar de a mentira ser um fenômeno recorrente ao longo da história, esta ganha novos contornos na era da comunicação digital, diferenciando-se pela “poluição” na informação a nível global; por um complexo emaranhado de interesses em criar, disseminar e consumir esses conteúdos “poluídos”; pelas novas formas e técnicas de ampliar a disseminação destes conteúdos; pelas inúmeras plataformas que recebem conteúdos; e pela velocidade de comunicação entre conhecidos (Wardle; Derakhshan, 2017, p. 4).

O Relatório do Conselho da Europa destaca a necessidade de abordar o tema a partir de oito frentes: 1) elaborando definições, para compreender a complexidade do fenômeno; 2) estudando implicações para a democracia; 3) identificando o papel de mídias tradicionais, especialmente a televisão, nesse novo panorama comunicacional; 4) avaliando as implicações do enfraquecimento de mídias locais; 5) observando os efeitos do *micro-targeting*, ou seja, do uso de dados para encaminhar conteúdo personalizado de acordo com o perfil do usuário; 6) identificando o funcionamento e efeitos da amplificação computacional, ou seja, o uso de *bots*, ou robôs, para ampliar o alcance de conteúdo; 7) discutindo e observando a formação de filtros-bolha e câmaras de eco; e 8) e os impactos da perda de confiança nas evidências (Wardle; Derakhshan, 2017, p. 4).

Conforme exploraremos adiante, optamos por adotar um conceito de *fake news* que relacione as características da comunicação digital discutidas nos tópicos anteriores, com o fenômeno da pós-verdade e a instituição de mídias hiper-partidárias nas mídias digitais.

Todavia, reconhecemos que o conceito de *fake news* não pode encerrar em si próprio todo o fenômeno da disseminação de desinformação ou mesmo veiculação de conteúdos que podem gerar violência, teorias da conspiração ou discurso de ódio nas plataformas digitais. Assim, ainda que sustentemos a especificidade do termo *fake news*, resultado de uma tentativa de reduzir a um conceito que, ainda que não possibilite a objetivação prática de cada conteúdo veiculado pelas mídias digitais, delimita o que entendemos enquanto fenômeno que requer/possibilita resposta jurídico institucional, é imperioso reconhecer que qualquer tentativa de regulamentação do tema deve abranger outros aspectos sobre a produção e disseminação de conteúdo pelas tecnologias de comunicação e informação digitais.

De maneira mais abrangente, o Relatório do Conselho da Europa propõe que o **fenômeno da desordem de informação** seja analisado a partir de **três categorias**, quais sejam: 1) *mis-information*, enquanto a informação compartilhada sem a intenção de prejudicar; 2) *dis-information*, enquanto a informação compartilhada com a intenção de prejudicar; e 3) *mal-information*, enquanto a informação verdadeira que é compartilhada com intenção de prejudicar, geralmente publicizando informações privadas. Esses são os critérios utilizados pelo Relatório para identificar a natureza dos conteúdos do fenômeno da desordem de informação, representados graficamente pelas imagens a seguir:



(Wardle; Derakhshan, 2017, p. 5).

Quanto às categorias propostas no relatório, a palavra *misinformation* surgiu em campos diversos e com conotação diferente da atualmente dada no campo da produção de mentiras na

internet, surgindo para caracterizar informações irrelevantes nos gerenciadores de sistemas, no campo administrativo, e na psicologia para designar informações enganosas (Dourado, 2020, p. 48-49).

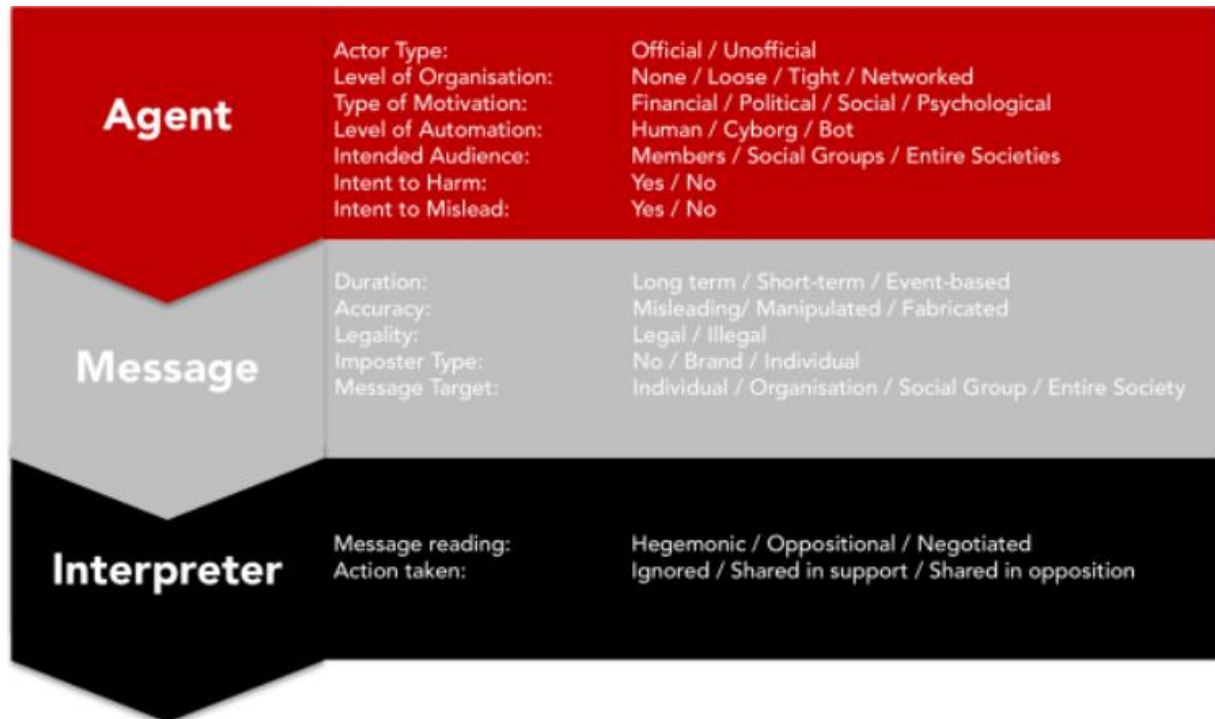
Já a palavra *disinformation* surge em relação ao uso de mecanismos para manipular a opinião pública, surgindo na Rússia durante a Guerra Fria, e diz respeito ao uso por governos para influenciar países estrangeiros, sendo adotada na língua inglesa em 1972 e na francesa em 1974. A adoção da palavra na língua francesa ressaltou o uso de técnicas de desinformação em massa para enganar e reconheceu o uso da palavra de maneira estrita, para designar a influência de opinião, ou em sentido lato, de maneira a designar qualquer mentira. A desinformação distingue-se da propaganda porque esta última, mesmo que pautada em inverdades, não se preocupa em omitir sua natureza. No conceito proposto por Volkoff, assim, desinformação envolve a manipulação da opinião pública, com fins políticos e por meio de processos ocultos (Dourado, 2020, p. 49-51).

Com o avanço da temática das *fake news*, entretanto, o uso das palavras *misinformation* e *disinformation* passou a ser ressignificado. *Misinformation* passou a se referir a disseminação de informações falsas sem a intenção do remetente de enganar, podendo englobar, por exemplo, erros de apuração no jornalismo. A despeito da ausência de intenção de enganar, a propagação da chamada *misinformation* pode ser prejudicial, a depender do caso, uma vez que podem ser rapidamente popularizados, tornando-se *trending topics*, memes ou *fake news* (Dourado, 2020, p. 51).

Disinformation, em contraposição, é o conteúdo falso deliberadamente produzido, com vistas à ampla divulgação, com finalidade de influenciar a opinião pública ou gerar ganhos financeiros. É, portanto, conceito mais amplo do que *fake news*, abarcando também estas, uma vez que pode se consubstanciar em outras espécies. A distinção entre os conceitos de *misinformation* e *disinformation*, portanto, reside na intenção no agente. Nesse sentido, no que diz respeito às *fake news*, por serem compreendidas enquanto o conteúdo produzido para parecer transmitir informações verdadeiras e conferidas, a intenção do produtor em falsear fica evidente, podendo ser até mesmo presumida, entretanto a intenção dos sujeitos que compartilham é mais ambígua (Dourado, 2020, p. 51-52).

Além de elaborar conceitos que permitam a delimitar a natureza dos conteúdos que causam a desordem da informação que ocorre nas mídias digitais, analisar os elementos e fases da produção deste tipo de conteúdo é indispensável para a elaboração de estratégias de resposta jurídico-institucional. Assim, o Relatório do Conselho da Europa propõe que a produção de conteúdo que produz desordem da informação é constituída de **três elementos**: 1) agentes; 2)

mensagem; 3) intérprete; e **três fases**: 1) criação; 2) produção; e 3) distribuição (Wardle; Derakhshan, 2017, p. 5-6).



(Wardle; Derakhshan, 2017, p. 6).



(Wardle; Derakhshan, 2017, p. 6).

Através destes critérios, é possível identificar especificidades do fenômeno da disseminação de conteúdo *online*, como, por exemplo, que o agente que produz um conteúdo pode ser diferente do agente que dissemina o conteúdo, inclusive quanto à motivação, que pode ser financeira para o primeiro e política para o segundo, por exemplo. Através deste recorte,

permite-se uma análise quanto a qual o nível de participação de cada agente na disseminação de conteúdo prejudicial e a melhor maneira de responsabilizá-los (Wardle; Derakhshan, 2017, p. 6-7).

Nesse sentido, tomando-se como exemplo um caso das eleições presidenciais estadunidenses de 2016, o site *WTOE 5* publicou uma notícia falsa que indicava que o Papa Francisco declarou apoio ao então candidato Donald Trump. Em uma análise das fases do processo de disseminação de uma notícia falsa, conforme proposto pelo Relatório, pode-se identificar: a criação da mensagem, que no caso foi concebida por um sujeito indeterminado; a produção da mensagem, que foi publicada pelo *WTOE 5 News*, que faz parte de uma rede de 43 sites de notícias fabricadas que publicaram mais de 750 artigos; a ampla distribuição nas plataformas digitais, especialmente o *Facebook*, por parte desta rede; e a reprodução do conteúdo por: pessoas conectadas à rede de sites de notícias fabricadas; apoiadores de Trump; atores que tinham interesse na vitória de Trump, como robôs ou *trolls* financiados pelo governo russo; apoiadores de Hillary Clinton que pretendiam mostrar a facilidade com que os apoiadores de Trump eram enganados; e por agentes da mídia que, em busca de apontar a falsidade do conteúdo, ampliaram o alcance do material (Wardle; Derakhshan, 2017, p. 23-25).

Quanto aos agentes que atuam nas três fases da disseminação de desinformação, o Relatório sugere sete critérios para categorizá-los: 1) quanto ao tipo: o agente pode ser oficial, como serviços de inteligência, partidos políticos ou organizações que produzem notícias; ou não oficial, como um grupo de cidadãos que trata de um tema; 2) quanto à organização, os agentes podem ser individuais; grupos de longa data; organizações bem estruturadas, como empresas ou grupos de *lobby*; ou grupos recém criados em torno de interesses em comum; 3) quanto à motivação, as principais categorias apontadas são: financeira, por meio do lucro com a produção do material; política, para influenciar a opinião pública; social, para estabelecer conexões com um determinado grupo, *online* ou *offline*; psicológica, para alcançar prestígio ou corroborar perspectivas pré-concebidas; 4) quanto à audiência projetada, ou seja, quanto ao grupo de usuários que se pretende atingir com o conteúdo, podendo ser uma lista de consumidores ou correspondentes, grupos sociais pautados em características socioeconômicas ou mesmo toda a sociedade; 5) quanto ao uso de tecnologia autônoma, ou seja, quanto ao uso de robôs para a disseminação do conteúdo; 6) quanto ao interesse em deliberadamente enganar a audiência ou não; 7) quanto ao interesse em deliberadamente gerar danos ou não (Wardle; Derakhshan, 2017, p. 25-26).

Quando o agente que atua na disseminação de desinformação é governamental, o nível de organização e os objetivos buscados, geralmente influenciar a opinião pública, para fins

estratégicos ou geopolíticos, adquirem especial relevância. Alguns exemplos de agentes governamentais podem ser encontrados na atuação do governo russo para influenciar a população estadunidense, especialmente durante as eleições presidenciais de 2016, e na votação do Brexit, que é caracterizada por produzir: conteúdo em larga escala e distribuído por diversos canais; disseminação rápida, contínua e repetitiva do conteúdo; não se pautar na realidade; inconsistente quanto à mensagem. O objetivo desta abordagem era fazer parecer que havia vários canais discutindo um determinado tema, com opiniões distintas, mas que promovessem no debate público posições favoráveis aos interesses russos. O mesmo se deu em outros países como a China, que pagou usuários para compartilhar conteúdos pró-governo, o Bahrein e o Azerbaijão, que criaram contas falsas para influenciar a opinião pública (Wardle; Derakhshan, 2017, p. 28-31).

A motivação dos agentes é, desta forma, um aspecto a ser considerado nas estratégias para combater a disseminação de desinformação. Na cadeia de produção de desinformação, há agentes cujo interesse é meramente financeiro, não tendo qualquer relevância para estes os impactos que o conteúdo pode desencadear. Assim, há sites ou páginas que produzem conteúdo prejudicial ou financiadores cujo objetivo é somente ou principalmente lucrar (Wardle; Derakhshan, 2017, p. 34-35). De fato, conforme exposto anteriormente, as próprias plataformas digitais lucram com a disseminação de conteúdos falsos. Mecanismos eficazes de governança digital devem levar em consideração não apenas a atividade dos sujeitos cujo interesse é derivado diretamente do impacto na disseminação do conteúdo, mas também dos agentes cujo interesse é indireto, nesse caso por aspectos financeiros.

Os agentes podem ser movidos por razões sociais ou psicológicas, desde a mera expectativa de conectar-se ou gerar entretenimento, até organizações em grupos extremistas que buscam corroborar uma visão de mundo, como comunidades *alt-right* que se utilizam da comunicação digital. No caso da formação de mídias hiper-partidárias, os conteúdos prejudiciais não são necessariamente falsos, mas podem ser mensagens de outra natureza que, pautada na polarização política e na identidade de classe, raça, gênero etc., buscando atuar sobre os sentimentos dos usuários (Wardle; Derakhshan, 2017, p. 35-36).

O uso de ferramentas de automatização se concentra na fase de distribuição de conteúdo com desinformação, especialmente no início, quando são utilizados robôs que simulam usuários reais, chamados *bots*, ou *cyborgs*, que são contas operadas por humanos com a ajuda de *software*, para disseminar o conteúdo para grupos de usuários com tendência a reproduzi-lo, mediante análise prévia do perfil dos usuários, bem como para influenciadores respeitados

dentro dessas redes de comunicação que poderão amplificar e dar significado à mensagem (Wardle; Derakhshan, 2017, p. 37-38).

Parte do processo de produção de desinformação é a mensagem, ou seja, a ideia que se pretende passar para o público-alvo de determinada campanha de desinformação. Há diversas formas pelas quais a mensagem pode ser disseminada, de forma que o conteúdo pode ser escrito ou audiovisual, como uma foto ou vídeo, o que impacta diretamente as tecnologias para a identificação automatizada de conteúdos com desinformação (Wardle; Derakhshan, 2017, p. 26-27).

Os critérios de classificação da mensagem são: 1) quanto à durabilidade, a mensagem pode ser elaborada para gerar efeitos imediatos e curtos, como em períodos eleitorais e conteúdo sobre eventos em tempo real, ou de longo prazo; 2) quanto à precisão, o conteúdo pode variar entre uma escala que vai desde falsas conexões até integralmente falsos. Ademais, neste ponto é importante destacar que, para um conteúdo ter potencial de gerar prejuízos, não é necessário que este seja falso, como no caso da disseminação de discurso de ódio pautado em diferenças religiosas, por exemplo; 3) quanto à legalidade, o material pode ser já reconhecidamente ilegal, a depender da jurisdição, como no caso de discurso de ódio, violações à propriedade intelectual, assédio ou violações à privacidade; 4) a mensagem pode ser produzida de maneira a fazer parecer ser oriunda de uma fonte oficial ou utilizando a imagem de uma pessoa conhecida ou instituição, de forma a serem classificadas como “conteúdo impostor”; 5) quanto ao alvo que a mensagem pretende influenciar, pode ser um indivíduo, como um político ou empresário, uma organização privada ou agência governamental, um grupo social ou toda a sociedade (Wardle; Derakhshan, 2017, p. 26-27).

Assim, correntes que optam por não utilizar o termo *fake news*, vinculam-se à perspectiva de que o termo “notícias falsas”, tem o condão de naturalizar que notícias podem ser falsas, confundindo a produção de conteúdo efetivamente jornalístico que podem conter erros pontuais, com a produção deliberada de conteúdos tendo em vista a desinformação. Entretanto, o termo também chama atenção para o fato de que as *fake news* são produzidas com a intenção de parecerem, efetivamente, notícias, e não conteúdos de ficção, unindo fatos verídicos com falsidades deliberadas (Dourado, 2020, p. 46-47). Nesse sentido, entendemos que o termo *fake news*, ao contrário, tem o mérito de reconhecer o objeto enquanto um problema social a ser enfrentado, apesar de seu uso indeterminado por parte de políticos e pela comunidade, e as particularidades do objeto justificam a manutenção deste enquanto categoria explicativa do fenômeno da mentira na esfera pública *online* (Dourado, 2020, p. 47-48).

Os conceitos de pós-verdade e *fake news* foram amplamente difundidos especialmente após dois fenômenos de larga expressão para as democracias contemporâneas, quais sejam o Brexit e a eleição de Donald Trump nos EUA. O termo pós-verdade foi classificado pelo Dicionário Oxford enquanto a maior utilização de emoções ou crenças pessoais, em detrimento dos fatos, para a formação da opinião pública, guardando relação direta com os meios de comunicação digitais e a desestabilização do sistema tradicional de informação e comunicação. Nesse sentido, para entender as implicações desses fenômenos é necessário partir de três premissas: de que a desinformação na política sempre existiu; que as tecnologias de informação e comunicação modificaram a forma como esse conteúdo é difundido; e, por fim, que não é possível imputar todo o contexto de crise democrática ao fenômeno da desinformação (Alves; Maciel, 2020, p. 147-148).

A difusão de informações falsas sob qualquer pretexto não é novidade, destacando-se especialmente, para os fins dessa pesquisa, a utilização de desinformação em massa em função de interesses políticos, como períodos de guerra. Dois exemplos do uso de desinformação para fins políticos são a criação de falsas rádios alemãs por parte dos ingleses para disseminar comentários contra Adolf Hitler ou a adulteração de fotos por parte da URSS para excluir membros que perdiam a simpatia do regime. Hannah Arendt destaca, nesse sentido, que há na política um campo de tensão entre a vontade de produzir o mundo como desejamos e entender a realidade como ela é, em um compromisso com a verdade. O revisionismo histórico é um exemplo da força que a ação política pode produzir para dar novos contornos à realidade fática, de forma que, ao reinterpretar fatos do passado com balizas não necessariamente pautadas sobre a realidade, pode-se ressignificar irremediavelmente a própria percepção social daquilo é ou não é verdade (Alves; Maciel, 2020, p. 148-149).

É importante destacar que, de fato, a produção de informações falsas para influenciar decisões públicas não é novidade, entretanto, diante das plataformas digitais, a desinformação adquire aspectos distintos, não apenas pela estrutura com que é disseminada, mas pela forma que se trata de verdadeira busca pela substituição da realidade, para além da mera manipulação dos fatos, no contexto da pós-verdade, inserida em uma infocracia.

Desta forma, não há consenso teórico sobre os elementos que constituem o que é chamado de desinformação, *fake news* ou notícias falsas, de forma que um esforço para a caracterização dos conceitos é necessário para compreensão do objeto a ser debatido e regulamentado. É importante destacar que esses conteúdos produzem, ao mesmo tempo, desinformação e induzem determinados comportamentos na sociedade. Assim, se delimita um ponto de reflexão, qual seja a produção deliberada de conteúdos com desinformação em busca

de influenciar determinados comportamentos. Aqueles que patrocinam a produção de tais conteúdos têm como objetivo, frequentemente, confundir a opinião pública para realizar interesses privados e assegurar a manutenção do poder (Amato; Saba; Loschiavo Leme de Barros, 2022, p. 542-544), mas a estrutura de produção e disseminação destes conteúdos tem a participação de diversos outros agentes, tanto aqueles que acreditam de fato na mensagem que está sendo difundida, quanto de publicitários ou agentes de *marketing*.

Uma das principais dificuldades derivadas do meio de comunicação digital é a indefinição dos sujeitos envolvidos na produção e divulgação de conteúdos, de forma que todos na “esfera pública política digitalizada” são potenciais mediadores, consumidores e produtores de conteúdo em massa, impossibilitando o uso dos controles de disseminação de desinformação aplicados aos meios de comunicação de massa tradicionais. Ao mesmo tempo, há uma indefinição no sistema brasileiro sobre o dever de veracidade e sua relação com a liberdade de expressão, decorrente da tradição liberal. Desta forma, ambas as indefinições demarcam o debate e ação institucional no combate à divulgação de desinformação nas plataformas digitais (Amato; Saba; Loschiavo Leme de Barros, 2022, p. 539-542).

A introdução das novas tecnologias de informação e comunicação modificou o panorama sobre o qual atuam as diversas formas de desinformação, especialmente por meio da criação e rápida expansão da rede mundial de computadores, a *Internet*, que conecta mais de 60% da população mundial, bem como pela popularização de *smartphones* com acesso imediato e constante à *internet*. A era da informação vivida hodiernamente é marcada por ferramentas de comunicação que, em contraposição à mídia tradicional, caracterizada por ser unidirecional, centralizada e verticalizada, são pautadas no modelo “todos para todos”, de forma que qualquer usuário pode produzir e compartilhar conteúdo, gerando uma espécie de ansiedade coletiva pautada na sobrecarga de informação, uma vez que não há como processar todo o conteúdo disponível (Alves; Maciel, 2020, p. 149-150).

Um aspecto da arquitetura das mídias sociais que prejudica a avaliação crítica das informações é que conteúdos oriundos de sites de notícias confiáveis são apresentados com o mesmo formato que sites que produzem conteúdos falsos. O foco dos *posts*, assim, é o conteúdo e não a fonte, o que privilegia o acesso a conteúdo por meio de compartilhamento dos demais usuários em detrimento da busca por fontes confiáveis. Nesse sentido, iniciativas de *source-checking*, assim como as amplamente disseminadas organizações que realizam *fact-checking*, bem como modificações na arquitetura das plataformas digitais, especialmente das redes sociais, que coloquem em evidência o perfil que está compartilhando o conteúdo, são medidas

importantes no combate à disseminação de informações falsas (Wardle; Derakhshan, 2017, p. 14-19).

Nas eleições presidenciais estadunidenses de 2016 é possível encontrar um dado que permite exemplificar o alcance das informações falsas disseminadas nas mídias digitais: nos últimos três meses da campanha eleitoral, os 20 principais conteúdos veiculados por sites que produziam conteúdos falsos e hiper-partidários tiveram pouco mais de 8,7 milhões de compartilhamentos, reações e comentários no *Facebook*, enquanto os 20 principais conteúdos originados de sites de notícias tiveram apenas 7,3 milhões de interações. No que diz respeito ao meio pelo qual as notícias eram compartilhadas nas plataformas digitais (redes sociais, pesquisa direta, outros *links* ou mecanismos de pesquisa), 41,8% dos conteúdos falsos foram distribuídos por meio das redes sociais, enquanto apenas 10,1% das notícias reais foram veiculadas por este meio (Wardle; Derakhshan, 2017, p. 11).

As redes sociais constituem importantes mídias no Brasil (e no mundo), fontes de comunicação e informação, sendo acessadas diariamente por milhões de cidadãos brasileiros e considerada fonte primária ou única de informações por parcela significativa dos usuários. A estrutura das plataformas digitais e da comunicação social por meio destas resulta na formação de grupos de usuários em torno de determinadas páginas e conteúdo, resultando na polarização do debate público.

A comunicação e busca de informação por meio de plataformas digitais assume características específicas pela capacidade do usuário de escolher os conteúdos e demais usuários que tem interesse em acompanhar. Inicialmente essa escolha era feita pelo usuário por meio de participação em fóruns, comunidades ou conversas em *chats* sobre assuntos de seu interesse, passando posteriormente a também ser feita por meio de algoritmos. O uso desses filtros personalizados para os usuários criou o que Sustain chama de câmaras de eco, modificando as interações interpessoais e editoração noticiosa, implicando em menor propensão à diversidade e pluralidade (Dourado, 2020, p. 60-62).

Algoritmos são procedimentos que, baseado em cálculos específicos, transformam dados em resultados desejados (Gillespie, 2018, p. 97). Eles cumprem o papel não apenas de ajudar o usuário a encontrar informações, mas meios para descobrir outros materiais, de forma que podem ser projetados para influenciar a opinião pública em determinado sentido, estruturando as informações a que o usuário terá acesso, ou seja, exercendo uma espécie de curadoria. Plataformas como *Facebook*, *Twitter* e *WhatsApp* utilizam algoritmos que levam em consideração dados como quantidade de interações do usuário, formato de conteúdo que o usuário mais interage, popularidade de determinada publicação, entre outros, para determinar

quais conteúdos serão apresentados preferencialmente em suas *timelines*. Desde 2018 o Facebook permitiu que o usuário editasse os conteúdos que gostaria de ver em sua *timeline*, editando preferências, ocultando contatos etc. portanto, o processo de personalização dos *feeds* de notícias das plataformas digitais são, em sua maioria, um híbrido entre a ação automatizada de algoritmos- que, reforce-se, é pautado nos conteúdos que o usuário acessa na plataforma- e as escolhas feitas deliberadamente pelo usuário (Dourado, 2020, p. 62-66).

As câmaras de eco, também chamadas de filtros-bolha ou bolha ideológica, constituídas por conteúdo personalizado a que os usuários têm contato nas plataformas digitais, tem sido consideradas peças fundamentais da propagação de *fake news*. Filtros-bolha, na concepção de Eli Pariser (*apud* Dourado, 2020, p. 66-68), é a segmentação de públicos mediante interesses, dentro dos quais circulam informações pautadas em crenças pré-existentes do grupo. Dentro das câmaras de eco circulam discursos limitados a uma determinada rede, confirmando o conhecimento dos próprios membros e o fornecimento das informações é feito por pessoas que já pensam de forma semelhante aos demais. São, portanto, ambientes propensos à propagação de *fake news*, ainda mais potencializados mediante o acirramento de ânimos (Dourado, 2020, p. 66-68).

Uma pesquisa de 2019 chamada *Trust in the Media*, do Instituto IPSOS mostra que a confiança nas mídias tradicionais, rádio, revista, jornais e televisão, diminuiu nos últimos cinco anos, pautado sobre a disseminação de *fake news* e as dúvidas quanto às boas intenções da mídia. Segundo a pesquisa, apenas uma fonte de notícias teve aumento na confiabilidade no período observado, que é a comunicação interpessoal, ou seja, conteúdos a que o indivíduo teve acesso por meio de pessoas próximas. A pesquisa identificou que, no Brasil, 65% confiam nas mídias tradicionais e 58% em sites de notícias e que 68% acreditam que *fake news* são distribuídas em mídias digitais, frente a 59% que acreditam que estas são oriundas de mídias tradicionais. Pesquisa do Instituto Reuters após as eleições de 2018 mostrou resultado ainda mais alarmante, identificando um aumento de 11% para 48% na desconfiança das notícias desde o ano anterior. A pesquisa *Digital Report* da Reuters em 2019 mostrou que o alinhamento dos usuários nas plataformas digitais está relacionado aos posicionamentos políticos, de forma que os usuários afirmaram acessar frequentemente sites de mídia alternativa vinculados a correntes políticas, como O Antagonista e a Rede Brasil Atual. Ou seja, há um declínio do jornalismo e das mídias generalistas, compondo o cenário de fragmentação nas plataformas *online* e busca por informações enviesadas, tendendo à maior permissividade dos usuários às *fake news* (Dourado, 2020, p. 74-76).

Ribeiro e Ortellado, ao analisarem os dados sobre as interações com as 500 principais páginas políticas brasileiras no *Facebook* em 2014, que concentraram as curtidas de 12 milhões de brasileiros, identificaram a criação de dois *clusters*, ou grupos de usuários, um aliado ao antipetismo e o outro ao anti-antipetismo. O primeiro concentra páginas de políticos e partidos de direita, liberalismo econômico e conservadorismo moral e o segundo, políticos e partidos de esquerda, feminismo, movimento negro, movimento LGBTQIA+ e ONGs de direitos humanos (Ribeiro; Ortellado, 2018, p. 73-74). Esses grupos estão representados graficamente na seguinte figura, de acordo com a quantidade de interações:

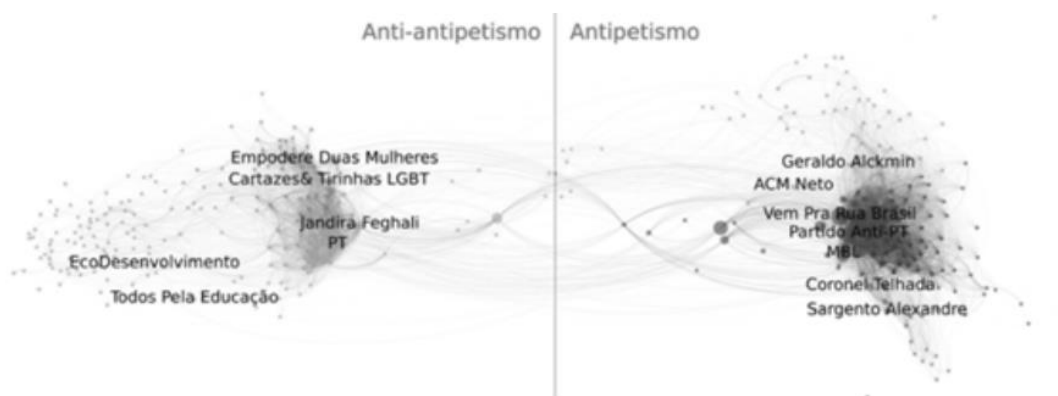


Figura (Ribeiro; Ortellado, 2018, p. 74).

O estudo de caso de Ribeiro e Ortellado demonstra o papel das plataformas digitais enquanto mídia, pautando o debate público, e a forma como as narrativas foram arregimentadas pelos dois polos de produção e disseminação de conteúdo.

Os autores identificaram que a narrativa do antipetismo gira em torno da concepção de que os maiores problemas do país seriam derivados da corrupção dos governos do Partido dos Trabalhadores, apontando-a como principal causa das crises econômica e ética e do déficit fiscal. Entendem, assim, que a Lava Jato é um mecanismo eficaz de combate à corrupção que favorecia políticos, grandes empresas beneficiárias de recursos de bancos públicos, de verbas de publicidade públicas e da Lei Rouanet e afins. Nesse sentido, entendem que a Lava Jato enfrenta a resistência de beneficiários da corrupção, nomeadamente o Congresso Nacional, meios de comunicação e o empresariado (Ribeiro; Ortellado, 2018, p. 75).

Já a narrativa do anti-antipetismo partia da premissa de que existe um movimento reacionário contra conquistas populares oriundas dos governos do PT. Tal movimento se revestiria de uma narrativa anticorrupção para impor contrarreformas, sendo o *impeachment* de Dilma Rousseff, a Lava Jato e a prisão de Lula meios ilegítimos para deslegitimar e destituir o Partido dos Trabalhadores, para promover uma agenda regressiva (Ribeiro; Ortellado, 2018, p. 75).

Ribeiro e Ortellado analisaram 150 manchetes sobre a política nacional durante o período de 21 a 27 de janeiro de 2018, semana da decisão em segunda instância sobre a prisão de Luiz Inácio Lula da Silva no caso do apartamento no Guarujá, que foram responsáveis por 3,5 milhões de compartilhamentos no período, o que corresponde a cerca de um terço de todo o conteúdo compartilhado pelas 500 páginas do *Facebook* observadas pelos autores. Dessas 150 manchetes, 67 estavam alinhadas à narrativa de golpe do anti-antipetismo, tendo sido compartilhadas quase 1,5 milhões de vezes, enquanto 45 estavam alinhadas à narrativa do antipetismo, compartilhadas 1,3 milhões de vezes, ou seja, foram responsáveis por quase 80% dos compartilhamentos da amostra analisada (Ribeiro; Ortellado, 2018, p. 75-77). Parcela dos conteúdos amplamente disseminados nesse período não continham necessariamente mentiras propriamente ditas, mas continham diversos níveis de distorção dos fatos tendo como objetivo influenciar o debate público (Ribeiro; Ortellado, 2018, p. 77).

Quanto ao contexto político favorável à tendência ao extremismo, especialmente caracterizado pela direita nos últimos anos, e o antipetismo, é necessário reafirmá-los enquanto instrumento que impulsiona o uso das plataformas digitais, para além das questões tecnológicas, ao uso indevido e à propagação de *fake news*. Polarização deve ser entendido enquanto um movimento que, diante do crescimento de sentimentos negativos em relação a adversários políticos, gera desconfiança social nas instituições democráticas, criando ambiente favorável à circulação de *fake news*. Trata-se, portanto, de um ciclo vicioso, em que a polarização implica em maior aceitação a mensagens com conteúdo falso e antidemocráticos os quais, por sua vez, implicam em mais polarização (Dourado, 2020, p. 95-97).

O modelo de negócio das mídias digitais, portanto, favorece a propagação da mensagem de líderes extremistas, enquanto dificulta que mensagens mais moderadas sejam veiculadas. Esse efeito midiático redireciona a atuação da política partidária, que busca vencer eleições e que, se antes precisava produzir discursos coerentes, hoje pode prescindir de raciocínio lógico, produzindo as mensagens de acordo com os dados dos usuários. Ao instigar mensagens mais extremas em grupos mais extremistas, e veiculá-las de maneiras semelhantes, porém mais suaves em grupos limítrofes, esse processo gera cada vez mais extremismo e polarização, na esteira de líderes populistas que defendem ideais intolerantes e preconceituosas. Por mais que tais líderes sejam personagens circunstanciais, as características desse sistema comunicacional pautado nas mídias de massa seguirão pressionando a política para os extremos, dificultando esforços pela deliberação política, de maneira que seus efeitos, ao contrário, podem ser muito mais duradouros (Empoli, 2019, p. 130-136).

1.6. A mídia-hiperpartidária desvelada pelo discurso de ódio

A construção de uma sociedade intolerante, dentro das democracias, pode implicar, no limite, na ruptura do Estado Democrático de Direito, implicando em danos tanto à primeira quanto à segunda característica do Estado: a adoção de autocracias e a negação a direitos políticos e humanos, especialmente das minorias. A expansão de fenômenos que levam ao extremismo tem diversas frentes e explicações. Se as mídias digitais ocupam espaço importante nesse processo na sociedade contemporânea, dois fatores com papel fundamental neste processo são a produção de desinformação, especialmente do que designamos *fake news*, e da propagação de discurso de ódio. O uso de discursos de ódio enquanto instrumento de política necessita de adequada compreensão, uma vez que se trata de processo específico, em muito relacionado à utilização das plataformas digitais, distinto do mero comportamento incivilizado.

Sponholz (2021, p. 221) traz o seguinte conceito de discurso de ódio: “é a comunicação pública que degrada simbolicamente grupos historicamente oprimidos ou sistematicamente discriminados”. O termo discurso de ódio, no original em inglês *hate speech*, foi criado por Mari Matsuda, jurista estadunidense, no contexto do racismo e de ações violentas como da Ku Klux Klan. Trata-se, portanto, de atuação consciente e/ou intencional de gerar iniquidade entre as pessoas em função de alguma categoria seletiva, podendo ser origem, cor da pele, gênero, religião, orientação sexual etc. Pode ser arregimentada com objetivos diversos como xingar; defender ideologias de desigualdade, tal qual o racismo ou sexismo; expressar crenças religiosas; ganhar visibilidade midiática; mobilizar online etc. (Sponholz, 2021, p. 221).

São três, portanto, os componentes do conceito de discurso de ódio, bem delimitados pela tabela abaixo, elaborada por Sponholz (2021, p. 223). i) O primeiro é que deve referir-se a grupo, ou seja, a uma coletividade. Ademais, este grupo deve ser objetivo de opressão crônica, fazendo parte de um processo de subordinação mais extenso; ii) o segundo é que a comunicação utilizada deve ser degradante, para além do estereótipo, incitando ameaça, discriminação, ódio ou violência; iii) por fim, realiza-se no espaço público, seja este o trabalho, a rua ou as mídias (Sponholz, 2021, p. 222-223).

Propriedades definidoras dos discursos de ódio

QUEM	O QUÊ		ONDE
Grupo	Degradação simbólica (comunicação)		Publicidade
Característica coletiva que corresponde a uma posição desprivilegiada de poder na hierarquia social	Disseminação de idéias discriminatórias		Espaços públicos (em oposição à esfera privada)
	Incitação de	▪ Ódio ▪ Desprezo ▪ discriminação ▪ violência	
	Incitação por meio de	negação pública de genocídios e crimes contra a humanidade	
	Ameaça		
	Justificação de	genocídios e crimes contra a humanidade	
	Expressão de	▪ xingamento ▪ escárnio ▪ insulto	

Fonte: Ilustração própria baseada em Committee on the Elimination of Racial Discrimination, 2013; Delgado und Stefancic, 2004; Matsuda, 1989, entre outros.

A emoção, portanto, não constitui elemento do discurso de ódio, o que implica em que este pode ser utilizado de forma racional, entendendo-se esta enquanto o oposto de emocional e, também, enquanto meio de justificação, ou seja, enquanto estratégia comunicacional. Nesse sentido, o discurso de ódio se caracteriza por manifestar-se enquanto agenda, no sentido do conjunto de pautas, apresentadas enquanto problemas que requerem ação. As pautas são elementos que se constituem por serem questões sobre as quais as pessoas se preocupam; problemas-chave a serem enfrentados pelo governo; existência de políticas diferentes para enfrentar o mesmo problema; e controvérsias públicas (Sponholz, 2021, p. 223-224).

No discurso de ódio, diferentemente da incitação à violência política ou apologia à ditadura, o elemento que caracteriza a iniquidade entre as partes é uma antinomia, ou seja, uma definição, e não o antagonismo, uma posição. Assim, as pessoas são atingidas pelo que são e não pelo que pensam. Por exemplo, as mulheres são alvos de discurso misógino, independentemente de qualquer característica que não seja o próprio fato de serem mulheres. Diferentemente da violência política, com origem em arranjos específicos (conjuntura), o discurso de ódio se origina de uma discriminação sistemática (estrutura). O problema, dentro do discurso de ódio, é caracterizado enquanto um grupo de pessoas, por características intrínsecas. A natureza das antinomias que constituem o cerne do discurso de ódio é social, não é criada pelo próprio discurso, mas está latente no tecido social, sendo apenas “ativada” pelos

atores sociais que utilizam do discurso de ódio (mídias, celebridades, políticos, partidos, organizações etc.). (Sponholz, 2021, p. 224-226).

Os reflexos do discurso de ódio são tanto de ordem individual quanto social, para o exercício da cidadania, constringendo o exercício da vida pública, provocando desconfiança e hostilidade entre grupos sociais, inibindo a participação em processos deliberativos, podendo ainda incitar, legitimar ou coordenar a violência física. Desta feita, as implicações do discurso de ódio são ainda mais profundas que outros fenômenos, como a polarização, uma vez que o discurso de ódio resulta em completa desintegração social e na impossibilidade de integração (Sponholz, 2021, p. 225-226).

A mídia, ao voltar sua atenção para determinados temas e convocar atores sociais para debatê-los, torna-os objeto de reflexão do público, passando a constituir-se enquanto pautas, no sentido exposto acima, gerando opiniões pró e contra. A atenção da mídia para determinado assunto, por sua vez, gera o que é chamado de capital midiático (com potencial de ser convertido em capital político), conferindo-se visibilidade, influência e prestígio aos atores sociais atuantes diante determinada pauta. Assim, veicular discursos de ódio e submetê-los a um processo de argumentação na mídia, ainda que de maneira jocosa, implica em naturalização daquela pauta, de forma a equipara-la a uma mera opinião (Sponholz, 2021, p. 227-228).

A mensagem transmitida pelo emissor da mídia de massa, entretanto, não penetra de maneira incólume nos espectadores, apesar de se tratar de uma comunicação em sentido único, vertical. De fato, aspectos de interpretação individual e coletiva e a ação política influenciam na maneira que as mensagens da mídia de massa são apreendidas pelo público. Assim, o poder da comunicação de massa reside na capacidade de pautar os temas e a forma como estes serão tratados na sociedade, modelando a linguagem da comunicação. Ou seja, para que um assunto ganhe relevância na sociedade informada pelos meios de comunicação, este deve estar presente na mídia, pressuposto sem o qual passa despercebido, entretanto, a maneira como a mensagem transmitida será recebida pode sofrer modificações, transformações ou mesmo subversões, a depender de outros aspectos culturais e sociais (Castells, 2011, p. 419-422).

O contexto político contemporâneo é de mitigação de ideologias, compreendidas enquanto uma visão de mundo sistemática e abrangente, em pautas. Ou seja, a afinidade ideológica e a identificação partidária têm sido gradualmente substituídas por pautas específicas enquanto elementos do debate público. Assim, as pautas arregimentadas por determinado ator social podem ser diversas, de forma a acomodar grupos distintos, de acordo com a importância que estes conferem a determinado problema, escanteando as posições do ator social em outras pautas. Desta forma, determinado ator político pode angariar capital político por sua pauta sobre

corrupção, por exemplo, ainda que seja adepto a uma pauta homofóbica. O ator social passa, então, a salientar pautas com maior aceitabilidade, como a corrupção, violência, criminalidade e ameaças à propriedade privada, podendo passar uma imagem suprapartidária do “cidadão de bem” (Sponholz, 2021, p. 228-230).

A propagação de discursos de ódio *online* deve levar em consideração a forma como as plataformas digitais funcionam. Quando os usuários interagem com conteúdo postado nas plataformas digitais, as empresas utilizam os dados gerados pela interação para estabelecer outros interesses e conteúdo a serem recomendados a este usuário, por meio de algoritmos. As plataformas digitais, portanto, não são meros expositores ou estoques de conteúdos postados por terceiros, mas devem ser identificadas enquanto um meio midiático propriamente dito. A popularidade de um tema nas plataformas digitais não é mero resultado da preferência do público, mas também resultado de fatores sociais e técnicos que envolvem o seu funcionamento (Sponholz, 2021, p. 230-232).

Nesse panorama, teorias conspiracionistas como campanhas antivacinas, a antipolítica e o incentivo à violência veiculam pelas mídias digitais, sem que seja do interesse das plataformas em si, mas em função da arquitetura que oferecem, uma vez que os algoritmos são desenvolvidos para manter os usuários conectados e engajados, independentemente do material que esteja sendo disseminado. Esse fato é relevante porque, uma vez que o que interessa não é exatamente a mensagem que está sendo transmitida, mas sim o alcance e viralização do conteúdo, materiais com intolerância, polêmicos, que instigam o medo ou a raiva, o insulto, o racismo e afins, tem maior capacidade de gerar reações (medida de sucesso de um conteúdo veiculado nas mídias digitais), seja pró ou contra, de forma que incentiva que atores que produzem conteúdo sobre estes temas tenham prevalência nas mídias digitais. Os engenheiros do caos atuam, portanto, através da raiva canalizada dos eleitores-consumidores, independentemente da razão, mas sim vinculado à emoção, conjuntamente com as máquinas de engajamento das mídias digitais, por meio de técnicas psicométricas, para influenciar o debate público e alcançar resultados tanto políticos quanto econômicos. Por meio desse processo, as mentiras e as conspirações passam a ser parte do suposto debate político, normalizando-os e, para alguns grupos, tornando-se “chaves de interpretação da realidade”, de forma a trazer para a arena política a vulgaridade, os insultos e os preconceitos, travestida de combate às elites em nome da liberdade (Empoli, 2019, p. 62-73).

É fundamental observar as plataformas digitais não apenas como um meio de tornar possível a publicação de conteúdos por qualquer indivíduo, mas especialmente pela criação de redes em torno de ideias. Quando essas redes são utilizadas para promover discursos

intolerantes e conteúdos falsos, conferem visibilidade e influência a grupos que, pela mídia tradicional, estariam escanteados (Dourado, 2020, p. 95-97). A construção de redes de usuários nas plataformas digitais por meio do interesse comum em determinado conteúdo é um dos elementos que constituem o fenômeno da disseminação de discurso de ódio e desinformação *online*.

O conjunto de interações dos usuários, os conteúdos veiculados e elementos técnicos como algoritmos de hierarquização constituem o que é chamado de objetos digitais. Os objetos digitais das plataformas de mídia social têm três dimensões: o objeto midiático, ou conteúdo, que pode variar em forma, como Gifs e memes; o objeto de conexão, resultado da interação na plataforma, que coaduna a ação humana com uma estrutura técnica oferecida pela intermediária, como “fazer amigos” *online* ou a mecânica de hashtags, priorizando determinados assuntos de acordo com a quantidade de curtidas ou compartilhamentos, por exemplo, criando a possibilidade de viralização de determinado conteúdo; e o objeto fático, que é constituído pelas relações construídas pelas interações do objeto de conexão, fornecendo informações sobre os atores e a comunicação do grupo nas redes sociais, de forma a possibilitar o fortalecimento dessas redes. É por meio destes objetos digitais que se estabelecem relações entre atores e conteúdos nas plataformas digitais, qualquer que seja seu tema. Quando utilizados para incitar, ameaçar ou humilhar grupos sistematicamente discriminados, verifica-se a propagação de discurso de ódio (Sponholz, 2021, p. 232-233). O uso deliberado dos objetos digitais para propagar discurso de ódio e *fake news* é uma das maiores ameaças às democracias contemporâneas.

O viés de confirmação e a teoria da dissonância cognitiva ajudam a compreender a preferência dos usuários das plataformas digitais por participar de grupos e acessar conteúdos em as opiniões são convergentes, ou seja, a exposição seletiva, uma vez que os indivíduos se sentem desconfortáveis em ambientes em que seu conhecimento, opinião e crenças não se adequam, de forma a buscarem ambientes em que se encaixam. As plataformas digitais, por meio dos algoritmos e das escolhas dos usuários, favorecem a segmentação dos debates produzidos em seu bojo. Quanto mais ao extremo as opiniões do indivíduo, portanto mais rígido e menos aberto a opiniões contrárias, maior a tendência a expor-se apenas a conteúdos semelhantes, gerando a tendência a tornar-se ainda mais extremo, dificultando esforços em busca da deliberação (Dourado, 2020, p. 68-71).

Algumas características da comunicação por meio das plataformas digitais, especialmente na forma como a informação é produzida, comunicada e distribuída, ajudam a compreender os dados: o surgimento e popularização de tecnologias acessíveis, baratas e

sofisticadas para a criação e distribuição de conteúdo; o consumo de informações se tornou público, uma vez que é feito nas mídias sociais, e não mais privado, ou seja, os usuários curtem e compartilham informações que acham interessantes, influenciando a forma como sua rede de amigos acessa conteúdos e vice-versa; a velocidade com que as informações são disseminadas foi acelerada pelo novo ciclo das notícias e o acesso a dispositivos móveis; e a informação é compartilhada por conhecidos em tempo real, diminuindo a propensão a serem analisadas criticamente (Wardle; Derakhshan, 2017, p. 11-13).

No mesmo sentido, algumas das principais características da comunicação por meio das plataformas digitais que a distingue das mídias tradicionais são: a facilidade de se criar um perfil independente e divulgar informações; o uso de redes sociais como fonte primária de informação; a capacidade do conteúdo ser amplamente difundido; o distanciamento e anonimidade da fonte de conteúdo; a maior propensão a aceitar irrefletidamente conteúdos que estejam de acordo com as pré-disposições do usuário; o compartilhamento de conteúdo entre os usuários, de forma a conferir maior confiabilidade a estes; o uso de robôs digitais para impulsionar conteúdos; e o uso dos dados dos usuários para direcionar conteúdos (Alves; Maciel, 2020, p. 150).

Algumas características que são frequentemente associadas ao termo *fake news* pela literatura especializada são, por exemplo: informações claramente falsas apresentadas como reais, com o interesse expresso de enganar; notícias enganosas que são espalhadas de maneira maliciosa por fontes que fingem ser legítimas; sites que publicam desinformação como se fossem conteúdos verídicos, geralmente utilizando redes sociais para ampliar o alcance; conteúdos criados para parecerem realidade, disseminados por audiências propensas a acreditar na sua veracidade e compartilha-los, normalmente sensacionalistas e divulgadas de forma intencional; notícias falsas criadas deliberadamente para enganar, diferenciando-se do conceito de *false news*, que é relacionado à incompetência ou irresponsabilidade jornalística (Alves; Maciel, 2020, p. 152)

Todas essas características parecem estar em consonância com o fenômeno das *fake news*, porém deixam de levar em consideração qual é o papel exercido pelo fenômeno no contexto de comunicação social por meio das mídias digitais. Assim, para além de elementos como a própria falsidade do conteúdo e a sua divulgação de forma intencional ou não, é necessário levar em consideração que as *fake news* são instrumentos que estão inseridos na disputa de narrativas da realidade fática característica da comunicação sociopolítica contemporânea e ao fenômeno da pós-verdade. Sua relevância no contexto do debate público, portanto, reside no objetivo de convencimento e fortalecimento de determinadas narrativas em

detrimento da qualidade da apuração e trabalho de investigação para a produção do conteúdo. É nesse sentido que o contexto de embate e disputa ideológica e a polarização característica das plataformas digitais, conforme estudado neste capítulo, constituem elemento central do conceito de *fake news*, uma vez que a veracidade do conteúdo perde relevância frente à disseminação de conteúdos que corroborem narrativas pré-concebidas, independentemente da aferição de dolo do produtor do conteúdo (Alves; Maciel, 2020, p. 152-153).

Desta feita, a forma como os conteúdos são veiculados por meio das plataformas digitais, por meio da atuação de algoritmos pautados nos dados dos usuários, que privilegia a exposição de determinados conteúdos em detrimento de outros, com a consequente criação de filtros-bolha ou câmaras de eco, relativiza a necessidade de o conteúdo falso ser elaborado de maneira a parecer ser conteúdo jornalístico, uma vez que estes materiais passam a ser aceitos e compartilhados apenas por coadunarem com as narrativas preconcebidas dos usuários e grupos de usuários, prejudicando a avaliação crítica da veracidade do conteúdo. Assim, o que é essencial para a disseminação das *fake news* é a capacidade do conteúdo de ser amplamente compartilhado, com o objetivo de disputar narrativas da realidade fática. A forma como o conteúdo se apresenta, entretanto, ainda que não seja simulacro de material jornalístico, é elemento acessório do conceito de *fake news*.

O fenômeno do compartilhamento em massa, seja este orgânico ou feito por meio de contas falsas, chamado propagação viral, é um dos elementos das *fake news*, uma vez que o alcance das informações falsas socialmente veiculadas é uma variável importante no impacto que estas podem causar ao tecido social. De fato, a “compartilhabilidade” tem sido considerada no desenvolvimento dos conteúdos, gerando o efeito de conferir um véu de credibilidade ao material. A transmissão de conteúdo dentro de uma câmara de eco favorece o fenômeno denominado cascatas de informação, que nada mais é do que o compartilhamento em série de determinado material, os quais alimentam a transmissão viral de uma mensagem. De fato, o efeito das cascatas de informação tem maior influência para a criação de conteúdos virais que a polarização da opinião (Dourado, 2020, p. 76-82).

Stuart Hall (*apud* Wardle; Derakhshan, 2017, p. 41), sociólogo, defende que a comunicação por meio da mídia é estruturada em um processo de codificação da mensagem por parte do transmissor e consequente decodificação por parte da audiência, cuja reação pode ser de três naturezas: hegemônica, ou seja, a aceitação da mensagem tal qual ela foi codificada; negociação, aceitando-se parte da mensagem, mas não exatamente da maneira como ela foi codificada; ou oposição, quando a maneira que a mensagem foi codificada é rejeitada. Ao mesmo tempo, conforme assevera George Lakoff (*apud* Wardle; Derakhshan, 2017, p. 41-42),

especialista em linguística cognitiva, a razão não é desconexa das emoções, de forma que o ser humano compreende o mundo por meio de metáforas e narrativas que dão sentido à realidade.

Assim, Lakoff diferencia o que chama de “falsa razão” da “razão real”. A primeira seria uma espécie totalmente consciente, literal e absolutamente pautada na lógica. Ocorre que, especialmente no que diz respeito às concepções políticas, a segunda espécie, a lógica real, se impõe, caracterizada por metáforas, narrativas, emoções e perspectivas de mundo. Os sistemas morais que emergem do uso da linguagem, assim, são construções subconscientes, que agem automaticamente e são adquiridas por meio da repetição (Wardle; Derakhshan, 2017, p. 41-42).

A opinião pública, conceito que deriva do iluminismo e da revolução francesa que se deu ao final do Século XVIII, é eivada de uma espécie de mistificação, desde seu princípio, em que a verdade poderia ser revelada por meio do debate público, seja por ser entendida enquanto preexistente e, então, revelada pela preponderância do argumento verdadeiro, seja enquanto resultado do próprio choque de opiniões. A liberdade de expressão irrestrita exerce a função, nesta lógica, de permitir que a verdade seja encontrada. Compreende-se, nesse sentido, que se exercido com liberdade, o debate público não se equivocará, em uma espécie de fé na razão que dá origem ao mito de uma opinião pública iluminada (Bucci, 2021, p. 60-70).

Para o agir comunicacional pautado na razão em Habermas, os sujeitos racionais deveriam vincular-se, concomitantemente, à verdade proposicional, ou seja, a enunciar aquilo que corresponde efetivamente aos fatos narrados; à retidão normativa, em que as proposições devem corresponder ao contexto normativo derivado do direito positivo, do costume e da linguagem; e à veracidade subjetiva, em que somente deve ser enunciado aquilo em que verdadeiramente se acredita. Há, assim, uma espécie de devoção (ou fé) na razão, derivada do raciocínio iluminista, no conceito de esfera pública e opinião pública que, entretanto, exige um grau de racionalidade da ação humana que é incondizente com a realidade. Para superar a irracionalidade, parcialmente admitida no cerne de cada indivíduo, Habermas fia-se à ação comunicativa e à consciência compartilhada enquanto processo e resultado de uma racionalidade inabalável. A esfera pública é um conceito que depende, portanto, tanto da razão individual, ainda que imperfeita, e da intenção de racionalidade dos indivíduos, quanto da razão comunicativa (Bucci, 2021, p. 83-88).

A grande questão que se impõe ao mito da opinião pública é a de que a ideia está pautada no império da razão, enquanto os indivíduos agem de maneira irracional. Pela razão, o ser agiria necessariamente pautado na verdade e na moralidade, cuja conduta resultante seria voltada ao bem comum enquanto direcionamento da verdade coletivamente percebida e da moralidade coletivamente valorizada, o que não necessariamente se verifica na ação comunicacional

efetiva. Entretanto, para além do descompasso entre a irracionalidade das decisões individuais e do bem comum, há ainda uma outra irracionalidade, que reside dentro do próprio sujeito, que nem sempre age em seu próprio benefício, contrapondo internamente o desejo e o medo à consciência e a intenção (Bucci, 2021, p. 71-77), transpondo tanto na sua vivência consumidora, quanto cidadã, a complexidade da existência humana.

Nesse panorama, podemos compreender o papel de aspectos sociais e psicológicos na produção e disseminação de conteúdos por meio das plataformas digitais, a atuação dos usuários das plataformas digitais na reprodução de conteúdo que contém desinformação e a importância da exposição a conteúdos que confirmam narrativas pré-concebidas por meio de câmaras de eco e filtros-bolha na ampla disseminação destes conteúdos (Wardle; Derakhshan, 2017, p. 41-42).

As redes de que se constituem as plataformas digitais, portanto, não são apenas entre os usuários conectados, mas também de grupos de usuários conectados com outros grupos, que se relacionam entre si, constituindo, também, redes. No *WhatsApp*, a participação dos usuários em vários grupos os torna mediadores do processo de viralização, repassando as mensagens entre os grupos, desde grupos segmentados pela posição política até grupos periféricos. Assim, o conteúdo viral depende do difusor, ou fonte, e do receptor, ou destinatário, e de uma progressão por etapas. Grupos periféricos, como grupos de família, constituem papel importante, por serem numerosos, a despeito de estarem distantes do centro. A importância destes grupos revela-se pelo dado de que, nas eleições de 2018 no Brasil, 90% dos grupos de família foram usados para compartilhar informações sobre candidatos e partidos, das quais 60% eram parcial ou totalmente falsas (Dourado, 2020, p. 82).

A influência sobre o debate público surge, então, de uma abordagem direta sobre os interesses dos eleitores-consumidores, através de uma análise de técnicos em *marketing* digital sobre as interações dos próprios usuários nas mídias digitais. O engajamento dos usuários e a velocidade da comunicação digital se contrapõem à lentidão e desilusão com a política tradicional, fornecendo combustível para a narrativa anti-*establishment* adotada pelas campanhas populistas contemporâneas, que consegue transferir a participação digital para as ruas. Os movimentos populistas digitais, como o Movimento 5 Estrelas, se utilizam de uma estrutura dúplice, com um centro que manipula os discursos a partir da análise dos dados, que se mantém invisibilizado nos bastidores, e de uma organização visível, que dá a impressão de ser organicamente direcionada, composta por muitos participantes, que parecem se unir “por acaso” e que, diferentemente do centro consciente, que são os especialistas em *marketing* digital, não estão conscientes do todo (Empoli, 2019, p. 32-38).

O centro produz conteúdos sistematicamente, de acordo com os dados coletados dos usuários, que serão posteriormente encaminhados por meio de *blogs* e bolhas em mídias digitais, conclamando os usuários para, então, compartilhar esses materiais, com o objetivo de viralizar. A impressão que os usuários têm, ao compartilhar ou comentar, é de participação política, prescindindo da burocracia da política partidária tradicional, reforçando a ojeriza a esta, sugerindo uma espécie de política digital direta em oposição à democracia representativa lenta e ineficaz. Campanhas que apostam em narrativas anti-*establishment*, assim, como o Movimento 5 Estrelas, atuam como uma espécie de partido-empresa que tem a finalidade de atender os interesses do eleitor-consumidor, muito mais rapidamente que a política tradicional, oferecendo um serviço que afirma ser a superação da democracia liberal-representativa, através de uma suposta democracia direta que abarca todo o povo (Empoli, 2019, 38-51).

Assim, as estratégias adotadas pelos produtores de conteúdos com desinformação são dotadas de quatro características, pautadas em características psicológicas e sociais: provocam reação emocional; têm um forte componente visual; são dotadas de uma narrativa forte; são repetidas. Desta feita, a estratégia de campanhas de desinformação é delimitar uma mensagem cuja intenção é construir uma determinada narrativa, para então produzir os conteúdos que serão distribuídos, que pode adotar diversas formas, como textos, fotos ou vídeos. Os impactos gerados por campanhas de desinformação com motivação política são, por exemplo, a normalização de preconceitos, o fortalecimento de narrativas de nós-contra-eles e até mesmo violência (Wardle; Derakhshan, 2017, p. 38-41).

Conforme observamos anteriormente, a perspectiva de que os usuários das plataformas digitais são consumidores racionais e que analisarão criticamente os conteúdos a que são expostos nas plataformas digitais não se sustenta, tanto pela vasta quantidade de material que é compartilhado diariamente, quanto pela priorização de conteúdos de acordo com os dados do próprio usuário, quanto pelos vieses derivados da posição sociocultural, política e experiências pessoais. Assim, é importante refletir, tendo em vista a importância do compartilhamento do conteúdo pelos próprios usuários para a reprodução de desinformação, ampliando o alcance dos *posts*, de que forma os intérpretes da mensagem, ou seja, os próprios usuários, interagem com o conteúdo, seja por meio de curtidas, comentários ou compartilhamentos, podendo reagir de maneira a se opor, concordar plenamente, discordar parcialmente, compartilhar em apoio ou compartilhar em oposição (Wardle; Derakhshan, 2017, p. 27-28).

Destaca-se, ainda, o papel da comunicação, que deve ser entendido como mais que mera transmissão de informação entre os indivíduos, mas enquanto a construção social de crenças compartilhadas, perspectiva que permite adotar uma visão psicológica e social para

compreender a disseminação e consumo de desinformação. A estrutura das plataformas digitais, nesse sentido, ao gerar incentivos para a socialização entre os grupos de usuários por meio de curtidas, comentários ou compartilhamentos, privilegia a disseminação de conteúdos com alta valência emocional, em detrimento de uma avaliação crítica de veracidade do conteúdo (Wardle; Derakhshan, 2017, p. 7-15).

Aspectos sociológicos e psicológicos ajudam a compreender o motivo pelo qual as pessoas tendem a compartilhar conteúdo *online*, segundo Berger e Milkman (*apud* Dourado, 2020, p. 79-80), especialmente pautados na valência emocional, ou seja, no fato de o conteúdo conter apelo emocional, seja este positivo ou negativo; e o papel da ativação. Emoções positivas são mais bem recebidas no contato interpessoal, mas emoções como temor, raiva e ansiedade também aumentam o nível de disseminação, diferentemente da tristeza. O conteúdo com mais raiva ou diversão é mais compartilhado e é impulsionado pelo nível de ativação que produz. A importância da emoção para a disseminação de conteúdo *online* implica em que aqueles mais moderados têm menor tendência a participar e, no que diz respeito a conteúdos políticos, os indivíduos têm maior tendência à conformação ideológica em relação aos conteúdos compartilhados (Dourado, 2020, p. 79-80).

Pela natureza da comunicação *online*, facilitada, há maior visibilidade a conteúdos que instigam o conflito. Chadwick e Vaccari (*apud* Dourado, 2020, p. 80-81), pesquisando sobre os motivos pelos quais as pessoas publicam informações falsas *online*, descobriram que 65,5% dos usuários no Reino Unido alegaram ter feito como uma forma de expressar seus sentimentos, 51,1% que gostariam de saber a opinião de outras pessoas, 43,9% para influenciar outros e 43,7% para provocar discussões. As motivações, portanto, foram tanto por conta de um meio de expressão, lugar que as plataformas digitais ocupam, quanto pela intenção de estimular posicionamento em terceiros (Dourado, 2020, p. 80-81).

Há, ainda, dois outros instrumentos cujo papel é fundamental no compartilhamento de informações falsas na esfera digital: os influenciadores e os perfis falsos. Os influenciadores e figuras públicas são perfis que contêm audiência exponencial e que podem ser políticos, artistas, escritores, intelectuais, profissionais do direito, jornalistas, ou pessoas que ganham fama na própria *internet*. Esses usuários constituem nó central da rede de disseminação de *fake news*, pois atuam como intérpretes que dão sentido à mensagem, agindo enquanto líderes de opinião (Dourado, 2020, p. 84-85).

O uso de perfis falsos, também chamados robôs ou *bots*, da mesma forma, é associado à disseminação de conteúdo falso *online*. Há os designados robôs sociais, que simulam um usuário real nas plataformas digitais, simulando movimentos populares espontâneos e,

consequentemente, influenciando o funcionamento do algoritmo, e sua atividade passou a ser identificada desde 2010, no *Twitter*. Outro tipo de perfil falso são os chamados *troll*, que são administrados por humanos e que tem como função organizar ataques sistemáticos a terceiros (Dourado, 2020, p. 87-88).

Há registros de ampla participação de perfis falsos na disseminação de *fake news* e conteúdo em geral na *internet*, por exemplo no caso da influência russa em eleições em outros países. No Brasil, durante as eleições de 2014, constatou-se a utilização de perfis falsos para criticar a candidata Dilma Rousseff e, durante um debate eleitoral ocorrido na Rede Globo, verificou-se que 11% das publicações no *Twitter* eram oriundas de robôs, o que significa mais de 1,5 milhão de interações, da mesma forma que no processo de *impeachment* da Presidente Dilma Rousseff, no qual durante a maior manifestação, em 13 de março, 10% das interações eram oriundas de robôs (mais de 800 mil interações) e durante a greve geral de 28 de abril de 2017 eram 22,39% do total de interações (Dourado, 2020, p. 88-91).

Na eleição dos Estados Unidos da América em 2016 constatou-se a participação de mais de 400 mil robôs em meio a 2,8 milhões de usuários. Esses robôs foram responsáveis por 3,8 milhões de *tweets* sobre as eleições. O próprio *Facebook* admitiu a atuação do governo russo na plataforma, gerando cerca de 80 mil conteúdos que atingiram em torno de 126 milhões de estadunidenses. O *Twitter* também reconheceu ação do governo russo, encontrando quase 4 mil contas vinculadas a este, manipuladas pela *Internet Research Agency*, agência russa responsável por financiar *bots* e *trolls* para influenciar discussões políticas (Dourado, 2020, p. 91-92).

Desta feita, um elemento que também constitui o conceito de *fake news* diz respeito à capacidade deste material se disseminar nas plataformas digitais, conhecido como efeito viral, que, pela ampla propagação, acaba por conferir maior confiabilidade ao material. Uma das características geralmente presentes nas *fake news* são o forte apelo emocional, pautado na ideia de acontecimentos urgentes, escandalosos e graves, que promovem a ampla divulgação deste tipo de conteúdo (Dourado, 2020, p. 57).

Algumas características contidas nas *fake news* aumentam sua aceitabilidade perante os usuários, como mensagens que exploram estereótipos; que estimulam emoções como raiva ou ansiedade; que tratam de um assunto atual; que viralizam; e mensagens audiovisuais, que tem mais apelo que outros formatos. A oferta de informações nas redes sociais passa, portanto, a se relacionar com o que é chamado de pós-verdade, fenômeno que explica a preferência por conteúdos emocionais em detrimento dos fatos objetivos (Dourado, 2020, p. 97).

Portanto, as *fake news* não necessariamente apresentam a estrutura formal de um material jornalístico, podendo ocorrer em formatos diversos como fotos, vídeos, *cards*, textos

e áudios, bastando que se apresente enquanto informação. Nesse sentido, fica evidente a distinção entre informação falsa e *fake news*, uma vez que materiais que contém informação falsa, mas que não pretende apresentar-se como fato, não se caracteriza enquanto *fake news*, por exemplo, uma sátira, comentário, opinião, teorias da conspiração ou reportagens com erros não intencionais (Dourado, 2020, p. 55-57).

É importante, ainda, salientar que somente as características da comunicação digital não são suficientes para compreender o fenômeno da desinformação *online*. Assim, os efeitos da desinformação por meio das novas mídias devem ser reconhecidos enquanto parte de um complexo emaranhado de fatores sociais, econômicos, políticos e tecnológicos das sociedades contemporâneas, especialmente das democracias ocidentais, citando-se: o aumento da segregação de raça, classe, religião e política; a falência do estado de bem-estar social; a incapacidade do Estado de oferecer serviços públicos; as mudanças climáticas, entre outros (Wardle; Derakhshan, 2017, p. 14).

Desta feita, há três dimensões explicativas que ajudam a compreender o poder de disseminação e, conseqüente, os efeitos produzidos pelas *fake news* na sociedade: o contexto político pautado na polarização; o ambiente comunicacional, que está centrado no digital e implica em mudanças no consumo de notícias; as condições sociotécnicas, tendo em vista a personalização do conteúdo, a mediação algorítmica e a automatização na difusão de mensagens (Dourado, 2020, p. 58-59).

1.7. *Fake news*: uma proposta de conceito

É preciso ter em mente que o fenômeno das *fake news* está inserido em um contexto político e cultural marcado pela radicalização política, pelas crises econômicas e a desconfiança em instituições políticas e midiáticas. Nesse panorama de crises política e econômica, as *fake news* exercem um papel de estratégia comunicacional, uma espécie de “informação de combate”, em que para além da produção de notícias falsas em si, há a instituição de mídias hiper-partidárias (Alves; Maciel, 2020, p. 151). Sob esse prisma, as *fake news* devem ser entendidas enquanto uma espécie de “informação de combate”, ou seja, conteúdos que, independentemente de serem sensacionalistas, tirados de contexto, exageros, especulações tratadas como fatos ou mesmo mentiras propriamente, sejam elaborados de forma a disputar sentidos da realidade, veiculados por mídias hiper-partidárias, em um contexto de polarização política (Ribeiro; Ortellado, 2018, p. 73).

A definição de *fake news* pode ser a de elaboração de materiais que pretendem gerar engano ou desinformação sobre determinado acontecimento, construindo versão inverídica sobre este. Pressupõe, portanto, que é material que se refere a fatos, mas que busca disputar sentidos pautado na ignorância dos fatos reais, resultando em má informação. *Fake news*, portanto, se distingue do conceito de desinformação, opinião e comentário, de forma que é necessário especificar seus sentidos e impactos na esfera pública política *online*, para possibilitar o combate à prática antidemocrática, por um lado, e, por outro, proteger a liberdade de expressão (Dourado, 2020, p. 40-41). As *fake news* são, nesse sentido, uma espécie dos conteúdos que podem gerar distorções na *internet*, simulando tratar-se de notícia ou novidade (*news*). Faz-se crer notícia não apenas pelo seu formato, simulacro jornalístico, que é dispensável para caracterizar-se enquanto *fake news*, mas principalmente por seu conteúdo que alega narrar fatos, não ficção (Dourado, 2020, p. 53-54).

Assim, na forma, tais materiais podem conter o uso de fontes, formato, manchetes, linguagem direta e distribuição via sites, simulando tratar-se de matéria jornalística, para conferir confiabilidade, mas tais elementos são apenas acidentais. Outros elementos textuais que podem acompanhar *fake news*, mesmo em conteúdos audiovisuais, é o uso dos termos urgente, exclusivo, ou ainda o pedido de compartilhamento do material, dando a impressão de tratar-se de notícias. Uma vez produzido o material, deliberadamente com o objetivo de confundir, este passa a circular por meio de robôs e impulsionamento, mas também de maneira orgânica entre os usuários que confiam na veracidade daquele conteúdo, de forma a conferir mais um véu de confiabilidade, mediante o recebimento de conteúdos por parte de pessoa conhecida (Dourado, 2020, p. 54-55).

Ademais, além da forma que as mídias tradicionais produzem e divulgam conteúdo, de maneira distinta da que ocorre nas novas mídias, em razão de aspectos técnicos derivados do próprio funcionamento das diferentes mídias, aquelas estão sujeitas a normas que regulamentam seu adequado desempenho, de forma que há procedimentos editoriais que devem ser cumpridos, bem como as empresas podem ser responsabilizadas pelo conteúdo que publicam e estão sujeitas a Códigos de Ética e leis que garantem a integridade da informação, como por exemplo a Lei nº 13.188/2015, que dispõe sobre o direito de resposta a pessoa ofendida. O mesmo nível de regulamentação, entretanto, não ocorre sobre a publicação de conteúdos pelos usuários nas chamadas mídias digitais (Alves; Maciel, 2020, p. 150).

No ano de 2013 o Fórum Econômico Mundial já reconheceu a disseminação massiva de desinformação digital enquanto um dos maiores riscos mundiais para o próximo decênio, relacionando o fenômeno a outros 18 dos 50 riscos apontados pela instituição: aumento do

fanatismo religioso, reação contra a globalização, fracasso nos regimes de propriedade intelectual, terrorismo, falha crítica do sistema, fracasso da governança global, incidentes massivos de fraude ou roubo de dados; “falha financeira sistêmica grave, ciberataques, fracasso na resolução de conflitos diplomáticos, má gestão do envelhecimento da população, disparidade de renda, crises de escassez de comida, fracasso na adaptação à mudança climática, desequilíbrio crônico do mercado de trabalho e consequências imprevistas das novas tecnologias de ciências da vida” (Dourado, 2020, p. 109-110).

Pela caracterização do fenômeno das *fake news* e do discurso de ódio e sua disseminação por meio das plataformas digitais, demonstra-se a relevância de algum tipo de regulamentação sobre o fluxo de conteúdos digitais, respeitando-se a liberdade de expressão. Pela natureza do setor de tecnologia, trata-se de desafio transnacional, implicando em, ao menos, atuação das próprias plataformas na moderação de conteúdos e elaboração de legislação em cada país (Dourado, 2020, p. 109).

Ao longo deste primeiro capítulo discorreremos sobre fatores socioeconômicos, políticos e tecnológicos que caracterizam a comunicação social intermediada pelas mídias digitais, desde sua popularização a partir dos anos 1990, especialmente por meio do *www*, até a criação de novas plataformas de comunicação e informação digitais, notadamente as redes sociais. Buscamos compreender de que forma as mídias digitais se diferenciam das mídias de massa tradicionais, gerando a segmentação dos conteúdos acessados pelos usuários, processo intensificado pela atuação das plataformas digitais por meio da psicométrica, utilizada para realizar o chamado *microtargeting*, ou seja, a veiculação de conteúdo de acordo com os interesses e perfil dos usuários, do que decorre a capacidade de influenciar o debate público de maneira inadvertida, por meio da própria formação de vontade, que deixa de ser livre. O resultado é a formação de mídias hiper-partidárias, caracterizadas pela formação de filtros-bolha ou câmaras de eco, ambiente propício para a veiculação daquilo que designamos *fake news* e discurso de ódio.

O conceito de *fake news* delimitado ao longo deste primeiro capítulo permite explicitar o desafio que a veiculação de desinformação nas plataformas digitais apresenta às democracias contemporâneas, evidenciando fatores sociais como a própria utilização das tecnologias de comunicação e informação pelos usuários, ou econômicos, como o interesse de determinado produtor ou disseminador de conteúdo nas mídias digitais, que pode ser indiferente quanto ao resultado, como influenciar o debate público em determinado período eleitoral, mas que lucra por meio da atuação neste processo, contratado por quem efetivamente tem interesse direto no

resultado. Este processo, entretanto, acontece nos bastidores e dificilmente é de conhecimento público.

A produção de conteúdo para ser veiculado pelas mídias-hiperpartidárias é pautada na racionalidade digital, ou seja, no objetivo de que o material viralize e não, necessariamente, esteja pautado na realidade. Desta feita, veicularam por estas redes materiais que buscam disputar sentidos da realidade fática, criando grupos de usuários em torno de determinadas redes, fora das quais todo o conteúdo é considerado falso, em um fenômeno chamado epistemologia tribal, do qual deriva a formação de identidades irreconciliáveis, dificultando o debate pautado na racionalidade discursiva, pressuposto da democracia liberal representativa.

No segundo capítulo discutiremos, mediante as reflexões elaboradas ao longo deste primeiro capítulo, as possibilidades de regulamentação da comunicação digital cujo objetivo é minimizar os efeitos perniciosos da disseminação de *fake news* e discurso de ódio nas plataformas digitais nas democracias contemporâneas. Nesse sentido, algumas perguntas que surgem ao longo desta reflexão sobre a estrutura da comunicação digital podem ajudar a guiar o estudo das possibilidades de regulamentação:

- 1) O sistema de isenção de responsabilidade das plataformas sobre os conteúdos postados por terceiros, com a consequente delegação ao Poder Judiciário da tarefa de decidir caso-a-caso, instituído pelo Marco Civil da Internet, é meio adequado/desejado para coibir a veiculação de conteúdos perniciosos pelas mídias digitais, ao mesmo tempo em que protege a liberdade de expressão? De que maneira a novas estrutura comunicacional e a racionalidade digital interferem no conceito jurídico de liberdade de expressão?
- 2) Nesse mesmo sentido, a estrutura técnica das comunicações digitais difere das mídias de massa tradicionais, como o rádio e a televisão, sobre os quais, mediante o interesse público derivado da capacidade de comunicar a milhões de expectadores, recaem responsabilidades constitucionais e normas de ética jornalística sobre o material veiculado. Assim, a isenção de responsabilidade das plataformas gera desigualdades na normatização das diferentes mídias, tendo em vista que as plataformas digitais também são acessadas por milhões de usuários diariamente, de maneira a facilitar a disseminação de conteúdo pernicioso pelas mídias digitais?
- 3) Há algum tipo de material ou discurso que não é protegido, na tradição constitucional brasileira, pelo princípio fundamental da liberdade de expressão, sopesados outros bens públicos? Quais são esses conteúdos? É possível estabelecer critérios objetivos para a identificação desses materiais *a priori*? A qual agente cabe

realizar a complexa tarefa de avaliar os conteúdos veiculados digitalmente, tendo em vista o excesso de conteúdo publicado diariamente?

CAPÍTULO 2- DEMOCRACIA, LIBERDADE DE EXPRESSÃO E GOVERNANÇA DOS DISCURSOS NAS MÍDIAS DIGITAIS

A comunicação digital modificou a esfera pública, de forma que empresas detentoras destas plataformas, como o *Facebook* e o *Google*, passaram a automatizar processos e decisões que anteriormente, nas formas de mídia de massa tradicionais, eram tomadas por editores de jornais ou diretores de televisão. De fato, foi criada uma espécie de “esfera pública automatizada”, mediante a qual os tradicionais mecanismos de controle e proteção social são ineficientes, tornando necessário que o monitoramento e regulação da comunicação digital acompanhe seus avanços tecnológicos (Pasquale, 2017b, p. 1).

Durante os anos 2000, o exponencial aumento do uso da *internet* foi objeto de bastante otimismo, pelas possibilidades de acesso e veiculação de conteúdos e a possibilidade de gerar efeitos socialmente benéficos. Essa visão otimista resultou em uma espécie de receio do Poder Público, especialmente do estadunidense, onde se desenvolveram as principais empresas de tecnologia de comunicação e informação digital, em refletir sobre formas de regulação. Essa expectativa, entretanto, foi frustrada ao longo dos anos 2010, uma vez que o mercado digital se caracteriza por um oligopólio, com a concentração de cerca de 75% da receita do mercado de propaganda digital em 2017, um mercado no valor de 73 bilhões de dólares, em apenas duas empresas, *Google* e *Facebook*, que atuam tendo como objetivo o lucro e cujo funcionamento dos algoritmos é secreto. Aquela visão positiva de uma comunidade digital vibrante foi paulatinamente substituída por uma outra realidade, a de uma “sociedade da caixa-preta”, em que *trolls*, robôs e até mesmo governos estrangeiros atuam para manipular os ambientes digitais (Pasquale, 2017b, p. 2).

O funcionamento das plataformas digitais permite a anonimidade de agentes que financiam campanhas de disseminação de conteúdo *online* que podem enganar o público, os legisladores e reguladores, ao mesmo tempo em que a fragmentação da audiência concentrou nas plataformas digitais um poder sem precedentes de direcionamento de conteúdo. As plataformas digitais, assim, tem a capacidade de decidir qual tipo e qualidade de conteúdo a audiência acessa, pautando-se no que é denominado *big data*, ou seja, a capacidade de criar um perfil detalhado dos usuários a partir dos dados que compartilham, que permitem monitorar, monetizar, controlar e prever o comportamento humano, o que é de interesse de entidades comerciais, campanhas políticas e governos, por exemplo, prejudicando processos

democráticos como a liberdade da formação de vontade, o pluralismo político e o diálogo deliberativo (Pasquale, 2017b, p. 2-3).

Uma vez que as plataformas medem o sucesso de uma publicação pelo engajamento que ela gera, ou seja, a capacidade do material “viralizar”, de acordo com o perfil dos usuários, em detrimento da veracidade, automatizando a esfera pública, perde-se uma das principais funções do discurso público, qual seja o debate entre diferentes interesses e aspirações, substituído por algoritmos que buscam maximizar os lucros. Tal estrutura possibilita a manipulação de eleitores, especialmente aqueles caracterizados por serem mais desinformados e indecisos, os quais costumam ser fundamentais para vencer eleições (Pasquale, 2017b, p. 3-4).

A preocupação com o impacto das mídias na esfera pública remonta o surgimento e popularização da mídia de massa tradicional, a qual alterou, ao concentrar em corporações e Estados o controle sobre as mídias, a formação de opinião e de vontade na esfera pública. As decisões que eram tomadas pelas mídias de massa tradicionais passaram a ser automatizadas nas mídias digitais, e a estrutura resultante da comunicação digital vem sendo utilizada para influenciar decisões políticas, como ficou evidente nas eleições presidenciais americanas de 2016 em que foram veiculadas informações falsas sobre a candidata Hillary Clinton, gerando oportunidades de lucro para produtores de conteúdos que criam e disseminam mentiras e incentivando o uso de *bots* para disseminar determinada mensagem, fazendo parecer que se trata de um processo orgânico (Pasquale, 2017b, p. 4-7).

Para além, o *Facebook* foi utilizado, durante a campanha eleitoral estadunidense de 2016, para desincentivar o voto em comunidades afro-americanas, por exemplo. De fato, diversos casos demonstram como os mecanismos das plataformas digitais levam a resultados prejudiciais, como a apresentação de resultados racistas ou misóginos em pesquisas ou a informações falsas. Em um caso paradigmático, o autor de um atentado em uma igreja na Carolina do Sul em 2015, que resultou na morte de nove afro-americanos, disse que quando pesquisou no *Google* sobre crimes de cunho racista, anteriormente ao cometimento do crime, obteve como primeiro *link* um *site* de uma organização supremacista branca, e que isso mudou para sempre a sua vida. A violência e desinformação resultante da comunicação digital é diretamente derivada do modelo de funcionamento das plataformas, a esfera pública automatizada, e a gravidade e o potencial de ruptura democrática derivado destas escolhas tomadas pelas grandes empresas de tecnologia justifica a demanda por maior transparência e responsabilidade para o setor, após um período de inércia (Pasquale, 2017b, p. 4-7).

2.1. A racionalidade digital e a fragmentação da esfera pública: impactos no princípio democrático da liberdade de expressão

O direito fundamental de votar é o mais básico princípio da cidadania em sociedades democráticas. A manifestação efetiva deste direito, entretanto, depende da realização de direitos e garantias democráticos, tais como a liberdade de expressão, opinião, informação e reunião. Mediante a ampla veiculação de conteúdos com *fake news* durante períodos eleitorais, impactando no direito à informação dos usuários e prejudicando o debate público, os Estados têm buscado adotar medidas contra a disseminação destes materiais. Todavia, a tarefa de moderar conteúdo é complexa, uma vez que se discute a legitimidade de o Estado indicar quais conteúdos constituem ou não *fake news*, gerando repercussões negativas a outro direito fundamental, a liberdade de expressão. Há, assim, uma tensão entre o direito à liberdade de informação e da liberdade de expressão, ambos princípios fundamentais da democracia (Horbach, 2022, p. 187-188).

A despeito de a própria atividade de se produzir mentiras e notícias falsas, especialmente durante períodos eleitorais, não ser novidade, a inserção das tecnologias de comunicação digital modificou a maneira como este processo ocorre. O rápido crescimento do acesso à *internet*, que alcança cerca de 62% da população global e quase 70% da população brasileira, ou 149 milhões de brasileiros, juntamente com a vasta quantidade de atividades desenvolvidas virtualmente, desde a compra de produtos, a comunicação, a informação e encontros, impôs a necessidade de se questionar qual seria o tratamento jurídico conferido a cada uma destas atividades.

A *internet* se tornou uma agregadora de diferentes atividades cotidianas, desde banalidades até temas de alta complexidade, sendo apontada, conforme referido alhures, como um dos principais meios de acesso à informação. Não é de se estranhar, portanto, que tais instrumentos tenham sido adotados enquanto ferramentas da disputa política, uma vez que a vida em comunidade é um dos aspectos basilares da humanidade. Desta feita, partidos, candidatos e pessoas públicas passaram a utilizar os meios digitais para promover suas ideias, criando uma conexão direta com os eleitores, sem a necessidade de intermediários. Ao mesmo tempo, os próprios usuários podem se utilizar desses meios para divulgarem e promoverem ideias com as quais concordam, de maneira espontânea e voluntária (Horbach, 2022, p. 188-190), de forma que a disseminação de conteúdos na *internet*, conforme discutimos no primeiro capítulo, não deriva apenas da manipulação por meio da chamada psicopolítica dado-pulsional, mas também por interesses e manifestações legítimas, protegidas pelo princípio constitucional da liberdade de expressão.

A interpretação dada ao princípio democrático da liberdade de expressão é moldada de acordo com as tecnologias de comunicação disponíveis em determinando tempo e espaço e pela forma como as pessoas efetivamente usam essas tecnologias. Independentemente do momento tecnológico das comunicações, o objetivo da proteção à liberdade de expressão é garantir aos indivíduos a possibilidade de participar das decisões e encontrar sentido de vida, enquanto pressupostos democráticos. Desta feita, a proteção à liberdade de expressão surge enquanto instrumento da Democracia para a promoção da participação em comunidade de um debate democrático e livre.

O conceito de esfera pública traz, em seu âmago, uma contraposição ao privado, conforme discutimos no primeiro capítulo, de forma que um de seus principais elementos conceituais é que sua ocorrência se dá em espaços socialmente compartilhados. Sua constituição se dá por meio da formação de uma espécie de cultura compartilhada, inicialmente nas grandes cidades, e na qual os meios de comunicação, como livros, jornais, periódicos, rádio, televisão e redes sociais exercem um papel “para o estado agregado da esfera pública”. A forma como nos comunicamos enquanto sociedade, portanto, estabelece os contornos do que é denominado esfera pública, de forma que é fundamental, assim, compreender as mudanças da esfera pública pela comunicação digital enquanto pressuposto para um esforço normativo compatível à nova esfera pública digital (Vesting, 2019, p. 195-196).

O surgimento da esfera pública se dá com o estabelecimento das grandes cidades no oeste europeu durante o século XVIII, iniciando-se especificamente na cidade de Londres, em que desde o século XVII começaram a se desenvolver princípios de liberdade republicana, como a impressão sem necessidade de licença, resultando em um novo paradigma para a cultura política moderna, em que a hegemonia cultural se desloca das Cortes para as cidades. Se antes as relações eram constituídas a partir de um centro (o rei, a família real, a Corte ou o clero), o surgimento e popularização da imprensa estabelece uma nova forma de comunicação, horizontal, institucionalizando redes de comunicação e relações sociais nas cidades que não são uniformes, resultando na formação da opinião pública em um poder anônimo e invisível que independe das estruturas de poder e que influencia decisivamente nos assuntos públicos. Assim, a confluência de fatores como o exponencial crescimento urbano, a ascensão de escolas e universidades, a literalização da população e as novas técnicas de imprensa, intimamente relacionadas com a contemporânea ascensão da burguesia, resultam na centralidade de um novo modo de sociabilidade comercial/burguês, que floresce na sociedade civil (Vesting, 2019, p. 196-198).

Desenvolve-se, a partir do imperativo de instituir-se uma esfera pública urbana independente de hierarquia, o aparato jurídico dos direitos de liberdades sociais, chamados pelos ingleses de direitos naturais, pelos franceses de direitos humanos e na Alemanha de direitos subjetivos ou direitos fundamentais, notadamente o direito à liberdade de opinião, imprensa e associação. Esse arcabouço jurídico tem como pressuposto a individualidade burguesa, tendo como ideal de homem burguês, o *gentleman*, com aptidão à vida em sociedade, fazendo parte da comunidade, desenvolvendo e cultivando boas maneiras, a capacidade a conversas agradáveis e formas elegantes de trato social. O surgimento da esfera pública trata-se, portanto, não somente de uma estrutura que age politicamente, enquanto um direcionamento comum às decisões do Parlamento, mas da formação de uma cultura que transcende a política, gerando novas possibilidades de comunicação e estabelecimento de uma vida comum (Vesting, 2019, p. 198-199).

No fim do século XIX e início do século XX, as condições sociopolíticas e tecnológicas que primeiro estabeleceram a esfera pública modificam-se, substituindo-se o individualismo burguês pela nova cultura de massa, pautada em meios de comunicação como jornais diários, fotografia, filmes mudos, rádio e livros de bolso, que substituem a literatura, pintura, teatro e arquitetura. A liberdade igualitária enquanto princípio de sociabilidade é substituída pelo pluralismo de grupos, centralizando em grupos e organizações sociais formais, como partidos políticos, associações sociais, igrejas, sindicatos, grandes editoras de livros e emissoras de rádio, a formação e expressão da opinião pública. Estes grupos oferecem uma “pré-estruturação do intercâmbio de opinião”, ao que se segue o reconhecimento de direitos coletivos de liberdade como a liberdade de associação, expressão e o direito de participação dos partidos políticos na formação da vontade popular (Vesting, 2019, p. 199-201).

Esse modelo de esfera pública de pluralismo de grupos se organiza em três círculos concêntricos assegurados institucionalmente, constituídos por um primeiro círculo de opiniões, mais abrangente, que são expressas e ouvidas de maneira difusa, nas escolas, vizinhanças, locais de trabalho, praças, etc.; um segundo círculo de comunicações politicamente relevantes veiculadas pelas mídias de massa, que pautam os temas a serem discutidos, informadas pela comunicação organizada, partidária e associativa; e um terceiro círculo, central, de formação da vontade política, exercido no seio dos órgãos estatais que absorvem as opiniões agregadas para resultar em decisões vinculantes, realizadas especialmente pelo Parlamento. Apesar de a liberdade individual burguesa manter-se enquanto pressuposto na esfera pública pluralista de grupos, o indivíduo enquanto elemento central de sociabilidade é substituído pela organização em grupos, de forma que o ideal de homem burguês, pautado no *gentleman*, é substituído pelo

empregado, funcionário, servidor público ou membro do partido, ou seja, enquanto parte de um grupo ou organização e que age de acordo com estas (Vesting, 2019, p. 201).

A questão da moderação de conteúdos na comunicação, nesse sentido, antecede as mídias digitais e contrapõe princípios constitucionais democráticos, especialmente a liberdade de expressão e direitos como a privacidade, autorais e a honra. A forma como o tema é regulamentado deve levar em consideração aspectos sobre as tecnologias de informação e comunicação socialmente relevantes, em busca de estabelecer mecanismos de sopesamento adequados, protegendo-se a democracia. Nesse sentido, a inserção dos meios de comunicação de massa, notadamente do rádio e da televisão, modificou o panorama da comunicação social, gerando a necessidade de novas perspectivas e regulamentação no que diz respeito à moderação de conteúdo (Hartmann, 2020, p. 1-2).

Um primeiro aspecto a ser destacado sobre a moderação de conteúdos na era da comunicação de massa tradicional é a concentração da decisão sobre o conteúdo a ser disseminado nas mãos de pequenos grupos detentores dos meios de comunicação, resultando em uma estrutura comunicacional denominada de um-para-muitos (*one-to-many*). Desta feita, há uma severa diferença na capacidade de disseminar conteúdo, que se concentra na mão dos poucos proprietários dos meios de comunicação, de forma que, enquanto os indivíduos de maneira geral conseguem se comunicar apenas com um certo grupo de pessoas conhecidas, as mídias de massa passam a comunicar para milhões de espectadores. Essa capacidade de comunicação em massa criou obrigações adicionais às mídias tradicionais, justamente em função do alcance que as informações veiculadas por estas tem e os eventuais danos que podem ser causados a terceiros (Hartmann, 2020, p. 2-3).

As decisões quanto aos conteúdos veiculados nos meios de comunicação de massa tradicionais, portanto, cabiam aos proprietários dos meios de comunicação, seja por meio da atuação de editores ou da escolha de repórteres, apresentadores ou outras profissões dos meios de comunicação tradicional, que se responsabilizavam por sua legalidade, exercendo um trabalho de curadoria. Neste sistema de comunicação, ao Estado não cabia atuar na moderação de conteúdo, ocorrendo apenas a responsabilização posterior pela via judicial. Tendo em vista que a quantidade de veículos de comunicação e de conteúdos disseminados, neste contexto, é relativamente pequena, em comparação às plataformas digitais, a opção pela via judicial caso a caso para determinar a legalidade/ilegalidade dos conteúdos era adequada à estrutura tecnológica comunicacional.

Os meios pelos quais o discurso era moderado estavam restritos à compensação monetária e o direito de resposta, aplicados majoritariamente pelo Judiciário, enquanto

mecanismos como remoção ou restrição prévia eram entendidos enquanto censura. Outro aspecto importante que foi modificado com as plataformas digitais é que as mídias de massa tradicionais não tinham como controlar especificamente quem seria alcançado pela mensagem que produziam e disseminavam, de forma que ao publicar um jornal, ou transmitir um programa de televisão ou de rádio, o máximo que poderia ser controlado era o alcance de onde estes seriam distribuídos, mas não era viável controlar quem lia, assistia ou ouvia àquele material (Hartmann, 2020, p. 3), como é possível pela estrutura de vigilância estabelecida na comunicação digital, com o direcionamento de conteúdo.

A inserção da comunicação digital modificou diversos aspectos relacionados à mídia tradicional de massa. Inicialmente, a concentração na capacidade de produzir e disseminar conteúdo foi subdividida em dois aspectos: se por um lado bilhões de usuários das plataformas digitais podem decidir quais conteúdos publicam e compartilham, por outro as grandes empresas de tecnologia concentram ainda mais a capacidade de tomar decisões sobre a arquitetura das plataformas e a forma os conteúdos serão veiculados em suas plataformas, influenciando decisivamente no conteúdo que será amplamente veiculado: ou seja, se aos indivíduos em geral foi possibilitado produzir e disseminar conteúdo, o profundo desequilíbrio no poder de decidir como os conteúdos serão disseminados, que já ocorria na época da comunicação de massa tradicional, foi mantido, transferindo-se às *big techs* um poder ainda mais abrangente (Hartmann, 2020, p. 3-4).

A estrutura do modelo de esfera pública pluralista de grupos, calcada nas mídias de massa tradicionais, apesar de ser pautado na ideia de permeabilidade das perspectivas individuais que podem ascender à importância social, submerge estas expressões por meio da participação em grupos ou organizações, em arenas políticas nas quais se realiza o debate público animado pela racionalidade discursiva. Na era da comunicação digital, entretanto, a comunicação não é mais pautada em grupos ou organizações, como partidos políticos ou sistemas de rádio, de forma que se dá por meio de tecnologias que dispõe da ampla possibilidade de formação de composições e agrupamentos, independentemente da pré-formação pluralística de grupos. Nesse panorama, são as empresas de comunicação digital que exercem uma espécie de curadoria, em que estruturam por meio de algoritmos a forma como se dá a comunicação social (Vesting, 2019, p. 202-203), sem que haja a submissão ao debate público e, conforme destacamos no primeiro capítulo, sem sequer a relação com o outro.

Os próprios limites materiais de veiculação de material, como o tamanho da folha de jornal ou o tempo de um programa na televisão ou no rádio, que impediriam a veiculação de outros conteúdos, foi flexibilizada, de forma que a quantidade de conteúdo que pode ser

veiculado e de indivíduos que podem circular informações foi ampliada, além de os custos para tanto serem proporcionalmente muito baixos, uma vez que disponibilizar um *site* na *internet* ou participar de uma rede social, por exemplo, tem custos muito baixos ou são, de fato, gratuitos. Esse novo panorama comunicacional possibilita que um conteúdo postado por um indivíduo, que não teria acesso aos meios de comunicação de massa tradicionais, possa ser largamente compartilhado, no fenômeno de viralização (Hartmann, 2020, p. 4).

Ou seja, o excesso de informações veiculadas nas plataformas digitais e o acesso individualizado aos conteúdos de interesse, possibilitam às plataformas digitais, por meio de seus algoritmos, baseado na análise dos dados dos usuários, moldarem os conteúdos que serão apresentados aos usuários. A soma das preferências dos usuários com a atuação de algoritmos resulta na formação de câmaras de eco que dificultam o debate plural de ideias e reforçam ideias e preconceitos do grupo. Quando a estrutura das plataformas digitais é utilizada para influenciar em questões publicamente relevantes, como em disputas eleitorais, como no caso da eleição presidencial dos EUA em 2016 e do Brexit, há um grave prejuízo à democracia, não apenas ao influenciar o próprio resultado das eleições, mas também colocando em questão todo o sistema eleitoral ou incentivando uma polarização radicalizada (Macedo Júnior, 2020, p. 237-240).

Uma vez que o funcionamento dos algoritmos é secreto, especialmente em razão de se tratar de um segredo empresarial, torna-se muito difícil identificar de que forma resultados prejudiciais das tecnologias de informação e comunicação digital se dão, ocultando-se vieses discriminatórios, práticas anticompetitivas ou condutas negligentes sob um véu de inescrutabilidade técnica. Sob essa perspectiva, agentes que queiram manipular o debate público são protegidos pela opacidade desses sistemas, do que se depreende a necessidade de que seja possível a realização de auditorias sobre os sistemas adotados pelas plataformas digitais, no mínimo no que diz respeito aos dados que alimentam o sistema, para que os reguladores possam entender de que forma estes funcionam. A ideia de que as plataformas digitais propiciam um “mercado de ideias” no qual prevalecem aquelas mais convincentes não se sustenta, conforme foi-se percebendo ao longo dos anos 2010, com a efetiva utilização das plataformas digitais, uma vez que através do investimento em *marketing* digital e do teste quanto a qual conteúdo tem maior capacidade de viralização, é possível manipular o conteúdo veiculado digitalmente (Pasquale, 2017b, p. 7-11).

Conforme discutimos no primeiro capítulo, a comunicação digital é dotada de características que a especificam e que a torna um ambiente propício para a disseminação de *fake news* e discurso de ódio, gerando potencial prejuízo às democracias.

A velocidade com que as informações são compartilhadas é muito maior que nas mídias tradicionais, em que havia um processo de editoria que confirmava a veracidade do conteúdo antes de veiculá-lo. Nas mídias digitais, a necessidade de uma notícia ser veiculada em tempo real gera uma lógica distinta, em que se publica antes e se verifica depois. A estrutura das plataformas também iguala fontes confiáveis com outras fontes, por vezes desconhecidas, diminuindo a capacidade dos usuários de realizarem uma análise crítica do material;

Os usuários podem customizar os conteúdos que acessam por meio das mídias digitais, se expondo apenas a conteúdo e opiniões que coadunam com suas perspectivas políticas e concepções. Se nas mídias tradicionais já havia diversos canais, de maneira que o espectador poderia realizar uma escolha, ao menos os temas veiculados eram semelhantes entre os diversos canais, de forma que eram apresentadas diferentes perspectivas sobre as mesmas questões, baseadas em fatos, diferentemente das mídias digitais;

O principal objetivo das plataformas digitais é manter o usuário interessado e conectado pelo maior tempo possível, de forma que oferecem entretenimento. Assim, junto à capacidade de customização do usuário, há a atuação de algoritmos de hierarquização de conteúdo que geram as chamadas câmaras de eco e filtros-bolha. Ao se informar por meio das mídias digitais, desta forma, o usuário está exposto apenas, ou ao menos majoritariamente, a conteúdos que coadunam com suas perspectivas. Desta forma, enquanto o objetivo do jornalismo veiculado pelas mídias tradicionais é, em si, informar, nas mídias digitais a informação é veiculada sob a lógica do entretenimento, com o objetivo de manter o usuário interessado e feliz;

A facilidade e o baixo custo com que os conteúdos podem ser produzidos e veiculados nas mídias digitais contrasta com os altos custos de produção do jornalismo tradicional. Assim, jornais e partidos políticos, por exemplo, enquanto instituições que atuam permanentemente na construção da democracia, elaborando atividades complexas, foram enfraquecidos pela velocidade e leviandade com que os temas são pautados nas mídias digitais, bem como pela pluralidade de fontes não-confiáveis, dificultando a capacidade de discernimento entre o que é fato, opinião ou juízo de valor.

Desta forma, na sociabilidade pautada na comunicação digital há uma intensa fragmentação da comunicação social, que resulta em uma esfera pública fragmentada e desagregada e no desaparecimento ou enfraquecimento de grupos sociais tradicionais e organizações, substituídos por uma “coletividade emergência” que surge em ocorrências momentâneas sem que se perpetuem de maneira duradoura, impactando até mesmo a forma como o sistema político-partidário tradicional se organiza, implicando em uma fragmentação dentro dos próprios partidos, que passa a funcionar na mesma lógica das redes sociais (Vesting,

2019, p. 203-204). No Brasil, o movimento pode ser identificado com a eleição do Presidente Jair Bolsonaro que, primeiramente no PSL, durante as eleições presidenciais de 2018, e posteriormente no PL, para as eleições de 2022 e, independentemente de ter saído vitorioso na primeira campanha e derrotado na segunda, gerou um êxodo partidário de políticos vinculados ao bolsonarismo a esses partidos, dos quais muitos foram eleitos, alheios a qualquer confluência ideológica com a organização partidária em si;

Tanto nas mídias tradicionais, quanto nas mídias digitais, o modelo de negócio é pautado na veiculação de propagandas de anunciantes que buscam alcançar o seu público-alvo por meio dos espectadores/usuários das mídias. Nas mídias tradicionais, canais com maior confiabilidade do público angariavam públicos maiores, de forma que, pelo modelo de negócio, a produção de conteúdo melhor, pautado na ética profissional, jornalismo responsável e na autorregulação dos canais de mídia, implicava em maior valor à publicidade veiculada. Entretanto, nas mídias digitais, os conteúdos que geram mais interação não são, necessariamente, aqueles com maior rigor jornalístico/baseado em fatos, mas sim conteúdos que chocam ou são apelativos, chamando a atenção dos usuários, conforme discutimos no primeiro capítulo, bem como aqueles que coadunam com as perspectivas do usuário. Assim, pela arquitetura das mídias digitais e seu modelo de negócio, normas procedimentais que garantiam a qualidade da informação socialmente veiculada foram suprimidas pela busca incessante por mais interação com os materiais veiculados, independentemente de qual sua natureza. Todas essas características da comunicação digital implicam em um terreno fértil para a disseminação de *fake news* que, apesar de não ser um fenômeno recente, adquire novas características e alcance pelas mídias digitais (Horbach, 2022, p. 190-192).

Mediante essas características, podemos identificar que a esfera pública digital (fragmentada) é profundamente marcada pela rápida sucessão de assuntos que adquirem importância na comunicação digital, gerando uma espécie de cultura da presença, em que um momento de intensidade substitui o outro, impedindo ou dificultando a atividade de reflexão aprofundada. Ocorre, por meio da atuação dos algoritmos, a construção de um “mundo próprio singular”, voltado para cada um dos indivíduos, que favorece a formação de fóruns situacionais de pessoas que pensam da mesma forma, alimentando as chamadas câmaras-de-eco, propensas à disseminação de *fake news* e discurso de ódio. O sujeito da sociedade digital, assim, não detém elementos de subjetividade do *gentleman*, participante em comunidade e, tampouco, do sujeito da sociedade de massas, membro do grupo, mas constitui uma nova espécie de sujeito, o *homo digitalis*, cuja subjetividade é diluída na rápida sucessão de episódios e referências situacionais. O “eu” da cultura das redes é desprovido de uma continuidade e se caracteriza pela estruturação

de uma “identidade narrativa” que somente pode ser elaborada posteriormente e sobre suas relações com sua rede, ou seja, trata-se de um ser-relacional (Vesting, 2019, p. 204-205).

2.1.1. Democracia em cheque: repensar a liberdade de expressão na era digital

Se no surgimento da esfera pública liberal a centralidade era a visão do outro enquanto um espelho de si, resultando em uma visão comunitária, pautado nos direitos fundamentais individuais, e na esfera pública pluralista de grupos tal elemento foi transferido à pré-formação da vontade política pautada na discussão pública de interesses e opiniões, lastreada nos direitos fundamentais coletivos, enquanto a materialização do debate público, estruturando as instituições democráticas como estabelecidas hodiernamente, é preciso entender quais estratégias de normatização são adequadas para lidar com a realidade da esfera pública digital, que traz em si o excesso do eu, em um componente narcisista, e a profunda fragmentação do debate público. É necessário, inicialmente, que exista a construção de um conhecimento comum, por meio da comunicação digital, para o qual a sociedade possa confluir, enquanto pressuposto democrático, por meio de mecanismos de governança voltados à promoção de uma “auto-organização social e tecnológica abrangente” (Vesting, 2019, p. 207-208).

As instituições democráticas contemporâneas estão sujeitas a um anacronismo derivado da fragmentação do espaço público e o novo sujeito derivado da sociabilidade da comunicação digital, excessivamente voltado ao eu, em oposição à participação comunitária ou grupal. Ocorre, ainda, uma proteção jurídica irrefletida aos direitos individuais que não leva em consideração a transformação das subjetividades e da comunicação social, da qual decorre uma adaptação do princípio da liberdade de expressão e a ampliação generalizada do direito à autodeterminação digital (Vesting, 2019, p. 205-208).

Assim, uma das principais questões que se apresenta na discussão da regulação da comunicação digital se refere ao potencial risco ao princípio constitucional da liberdade de expressão. Nesse ponto, a partir de uma análise das possibilidades regulamentares no que diz respeito ao exercício da liberdade de expressão, por meio da dualidade proibição/permissão, decorrem nas tradições democráticas quatro possibilidades: proibição da ação; proibição da ação em geral, exigindo-se outorga estatal para contextos ou indivíduos em específico; permissão incondicional; e, por fim, permissão condicional, limitada pelos eventuais danos causados a terceiros, podendo ser feito controle prévio ou *a posteriori*. A proibição, a necessidade de outorga prévia e o controle prévio são, tradicionalmente, considerados supressão

ou restrição à liberdade de expressão, enquanto a liberdade mais ampla, encontrada na tradição estadunidense, encontra dificuldades quanto ao sopesamento com outros direitos fundamentais, como a proteção aos direitos da personalidade (Maranhão; Campos, 2019, p. 327).

Diversos países têm reconhecido que a disseminação irrefreada de *fake news* e discurso de ódio, especialmente por meio das mídias digitais, tem amplo potencial lesivo à democracia, de maneira a buscar mecanismos de combate a este tipo de conteúdo. Nos EUA, a interpretação conferida à Primeira Emenda à Constituição, no que diz respeito à liberdade de expressão, é bastante ampla, conferindo direito a, inclusive, mentir, como restou decidido no caso *United States v. Alvarez* em 2012. Assim, as medidas que podem ser tomadas pelo Poder Público no combate à disseminação de *fake news* são muito limitadas neste país. Em razão de tal limitação, a discussão na experiência estadunidense se concentra especialmente na ação não-governamental, como, por exemplo, por meio da criação de agências de *fact-checking*, oferecendo maior acesso a conteúdo verídico aos usuários das plataformas digitais. A despeito de medidas dessa natureza serem importantes, seu alcance e efetividade é limitado, tanto pelos altos custos relacionados ao funcionamento de agências de *fact-checking*, quanto a aspectos como a lentidão na produção de materiais de qualidade, que dificilmente acompanham a velocidade da vasta quantidade de *fake news* que são veiculadas, ou ainda pelas acusações de parcialidade das agências em si (Horbach, 2022, p. 195-196).

O mesmo se aplica a outras medidas propostas na experiência estadunidense como, por exemplo, a modificação dos algoritmos para indicarem conteúdos que potencialmente são *fake news*, o que implicaria em que as plataformas continuariam apontando quais conteúdos são acessados ou não; ou a educação midiática que, além de levar tempo, não tem eficácia garantida, uma vez que até mesmo pessoas altamente instruídas estão sujeitas a acreditar em *fake news*. Uma vez que a tradição jurídica estadunidense protege de maneira quase irrestrita a liberdade de expressão, qualquer medida que possa implicar em restrições ao princípio é descartada. Todas essas medidas levam em consideração, entretanto, que os agentes que produzem e compartilham conteúdos nas plataformas digitais estão agindo de boa-fé, como no caso de um usuário que acredita em determinado conteúdo e compartilha de maneira deliberada, ainda que se trate de uma mentira (Horbach, 2022, p. 196-197).

Já na tradição constitucional europeia, a interpretação do princípio da liberdade de expressão, também garantia fundamental, encontra sopesamento por meio de outros bens publicamente fundamentais. Na Convenção Europeia dos Direitos do Homem, o artigo 10, que trata da liberdade de expressão, traz um rol de limitações, que incluem a previsão do primeiro parágrafo da permissão para que se exija licenciamento de empresas de transmissão, televisão

ou cinema. É no parágrafo segundo deste artigo, entretanto, que está expresso de que maneira o princípio deve ser entendido: inicialmente, não se trata de uma liberdade irrestrita de expressão, mas de uma liberdade onerada por deveres e responsabilidade que devem se submeter a formalidades, condições, restrições e penalidades descritas pela lei e necessárias à sociedade democrática, à segurança nacional, integridade territorial, segurança pública, à prevenção da desordem e do crime, proteção da saúde e da moral, da reputação e dos direitos de terceiros, para prevenir a divulgação de informações privadas e confidenciais e para manter a autoridade e imparcialidade do Judiciário (Horbach, 2022, p. 197).

Desta feita, na Europa, medidas institucionais públicas mais restritivas, como a indisponibilização de conteúdos, podem ser tomadas, o que aumenta o rol de ferramentas à disposição no combate à disseminação de *fake news*. O maior exemplo de uma norma que institui um mecanismo destes é a lei alemã *Netz-DG* (e da recém aprovada *Digital Services Act*, pela União Europeia), que determina o estabelecimento de mecanismos de notificação e exclusão de conteúdos no prazo de 24 horas até 7 dias, para casos mais complexos, sob pena de responsabilização das plataformas digitais pelos conteúdos que contém *fake news*, discurso de ódio ou que são evidentemente ilegais veiculados nestas mídias. Para proteger a liberdade de expressão, a lei alemã obriga que as plataformas voltem a disponibilizar conteúdos que tenham sido indevidamente indisponibilizados. Além do caso alemão, outra previsão, jurisprudencial, da tradição europeia, é o direito ao esquecimento, fixado pelo caso *Google Spain vs. Gonzales* do Tribunal de Justiça da União Europeia (Horbach, 2022, p. 197-198).

O reconhecimento da necessidade de as plataformas digitais realizarem valores públicos e protegerem direitos de privacidade tem se expressado nas experiências normativas recentes. O direito ao esquecimento, ou o direito de ter determinado material retirado dos resultados de mecanismos de pesquisa, conforme definido pela Corte de Justiça da União Europeia, é um exemplo de como se tem buscado um meio termo entre a completa anarquia da informação e a censura, nesse caso balizado pelo interesse público de ter acesso à informação e o direito individual à privacidade. Assim, a informação/notícia, na esteira do quanto decidido pela CJUE, continua disponível, evitando a censura e garantindo o acesso à informação, todavia, ao evitar que qualquer pesquisa sobre determinado indivíduo na *internet* direcione diretamente para aquele material, possibilita-se, ao mesmo tempo, o direito à privacidade (Pasquale, 2017b, p. 12).

No Brasil, a Constituição Federal prevê, em seu art. 5º, IV e IX, a liberdade de manifestação do pensamento, sendo vedada a censura e necessidade de autorização prévia. Entretanto, destaca uma condicionante ao exercício da liberdade de expressão, ao dispor que é

vedado o anonimato. Tal limitação, que se consubstancia em uma condicionante ao exercício da liberdade de expressão, deve ser interpretada conjuntamente com a realidade da comunicação midiática predominante à época da aprovação da Carta de 1988, ou seja, a mídia de massa tradicional. Ao prever a vedação ao anonimato, a Constituição possibilita o funcionamento dos mecanismos de responsabilização, regulação profissional e ética da produção jornalística característica desta estrutura comunicacional. Assim, o texto constitucional garante a possibilidade de responsabilização pelas manifestações, por meio do direito de resposta ou indenização por dano material e à imagem, como prevê o inciso V do art. 5º, bem como estabelece contornos ao que não é considerado censura, ao prever a inviolabilidade dos direitos da personalidade, quais sejam a privacidade, intimidade, honra e imagem, no inciso X do mesmo artigo. Assim, a vedação ao anonimato não é absoluta, mas dela decorre, na verdade, um dever de “identificabilidade”, ou seja, da possibilidade de identificação *a posteriori* daquele que, ao expressar-se, causa danos a terceiros (Maranhão; Campos, 2019, p. 327-328).

O art. 220 da Constituição Federal, que trata da comunicação social, no mesmo sentido dos incisos do art. 5º da Carta Magna mencionados anteriormente, não permite qualquer possibilidade de censura ou necessidade de autorização prévia para a manifestação do pensamento. O dispositivo institui, ainda, uma especial proteção a conteúdo jornalístico, ao qual não é permitido opor-se nenhuma espécie de “embaraço”, conforme inteligência do §1º do art. 220, uma vez que as normas de qualificação profissional, de dever de sigilo profissional e da proteção aos direitos da personalidade, previstas no art. 5º, XIII, XIV e IV, V, X da Constituição Federal, respectivamente, já criam uma série de obrigações e normas de boas práticas jornalísticas que permitem, suficientemente, a responsabilização por eventuais excessos (Maranhão; Campos, 2019, p. 328).

No que diz respeito à veiculação de informações não jornalísticas, ao contrário, são permitidos certos “embaraços” à sua veiculação, como por exemplo a indicação de horários, locais e faixa etária para espetáculos públicos, devendo-se informar o Poder Público sobre a natureza do evento (art. 220, §1º), a obrigação de advertir sobre potenciais malefícios de produtos nocivos à saúde ou ao meio ambiente, bem como o dever de ser oportunizado aos indivíduos que se protejam de comunicação potencialmente nociva (art. 220, §3º). Ou seja, a Constituição Federal protege a liberdade de expressão, vedando a censura, ao mesmo tempo em que prevê mecanismos que possibilitam a responsabilização pela manifestação que seja ilegal, levando em consideração ainda a natureza da comunicação e as especificidades da comunicação

jornalística nas mídias de massa tradicionais, comunicação midiática prevalecente à época da promulgação da Constituição Federal de 1988 (Maranhão; Campos, 2019, p. 329).

A tradição constitucional brasileira se aproxima, neste sentido, muito mais da europeia que da estadunidense. O princípio da liberdade de expressão, conforme consagrado pela Constituição Federal de 1988, no art. 5º, IV e V, bem como a interpretação jurisprudencial dos Tribunais Superiores, do Supremo Tribunal Federal, do Superior Tribunal de Justiça e do Tribunal Superior Eleitoral, comporta limitações em razão de outros princípios e garantias constitucionais democráticos. Todavia, em razão de a maioria das grandes empresas de tecnologia, com milhões de usuários mundo afora, estarem sediadas nos EUA, a regulamentação das plataformas digitais para combater a disseminação de *fake news* e discurso de ódio tem sido apontada como uma violação ao princípio da liberdade de expressão, gerando uma tensão entre a tradição estadunidense de quase irrestrita liberdade e as tradições brasileira e europeia que impõem limitações. Nesse sentido, é importante destacar que, por maior que seja a empresa de tecnologia, e independentemente de onde está sediada, sua atuação deve estar subordinada ao regime legal da localidade em que oferecem seus serviços. O regime constitucional brasileiro, na mesma esteira da União Europeia e de países da União Europeia, como a Alemanha, permite limitações à liberdade de expressão, conferindo a instituições públicas, especialmente ao Judiciário, a legitimidade para atuar no combate à disseminação de *fake news* e discurso de ódio, no que diz respeito à honra, intimidade, imagem e a dignidade da pessoa humana (Horbach, 2022, p. 198-199).

2.1.2. A liberdade que limita: o princípio da liberdade de expressão tolhido pelo excesso de vigilância

O princípio da liberdade de expressão foi forjado originalmente para blindar os cidadãos da vigilância estatal. Todavia, o desenvolvimento de novas tecnologias de informação e comunicação alteraram profundamente as características da sociedade democrática contemporânea. De fato, entre o final dos anos 1990 até este início dos anos 2020, o desenvolvimento das ferramentas de comunicação digital resultou no domínio de empresas detentoras de plataformas digitais de abrangência multinacional que desafiam as fronteiras do Estado e que são pautadas no uso de algoritmos e inteligência artificial para determinar o conteúdo veiculado nesses meios, exercendo, por meio do uso dos dados dos usuários, uma nova espécie de vigilância que é privada. Portanto, se o elemento central do princípio da liberdade de expressão é a proteção dos indivíduos e da comunidade como um todo da

vigilância e censura, é necessário que sua interpretação passe a abranger também as novas formas de controle exercidas por empresas privadas, sem que se renuncie, entretanto, à proteção em face a intervenção estatal (Balkin, 2018, p. 1151-1152).

Por meio da captação e uso de dados, a comunicação digital passou por uma revolução dentro de si mesma, de forma que hoje está estabelecida uma verdadeira sociedade algorítmica, em que há novas possibilidades de controle, vigilância, discriminação e manipulação, tanto por parte de governos, como de empresas privadas. A detenção dos meios de comunicação por empresas privadas implica em que os seus mecanismos e estruturas, notadamente os algoritmos e a obtenção de dados, pautam o debate público, criando uma espécie de governança privada dos discursos. Tendo em vista o reconhecimento desta realidade, há recentemente o desenvolvimento de mecanismos de regulação estatais voltados às plataformas digitais que visam proteger a liberdade de expressão e o debate público. Coexistem, portanto, duas ameaças à liberdade de expressão, em contraposição à teoria tradicional que abrangia apenas a atuação estatal na supressão de liberdade, incluindo-se, agora, a capacidade de vigilância da *big data*. Assim, a interpretação tradicional conferida ao princípio da liberdade de expressão, bem como os mecanismos derivados desta interpretação, não têm o condão de proteger a vigilância promovida pelas empresas privadas, de forma que a interpretação tradicional do conceito de liberdade de expressão é inadequada para a atual estrutura de comunicação social baseada nas mídias digitais, uma vez que esta não leva em consideração, e sequer poderia levar, a capacidade de uso de dados e algoritmos para moldar o debate público (Balkin, 2018, p. 1153-1154).

Mesmo que sem o reconhecimento do exato funcionamento das plataformas, determinados conteúdos deveriam ser impedidos de ser veiculados, independentemente de qualquer outro fator, como por exemplo conteúdo nazista ou que incentive genocídio. Apesar de o impedimento de determinadas manifestações de expressão ser constitucionalmente vedado na tradição estadunidense, pautado em uma espécie de fundamentalismo da liberdade de expressão, protegendo até mesmo o discurso mais repreensível, diversas outras tradições constitucionais, como a da França e da Alemanha, admitem tal possibilidade. Estender o princípio da liberdade de expressão para abranger, também, a capacidade das plataformas direcionarem conteúdo, através da captação de dados dos usuários e *microtargeting*, significa proteger o poder a organizações com capacidade econômica e organizacional de contratar os melhores programadores e elaborar os melhores programas para influenciar o debate público, conferindo especial proteção estatal a corporações que, muitas vezes, cooptam e subordinam ao próprio Estado. Tal extensão implica, na verdade, em um prejuízo aos valores

clássicos basilares do princípio da liberdade de expressão, notadamente a autonomia e a autodeterminação democrática (Pasquale, 2017b, p. 9-10).

A sociedade algorítmica se desenvolveu sobre as bases da sociedade digital, em que cada vez mais aspectos da vida cotidiana passaram a deixar marcas no ambiente digital, possibilitando a coleta e tratamento massivos de dados dos usuários e gerando a *big data*, composto por dados coletados e metadados, ou seja, dados sobre os dados, ou análises sobre os dados levantados. Os algoritmos e a inteligência artificial são os mecanismos dessa sociedade, enquanto os dados são seu combustível. O objetivo da sociedade algorítmica é a onisciência prática, ou seja, saber ao máximo possível o que, quando e onde as pessoas estão, e prever o mesmo sobre onde estarão. Essa possibilidade, inédita, implica na criação de uma nova economia do poder. Nessa nova estrutura, as fontes de poder são distribuídas, uma vez que estão pautadas nos dados gerados pelos usuários em seus aparelhos eletrônicos conectados à *internet*, porém a capacidade de utilizar este poder não é democrática, uma vez que apenas algumas poucas empresas são capazes de coletar, analisar e usar esses dados (Balkin, 2018, p. 1154-1157).

As novas tecnologias de comunicação digital e de análise de dados são os instrumentos que tornaram possível uma nova espécie de predição sobre o comportamento humano e de controle, arregimentadas por alguns agentes, notadamente governos, empresas ou algum grupo de influenciadores, para estabelecer relações de poder entre as pessoas de quem os dados são coletados e os agentes que tomam decisões com base nesses dados. Sem a possibilidade de livremente coletar os dados dos usuários, transmiti-los e analisá-los, a estrutura de poder derivada da sociedade algorítmica colapsa. É por isso que os detentores desses mecanismos de poder buscam proteger, a todo custo, a liberdade de expressão: não se trata de coadunar com uma perspectiva democrática de participação em comunidade, mas sim de garantir que os dados serão compartilhados e coletados para manter a estrutura de poder em funcionamento. Portanto, a interpretação que sugere a aplicação do princípio da liberdade de expressão sobre as atividades desenvolvidas pelas plataformas digitais diz respeito ao entendimento de se esta abrange o direito de estabelecer, por meio do uso de dados, algoritmos e inteligência artificial, relações de poder (Balkin, 2018, p. 1157-1160), e ainda promove-las por meio da proteção estatal contra a responsabilização das plataformas pelos prejuízos, individuais e coletivos, que possam derivar da comunicação digital.

A relação estabelecida entre as empresas e os usuários é assimétrica, uma vez que o usuário é transparente com a plataforma quando a utiliza, compartilhando seus dados, enquanto as plataformas não são tão transparentes com os usuários sobre a forma como esses dados serão

utilizados e, tendo em vista a complexidade do funcionamento de algoritmos e inteligência artificial, sequer seria possível estabelecer uma relação de transparência entre usuário e plataforma plenamente satisfatória, tendo em vista o desconhecimento dos usuários em geral das questões técnicas que envolvem o tema. Os usuários estabelecem, portanto, uma relação de confiança com as plataformas, acreditando que seus dados serão utilizados em seu benefício e não para prejudicá-lo de alguma maneira. Esta relação é muito semelhante com aquela estabelecida com médicos ou advogados, por exemplo: o cliente compartilha a informação porque precisa de um serviço para o qual não tem a estrutura técnica e, ao confiá-las ao profissional, este está subordinado ao cumprimento dos interesses dos clientes, não podendo usar essa informação para criar conflitos de interesse ou desvantagens a este. Se as plataformas são fiduciárias das informações (dados) dos usuários, em analogia à atividade médica ou advocatícia, essas não constituem parte do debate público em si, mas sim parte de uma relação de confiança usuário-empresa que pode ser regulamentada democraticamente, especialmente para proteger a parte hipossuficiente (Balkin, 2018, p. 1160-1161).

Essas relações assimétricas são pautadas na capacidade técnica das plataformas de lidar com os dados dos usuários e na necessidade dos usuários, para poderem utilizar os serviços oferecidos pelas plataformas, compartilharem suas informações. A atividade desenvolvida pelas plataformas digitais envolve níveis diferentes de responsabilidade sobre os dados dos usuários e os próprios dados têm papéis distintos a depender do serviço oferecido. A relação de confiança entre as plataformas e usuários é caracterizada: 1) pela monetização dos dados dos usuários, que é elemento central do funcionamento da maior parte das plataformas digitais; 2) pelo fato de que quanto mais dados forem compartilhados, maior a possibilidade de lucro das plataformas digitais, de forma que estas criam incentivos para que seus usuários compartilhem dados; 3) por, diferentemente da relação entre cliente e médico, em que há deveres positivos, ou seja, o dever de cuidado do médico com o cliente, na relação entre os usuários e plataformas não há essa expectativa, mas apenas que os dados não sejam utilizados para gerar malefícios aos usuários (Balkin, 2018, p. 1161-1163).

A novidade da sociedade algorítmica, portanto, enquanto um avanço na sociedade digital, é o uso abrangente de mecanismos de captação e processamento de dados dos usuários para influenciar a forma como os indivíduos agem. Esses dados, provenientes do uso de diferentes serviços de informação e comunicação digital, abrangem o uso das plataformas digitais mais comuns, como redes sociais ou mecanismos de busca, mas também dados do *smartphone* como localização, ou até mesmo informações sobre o cotidiano, por meio de assessorios inteligentes, que abrangem desde torradeiras conectadas à *internet* e robôs de

limpeza até assistentes pessoais digitais como a *Alexa*. Com a posse de um perfil tão detalhado dos indivíduos, as empresas são fiduciárias das informações fornecidas pelos usuários, de forma que os prejuízos gerados pelo uso inadequado destas informações deve ser responsabilidade destas (Balkin, 2018, p. 1168-1171).

A estrutura de poder do sistema de comunicação, nesse sentido, foi modificada pela inserção da comunicação digital, guiada pelo seu modelo de negócio. A instituição de mecanismos de moderação de conteúdo deve levar em consideração as relações de poder exercidas entre os sujeitos de acordo com a posição econômica e as tecnologias de comunicação e informação vigentes. A comunicação por meio das plataformas digitais, diferentemente da mídia tradicional, não passa por uma espécie de “aprovação prévia” feita pelos próprios meios de comunicação por meio do processo de curadoria e, além disso, é necessário reconhecer que não é mais caracterizada por ser exercida em um espaço público, mas sim por ser realizada em um “local” virtual cuja propriedade é particular, sendo estruturado por interesses particulares (Hartmann, 2020, p. 4-5).

2.1.3. Regulamentando a tecnologia: um desvio de rota pautado na justiça

Nesse sentido, uma das principais dificuldades na busca por uma justiça aplicável à inteligência artificial diz respeito à posição que as teorias políticas tradicionais relegam às tecnologias. Na teoria da justiça de John Rawls, apesar de as características socio-tecnológicas específicas serem consideradas para determinar critérios de justiça, como o funcionamento do sistema legal, a divisão econômica do trabalho, a capacidade de taxação, entre outros, o nível de tecnologia é excluído, o que é um equívoco. Já na teoria marxista do materialismo histórico, as condições materiais tecnológicas são reconhecidas como fatores determinantes para a estruturação do modo de produção, ou seja, das relações econômicas prevalentes, que forjam a superestrutura da sociedade, inclusive as leis e os sistemas de crença, porém, ao contrário da teoria de Rawls, a moral e as escolhas humanas têm seu impacto subdimensionado (Gabriel, 2022, p. 218-219).

Desta forma, retomando a discussão do primeiro capítulo, é fundamental compreender que a relação entre seres humanos e tecnologia é um processo de dois sentidos: as tecnologias moldam a maneira como nos organizamos socialmente, possibilitando, nas sociedades modernas, o controle centralizado, organizações em maior escala, a especialização econômica e determinando os objetivos, aspirações e autoconhecimento dos cidadãos; e, por outro lado, as

próprias tecnologias são desenvolvidas mediante um processo de *design*, experimentação, desenvolvimento, iteração e adoção, de forma que o poder de guiar este processo resulta nas novas estruturas que estão sendo e serão adotadas. Assim, as decisões tomadas pelos agentes dotados de poder para guiar a produção de novas tecnologias, ou seja, de tecnólogos, mercados de tecnologia, estruturas de governança e normas sociais, implicam em impactos profundos na organização social. Desta forma, mediante o reconhecimento destes dois sentidos do desenvolvimento tecnológico, fica evidente a necessidade de identificá-lo enquanto um importante objeto de valoração moral, conforme a teoria da justiça de Rawls, ao mesmo tempo em que se identifica o impacto material das tecnologias na organização social (Gabriel, 2022, p. 219).

A teoria política liberal, conforme proposta por Rawls, destaca que a constituição política, os regimes legais de propriedade, a organização da economia e a natureza da organização familiar constituem a estrutura básica da sociedade, a qual está sujeita aos princípios de justiça distributiva, pautada em uma determinada estrutura normativa que busque propiciar igualdade. Ou seja, outros fatores além destas estruturas podem diferenciar a posição social ocupada pelos indivíduos, o que decorre das diferenças inerentes à pluralidade humana, entretanto esses componentes da estrutura básica devem ser pautados em princípios de justiça distributiva, sob pena de que, caso sejam deixados sem interferência, gerem situações de iniquidade que implicariam em que aos sujeitos menos favorecidos não restaria outra opção que não seja aceitar o arranjo proposto pelos ricos e poderosos (Gabriel, 2022, p. 220-222).

A estrutura básica social não deriva apenas destes sistemas apontados por Rawls, mas também de sua interação com as tecnologias vigentes em um determinado tempo e espaço e, nesse sentido, as inteligências artificiais, cada vez mais presentes, conseqüentemente, exercem um papel fundamental na maneira como devemos compreender os princípios de justiça distributiva. Nesse aspecto, inteligências artificiais têm influenciado de maneira decisiva nas oportunidades oferecidas aos cidadãos, por exemplo por meio de sistemas financeiros que oportunizam empréstimos ou seguros de acordo com os perfis dos indivíduos ou mesmo a disponibilização de serviços públicos, como na disponibilização de moradias populares, ou, ainda, nos usos das inteligências artificiais da justiça criminal, as quais podem reproduzir -e agravar- vieses de raça. As inteligências artificiais, assim, têm forjado uma verdadeira “reputação digital” dos cidadãos, que definem possibilidades que lhe serão oferecidas dentro de determinada estrutura social, de maneira a influenciar decisivamente nas oportunidades dos indivíduos, de forma que a atuação destas tecnologias deve ser, também, objeto de princípios de justiça distributiva (Gabriel, 2022, p. 220-222).

Uma vez que as opções feitas pelas inteligências artificiais, assim como os outros elementos que constituem a estrutura básica da sociedade, implicam em impactos na vida dos cidadãos, a forma como os sistemas funcionam deveria ser objeto de escrutínio público. A publicidade dos fundamentos da estrutura social básica, assim, é pressuposto de um sistema de justiça distributiva, uma vez que possibilita a crítica pública dos elementos que guiam as possibilidades de vida, do que deriva uma obrigação aos desenvolvedores de tecnologias de inteligência artificial de explicar seus modelos, de forma que possam ser objeto de um debate público informado. Desta forma, a adoção de tecnologias em determinados contextos não pode ser justificada apenas por interesses privados, mas deve ser guiada, também, pelo interesse público (Gabriel, 2022, p. 222-223).

As liberdades básicas do sistema de justiça de Rawls são a liberdade de pensamento e consciência, liberdades políticas e de associação, assim como a integridade pessoal e os direitos e liberdades derivados do Estado de Direito. Eventuais distorções produzidas por algoritmos, pautadas em vieses, na avaliação dos perfis digitais, podem gerar prejuízos a estes princípios, afetando a igualdade da cidadania e deixando de ser imparcial. Assim, emerge da nova estrutura tecnológica uma nova necessidade por privacidade, para proteger a sociedade de formas de vigilância invasivas ou da manipulação do comportamento (Gabriel, 2022, p. 223-224).

Já no que diz respeito à justa igualdade de oportunidade, o sistema de justiça distributiva de John Rawls pressupõe não apenas uma igualdade formal, mas também que existam medidas corretivas que propiciem um “desenviesamento”, ou seja, uma estrutura que promova igualdade material, superando aspectos de classe, raça, sexo etc. Outro princípio da justiça distributiva de Rawls é conhecido como princípio da diferença, em que as diferenças de distribuição de bens sociais primários se dá tendo em vista conferir a melhor condição possível ao menos favorecido. Este princípio, à primeira vista, se aplica melhor a situações econômicas, em relação a bens materiais, portanto, mas Rawls já destacava a importância das “bases sociais do autorrespeito”, dando grande importância à autoestima. É importante que as estratégias de equidade não estejam restritas a redistribuição de bens sociais após a geração de iniquidades, mas que os próprios fundamentos do sistema sejam forjados de maneira a produzir equidade, diminuindo as hipóteses de iniquidade que poderiam surgir posteriormente. Considerar as condições socio-tecnológicas enquanto uma das estruturas básicas da sociedade permite, assim, visualizar possibilidades de promoção de igualdade substancial, ao evidenciar situações de desigualdade que são derivadas da própria tecnologia, além de exigir que o desenvolvimento das tecnologias seja objeto de justificação pública, promova os direitos civis e incorporem princípios de justiça distributiva (Gabriel, 2022, p. 225-227).

Algoritmos e inteligências artificiais atuam de acordo com o seu desenvolvimento e programação, ou seja, pautado nos materiais a que foi exposto, em como interagem o *software*, o *hardware* e o ambiente externo, o que resulta no comportamento apresentado. Desta forma, estes robôs reproduzem as escolhas feitas por seus criadores. Nesse sentido, a estratégia para regulamentar que tem sido cada vez mais adotada, como no caso da lei alemã e da União Europeia, é de regular não as tecnologias em si, mas sim seus criadores e proprietários, responsáveis pela forma como funcionam. Para que se adote um sistema eficiente, levando em consideração que as inteligências artificiais e algoritmos não são dotadas de significado e ação, é preciso que se impute aos desenvolvedores não apenas a responsabilidade pelos eventuais resultados negativos derivados do uso das plataformas, mas uma espécie de responsabilidade pelo *design* da plataforma, de forma a influenciar no desenvolvimento dos sistemas, barrando certas opções de *design* e encorajando outras (Pasquale, 2017a, p. 1252-1255), afeitas à democracia.

Assim, quando as inteligências artificiais estão integradas ao funcionamento das grandes instituições e práticas sociais, constituem parte da estrutura básica social, de forma que devem ser objeto dos critérios de justiça distributiva, mediante o impacto que causam na vida dos cidadãos, especialmente nas grandes instituições políticas e econômicas. No que diz respeito a mecanismos de busca ou sistemas de recomendação, como os que existem em redes sociais, há uma integração entre a tecnologia e práticas sociais epistemológicas prevaletentes. Ou seja, por meio destes sistemas há uma espécie de curadoria que guia o conhecimento, ou aquilo que é entendido enquanto verdadeiro, no caso de mecanismos de buscas, e que influenciam o desenvolvimento da moral, mudando a autopercepção, os desejos e as preferências, como ocorre especialmente nas redes sociais (Gabriel, 2022, p. 227-228).

Assim, ocorre uma exploração e dominação interpessoal cuja legitimidade é, no mínimo, questionável. Ainda que estas tecnologias não venham a ser consideradas parte da estrutura básica social, portanto estritamente sujeitas aos princípios de justiça distributiva, não estão isentas do imperativo de serem guiadas por um dever de justiça, para corroborar com as instituições básicas, e no caso do desenvolvimento de novas tecnologias, esse imperativo deve incidir como um dever de desenvolvimento para apoiar o funcionamento de instituições justas, e não para subvertê-las. Ademais, a justificação pública, ou seja, a possibilidade de a sociedade de maneira geral compreender o funcionamento de um determinado sistema tecnológico e influenciar nas decisões tomadas em relação a este, se fundamentaria até mesmo apenas pelo reconhecimento do impacto que as tecnologias causam na sociedade, sob pena de ceder demasiadamente o controle social à instituições privadas, ferindo a igualdade e liberdade,

portanto trata-se de uma maneira de se opor a formas de imposição tecnológicas injustas (Gabriel, 2022, p. 227-228).

2.2. Prática institucional: a governança digital no início dos anos 2020

As intermediárias *online*, incluídas nesse conceito diversos provedores de programas e serviços *online*, tais como mecanismos de pesquisa, redes sociais e aplicativos, têm grande relevância na vida contemporânea, representando instituições de grande alcance e impacto social e, conseqüentemente, são objeto de discussão política. Assim, na intersecção entre lei, tecnologia, normas e mercados, diversos países têm criado formas de governança das intermediárias que são estruturadas de acordo com as experiências locais. É importante destacar que a terminologia escolhida em cada país para determinar o que é designado intermediárias *online*⁶ é bastante diversa (Gasser; Schultz, 2015, p. 1).

A dificuldade terminológica em torno dos serviços prestados pelas chamadas intermediárias *online* explicita as características deste tipo de serviço que, por ser novo e bastante cambiante, engloba diversas perspectivas distintas. Dentro do direito, por exemplo, a concepção varia de acordo com a perspectiva adotada. Por exemplo, se o foco é a regulação da mídia, a concepção é mais relacionada ao impacto do serviço na opinião pública; se o foco for o regime de responsabilidade das intermediárias, estará mais relacionado à capacidade técnica de retirar conteúdos e no controle que as plataformas têm sobre o conteúdo postado por terceiros; se o foco for a liberdade de expressão, privilegia-se a perspectiva segundo a qual as intermediárias *online* criam um novo espaço para expressão individual *online*. O uso do termo intermediárias *online* é um esforço de determinação desse tipo de serviço, por meio da generalização, contendo dentro da definição diferentes serviços e perspectivas, em busca de diferenciar estas mídias digitais de outras formas de mídia tradicionais. Assim, podem ser consideradas intermediárias *online*: 1) plataformas que atuem “entre” o conteúdo e os usuários; 2) que apresentam relevância para a comunicação pública, não sendo meramente um instrumento privado; 3) que não são meios de comunicação tradicional. As principais plataformas digitais que contêm tais características são mecanismos de pesquisa, micro blogs,

⁶ No esforço de unificar a terminologia utilizada para designar os serviços oferecidos pelas plataformas digitais, a OECD- Organização para a Cooperação e Desenvolvimento Econômico (2010, p.9) - explica que o termo “*online intermediaries*” diz respeito a plataformas que facilitam a transação entre terceiros na *internet*, dando acesso, hospedando, transmitindo ou *indexando* conteúdo, produtos ou serviços originados por terceiros na *internet* ou provendo serviços de *internet* para terceiros.

redes sociais, e plataformas de conteúdos postados pelos usuários (Gasser; Schultz, 2015, p. 2-3).

A abordagem adotada pelo Direito difere de acordo com as características sociais de cada país ou região que, por suas especificidades, colocam em evidência diferentes fenômenos que guardam relação com as intermediárias *online*. Assim, na Índia, por exemplo, a legislação inclui no conceito de intermediárias *online* provedores de *internet* e serviços de telecomunicação e, na Turquia, plataformas de *e-commerce*, as quais representam importante papel na discussão sobre a responsabilidade das intermediárias no país.

A forma como a regulamentação das intermediárias *online* se dá em diferentes países ou regiões, portanto, está intimamente relacionada com as plataformas que assumem maior relevância social e econômica. Assim, tema caro à União Europeia tem sido os mecanismos de pesquisa, tendo em vista a relevância social da discussão sobre o direito ao esquecimento, chamado de direito de ser excluído das listas de pesquisa, que deriva da popularidade do *Google*, pautado não necessariamente na exclusão do próprio conteúdo prejudicial da *internet*, mas sim da exclusão dos resultados nos mecanismos de pesquisa, limitando o alcance dos conteúdos. Nos Estados Unidos da América, o foco da legislação tem sido plataformas de criação de conteúdo pelos usuários. Desta forma, destaca-se a diversidade de abordagens na normatização do tema de acordo com as especificidades de cada experiência (Gasser; Schultz, 2015, p. 3-4).

De maneira geral, entretanto, na tentativa de abranger a maior parte de serviços promovidos pelas intermediárias *online*, as legislações têm adotado definições genéricas e abstratas, como no caso da legislação estadunidense, por meio da *U.S. safe harbor*, ou no caso da legislação brasileira, pelo Marco Civil da Internet, em que ambos descrevem ações abstratas para definir quais são as intermediárias *online*, ou no caso da *European E-Commerce Directive* que elabora uma definição híbrida, definindo protótipos abstratos tais como provedores de hospedagem, provedores de acesso, etc. Devido à natureza volátil do ecossistema das intermediárias *online*, a lei, os legisladores, reguladores e os Tribunais deparam-se, regularmente, com as dificuldades de definição quanto às funções exercidas pelas intermediárias *online*, bem como com os possíveis efeitos de intervenções nas distintas áreas de atuação destas, dificultando soluções duradouras, exigindo e promovendo novos debates institucionais (Gasser; Schultz, 2015, p. 4).

Os sistemas de governança das intermediárias *online* podem ser divididos em duas categorias abrangentes: i) aqueles em que há previsões legais especificamente voltadas para as intermediárias *online*; ii) aqueles em que as previsões legais gerais são aplicadas às

intermediárias *online*. Um dos casos que deixam mais explícitas as diferenças entre os dois sistemas é o do Brasil. Antes da entrada em vigor do Marco Civil da Internet, não havia previsões específicas para as intermediárias *online*, o que foi modificado após a aprovação da referida legislação. Assim, no que diz respeito à responsabilização das intermediárias *online*, antes da entrada em vigor da norma a jurisprudência subdividia-se em três correntes distintas: a primeira eximia a intermediária *online* de qualquer responsabilidade pelo conteúdo postado por terceiros; a segunda instituía um regime de responsabilidade objetiva, tendo em vista o conceito de risco da atividade desenvolvida pelas intermediárias *online*; e a última previa a possibilidade de responsabilização mediante aferição de culpa. O Marco Civil da Internet instituiu previsão específica quanto à responsabilização das intermediárias *online*, especialmente no art. 19 da referida lei (Gasser; Schultz, 2015, p. 4-5).

Os Estados Unidos da América são pioneiros na instituição de normas específicas destinadas às intermediárias *online*, criando um sistema de proteção às intermediárias referidos na literatura como *safe harbor*, ou seja, porto seguro para as plataformas. As normas estadunidenses sobre o tema são a *Section 230 da Communications Decency Act* e a *Section 512 of the Digital Millennium Copyright Act*. O mesmo ocorria na União Europeia, por meio da *E-Commerce Directive*, que institui normas de proteção das intermediárias *online* e trata como excepcionais as possibilidades de responsabilização destas que, entretanto, teve o panorama radicalmente modificado pelo *Digital Services Act*, aprovado em 2019, na esteira do reconhecimento internacional da necessidade de regulamentação das mídias digitais.

Nesse panorama, os Estados Unidos, por serem a sede da maior parte das *big techs*, guiam, por meio da sua normatização e princípios legais, as principais regras aplicadas pelas plataformas digitais (como vimos, esse processo vem sendo relativizado ao longo dos últimos anos, com a inserção de empresas de outras origens, porém a própria estrutura comunicacional e principalmente mercadológica segue semelhante). O conceito de liberdade de expressão neste país é um dos mais liberais do mundo, abrangendo mais temas e conteúdos que podem ser veiculados ali do que em outras localidades, inclusive no Brasil, gerando tensões especialmente no que diz respeito a questões de moderação de conteúdo (Macedo Júnior, 2020, p. 235-236).

É importante destacar que os tipos de conflitos que existem nas diversas experiências locais resultam em sistemas de governança e responsabilização distintos. Por exemplo, enquanto no Vietnã, Tailândia e Índia grande parte dos grandes conflitos envolvendo as intermediárias *online* são a exclusão de conteúdos por solicitação do Estado, por exemplo pela “proteção da honra do rei da Tailândia”, na maioria dos demais países, especialmente aqueles com maior tradição democrática, as lides envolviam problemas entre usuários. Tal fato implica

na necessidade de reconhecer-se que, por mais que seja um desafio transnacional, a depender da extensão dos direitos individuais e da cultura de cada localidade, sistemas de governança distintos podem ser instituídos. É importante destacar, ainda, que no que diz respeito a questões de direitos autorais, a maioria dos sistemas de governança têm disposições específicas, distintas do mecanismo de responsabilização das intermediárias *online* em outros temas (Gasser; Schultz, 2015, p. 5), em função de *lobby* de setores econômicos.

Outra distinção entre os sistemas de governança é se a legislação protege as intermediárias *online* da responsabilização. O maior exemplo desta proteção, conforme supramencionado, é dos Estados Unidos da América, em que a *Section 512 of the Digital Millennium Copyright Act* protege as plataformas da responsabilização, sob a justificativa de que, se não há curadoria quanto ao material que é publicado, não deve haver responsabilização das plataformas. No mesmo sentido, a legislação indiana também protege da responsabilização as intermediárias que não exercem controle editorial. Ademais, a proteção à responsabilização das intermediárias *online* pode ter diferentes alcances, de forma a protegê-las da responsabilização em todos os casos ou apenas em casos específicos legalmente previstos (Gasser; Schultz, 2015, p. 6).

Todavia, conforme analisamos no primeiro capítulo, as plataformas digitais, como as redes sociais *Facebook* e *Twitter* e mecanismos de busca como o *Google*, dificilmente podem ser considerados isentos em relação à forma como o conteúdo é disseminado. Desta feita, a linha de distinção entre o que pode ser considerado curadoria e o que não é mostra-se cada vez mais tênue, revelando-se um aspecto fundamental nos debates sobre a regulamentação das plataformas digitais.

Outra característica que diferencia os sistemas de governança das intermediárias *online* é se este estabelece um “procedimento de notificação e exclusão” ou não, bem como se a existência deste procedimento é uma opção para a proteção da responsabilização das intermediárias, ou se é um mecanismo obrigatório, de forma que a falha na utilização do procedimento implica em responsabilização. Destacam-se dois sistemas que preveem tal procedimento, o primeiro de forma explícita, na Índia, e o segundo que foi construído de maneira implícita, na União Europeia. O caso que resultou na criação desta exigência implícita se deu na *Court of Justice of the European Union*, no caso *Google Spain SL, Google Inc. versus Agencia Española De Protección De Datos* (AEPD), em que a decisão considerou que a solicitação de retirada de um *link* dos resultados de buscas em mecanismos de pesquisa está associada ao direito ao esquecimento, de forma que estabelecer um “procedimento de

notificação e exclusão” é condição para a proteção da responsabilização das intermediárias *online*, sem a qual podem ser responsabilizadas (Gasser; Schultz, 2015, p. 6).

Os sistemas de governança das intermediárias *online* são estruturados, na maior parte dos casos, por meio de regulação indireta, ou seja, não se proíbe ou permite algum comportamento específico às intermediárias *online*, mas, ao contrário, estabelecem-se critérios segundo os quais estas serão responsabilizadas ou não pelos conteúdos postados por terceiros. Um exemplo desse sistema é o supracitado “procedimento de notificação e exclusão” de conteúdo em conflitos entre usuários nas plataformas. Já o outro sistema de governança, menos utilizado (e não democrático), estabelece normas que regulam diretamente as intermediárias *online*, como é o caso da Tailândia, em que há um regime que exige licenciamento prévio das plataformas (Gasser; Schultz, 2015, p. 6-7).

Os novos mecanismos de regulação do discurso têm como característica a cooperação ou cooptação público/privado, por meio da qual o Estado, que não é dotado da estrutura técnica das plataformas digitais, cria incentivos e obrigações para que os agentes privados exerçam regulação, podendo ser por meio de bloqueios e filtros de conteúdo *ex ante*, exclusão de conteúdo *ex post*, vigilância sobre os usuários ou provendo informações sobre os usuários. Essa relação pode variar em estrutura, por meio de regulação direta, ameaças, sugestões de que as plataformas serão mais bem vistas pelo Estado se cooperarem, ou negociações sobre os termos da cooperação. A depender de cada caso, as empresas podem cooperar deliberadamente ou serem cooptadas para fazê-lo.

Assim, a capacidade de vigilância que as plataformas digitais desenvolveram possibilitam mecanismos de governança e restrição de conteúdos inéditos, de forma que os Estados buscam valer-se destas habilidades para exercer a regulação de conteúdo. De fato, as próprias empresas criam mecanismos de governança privada por conta própria, seja para obrigar os usuários a cumprirem com seus termos de serviço; para manter seus usuários felizes e atrair novos usuários; para garantir aos seus parceiros comerciais que são capazes de localizar, filtrar, bloquear e remover conteúdo que não condiz com o interesse da plataforma; ou até mesmo para resolver disputas entre os usuários. Assim, as plataformas digitais são, em maior ou menor grau, não apenas empresas de tecnologia e de mídia em si, mas também empresas de governança privada, independentemente de incentivos legais para tanto. Ou seja, as plataformas digitais, em especial aquelas que têm mais usuários e atuam em diversos países, são objeto de pressão de regulação de várias normas nacionais e internacionais sobre o tema e constituem um verdadeiro sistema de governança transnacional e privado do discurso (Balkin, 2018, p. 1179-1182).

É importante destacar, ainda, que as estruturas dos sistemas de responsabilização das intermediárias *online* atualmente existentes são, raramente, resultado de uma estratégia regulatória, mas resultado da interação entre diferentes ações desconexas. O caso do *Google* na Corte de Justiça da União Europeia, supramencionado, que envolve o tema da exclusão de resultados de buscas em mecanismos de pesquisa, evidencia o fato de que, por meio de uma decisão pontual sobre a responsabilização da intermediária *online*, pode-se criar uma exigência sem que esta seja, necessariamente, resultado de um planejamento. Neste caso, a Corte estabeleceu, ainda que indiretamente, por meio de sua decisão, a competência para a solução de conflitos entre usuários para a plataforma, no caso o *Google*, uma vez que, para evitar a responsabilização, este deve estabelecer um procedimento de notificação de exclusão e, assim, sopesar o direito do usuário de ter sua intimidade protegida e o interesse da pessoa que postou o conteúdo, bem como o interesse coletivo de acessar determinada informação sem que, entretanto, tenham-se estabelecidos critérios legais para tanto (Gasser; Schultz, 2015, p. 7-8).

Evidencia-se, portanto, que a atuação das Cortes, por meio dos julgamentos, compõe o sistema de governança das intermediárias *online*, criando incentivos e desincentivos para a atuação do agente privado. Pela natureza do Judiciário, entretanto, antever tais reflexos de suas decisões é dificultoso, tornando-se necessário, no mínimo, que as decisões tomadas pelas Cortes sejam coerentes e uniformes. No caso brasileiro, antes da aprovação do Marco Civil da Internet, as Cortes aplicavam três correntes distintas, de forma a não unificar as decisões provenientes do Judiciário, tornando difícil a tomada de decisão das plataformas, que não sabem como devem agir diante da ausência de previsão legal específica (Gasser; Schultz, 2015, p. 8). De fato, mesmo após a aprovação do Marco Civil da Internet, os Tribunais brasileiros não unificaram a jurisprudência sobre o tema, conforme verificamos pelas diferentes fundamentações de decisões proferidas pelo Tribunal Superior Eleitoral mencionadas no primeiro capítulo, bem como pela pendente decisão sobre a constitucionalidade do art. 19 do Marco Civil da Internet, que prevê o sistema de isenção de responsabilização das plataformas digitais sobre o conteúdo postado por terceiros, sob julgamento no Supremo Tribunal Federal no Recurso Extraordinário nº 1037396/SP, Tema de Repercussão Geral 987, e no Recurso Extraordinário nº 1057258/RJ, Tema de Repercussão Geral 533.

Há, ainda, dois aspectos relevantes no que diz respeito aos procedimentos de notificação e exclusão. O primeiro é que os modelos variam entre aqueles em que não há qualquer previsão legal sobre a forma como serão realizados tais procedimentos até aqueles em que os procedimentos estão bem estabelecidos pela norma, como o caso da Índia. O outro aspecto é que raramente há a previsão da necessidade de informar previamente o usuário da exclusão do

conteúdo ou há a previsão de um procedimento de recurso para o usuário (Gasser; Schultz, 2015, p. 8-9).

Quanto à forma como as intermediárias *online* lidam com a exclusão de conteúdos, há três sistemas que são aplicados: i) sistemas automáticos, como o *ContentID* do *YouTube*, que identifica automaticamente violações ao direito de propriedade; ii) sistemas manuais, como o adotado pelo *Google* após o caso na Corte de Justiça da União Europeia; iii) sistemas híbridos, em que a maior parte dos casos são resolvidos por sistemas automáticos, mas aqueles mais complexos são resolvidos por sistemas manuais (Gasser; Schultz, 2015, p. 9). O Facebook, conforme será analisado no Capítulo 3, utiliza de sistemas híbridos para lidar com a moderação de conteúdos postados por terceiros, seja em função do cumprimento de políticas da própria empresa, seja em cumprimento a solicitações de terceiros como Governos, Tribunais, organizações privadas etc.

Juntamente com fatores tecnológicos e de mercado, a atuação estatal é um dos principais elementos que compõe a estrutura de funcionamento das intermediárias *online*. O Estado, nesse sentido, atua de maneira ampla, desde sua participação como usuário das intermediárias até o seu papel de regulador. Apesar da importância do Estado na regulação das intermediárias ser uma constante, a forma como este papel é desempenhado difere de acordo com cada experiência, variando em modelo e estratégia, de forma que em alguns sistemas a regulação é centralizada em agências governamentais, enquanto em outros as Cortes de Justiça são mais relevantes. Ao mesmo tempo, o enfoque dado pelos reguladores pode variar, desde a ênfase em aspectos técnicos até na definição de requisitos para o licenciamento das plataformas (Gasser; Schultz. 2015, p. 9).

Há diversos motivos que justificam a regulação das atividades das intermediárias *online*, podendo estar relacionados a externalidades, como no caso de determinar-se que as intermediárias devem ser responsáveis por todos os custos do serviço, sem repassar a terceiros; outra razão pode ser garantir um nível de atividades essenciais, como para criar um ambiente de informações diverso e baseado em múltiplas fontes; ou, ainda, estar relacionado com a proteção de interesses de populações vulneráveis, como as crianças. Assim, quanto às motivações e funções do sistema de regulação das plataformas digitais, identificam-se três categorias analíticas distintas e que podem ocorrer concomitantemente: permissões (*enabling*), sopesamento (*leveling*) ou restrições (*constraining*) (Gasser; Schultz. 2015, p. 10).

O sistema de governança das intermediárias *online* estadunidense, especialmente pautado na *Communications Decency Act*, é o mais relevante exemplo do caráter permissivo, uma vez que a norma restringe a possibilidade de responsabilização das intermediárias *online*

sobre o conteúdo postado por terceiros. A razão histórica para a inclusão de tal dispositivo na normativa estadunidense era o objetivo do legislador de permitir que as intermediárias exercessem a regulação dos conteúdos sem que estivessem expostas à responsabilização. Entretanto, os debates políticos recentes que envolvem a manutenção de tal previsão normativa guardam relação, especialmente, com duas outras justificativas: a primeira diz respeito aos benefícios econômicos do funcionamento das intermediárias nos Estados Unidos, por meio do incentivo à inovação, sem a preocupação com a responsabilização; e a proteção do direito fundamental à liberdade de expressão (Gasser; Schultz. 2015, p. 10), sobre o qual discorreremos anteriormente.

O sopesamento, ocorre quando o sistema de governança das intermediárias *online* estabelece mecanismos de sopesamento de direitos contrapostos, podendo ocorrer por meio do estabelecimento de mecanismos gerais de notificação e exclusão de conteúdos, como no caso estadunidense, por exemplo, quando da contraposição entre os direitos autorais e os direitos dos usuários, ou pode estar pautado em critérios mais abrangentes para análise caso a caso, como no caso da União Europeia, quando da contraposição entre o direito de ser esquecido e o eventual interesse dos usuários na informação (Gasser; Schultz. 2015, p. 10-11).

A restrição também é comum a diversos sistemas de governança, podendo ser identificada, por exemplo, no caso da exigência de implementação de mecanismos de notificação e exclusão de conteúdo para que as intermediárias *online* possam valer-se dos benefícios da isenção de responsabilidade sobre o conteúdo postado por terceiros. Entretanto, tal função pode assumir características mais restritivas (e não democráticas) em outros sistemas de governança, como no caso da Tailândia, em que há a previsão de responsabilização das intermediárias *online* para preservar a ordem pública e controlar o fluxo de informações, por meio de censura e vigilância, ou por meio de regras restritas para o licenciamento das intermediárias *online* no país, como no caso do Vietnã (Gasser; Schultz. 2015, p. 11).

Outro aspecto relevante no estudo dos sistemas de governança das intermediárias *online* é os ramos dos governos que atuam nesse sistema. Diferentes órgãos governamentais podem constituir o elemento central em sistemas de governança e a estrutura adotada, bem como o nível de envolvimento do governo no sistema de governança colocam em evidência questões fundamentais quanto a legitimidade, transparência, *accountability* e incentivos gerados pelo sistema. Outro aspecto que não pode ser negligenciado é o nível de conhecimento técnico que a estrutura governamental tem sobre o funcionamento das diferentes intermediárias *online*, uma vez que esta *expertise* pode estar distribuída desigualmente entre os diferentes órgãos governamentais (Gasser; Schultz. 2015, p. 11).

Muitos sistemas de governança estão pautados nas decisões das Cortes de Justiça, como era o caso brasileiro até a edição do Marco Civil da Internet e, de certa forma, mesmo após a aprovação desta. Outra possibilidade é a atuação estatal por meio de uma agência governamental especializada, como é o caso da Coréia do Sul, onde há uma agência para atuar sobre os casos que não envolvam direitos autorais, ou no caso mais extremo de países que instituem regras de licenciamento mais restritivas, como no já citado caso do Vietnã (Gasser; Schultz. 2015, p. 11-12).

As estruturas de incentivo podem, ainda, ser pautadas em diferentes técnicas de cumprimento dos preceitos normativos. Nesse sentido, uma técnica que vem se tornando mais comum se dá por meio do uso de algoritmos, especialmente nos EUA. Assim, por exemplo, os algoritmos são utilizados para lidar com a proteção dos direitos autorais em plataformas como o *YouTube*, enquanto um mecanismo privado, mas também são utilizados como forma de monitoramento governamental por parte de governos como na Tailândia, Turquia e Índia (Gasser; Schultz. 2015, p. 13).

Outro importante fator quanto aos mecanismos de cumprimento das normas de *compliance* diz respeito aos custos de acionamento dos diferentes mecanismos. No caso da Turquia, por exemplo, em que há a possibilidade de iniciar um procedimento criminal sem custos, aliado às dúvidas quanto aos mecanismos administrativos de notificação e exclusão de conteúdos, implica em uma preferência pelo acionamento do sistema judicial em detrimento de soluções administrativas. Outro caso que ilustra a importância dos custos quando do estabelecimento da regulamentação das plataformas digitais é a contraposição entre mecanismos automatizados como a proteção de direitos autorais realizada pelo *YouTube* e mecanismos de análise caso-a-caso realizado por seres humanos, como no caso do direito ao esquecimento da União Europeia. Desta forma, os custos para a adoção de um ou outro mecanismo estruturam o funcionamento do sistema de governança das intermediárias *online*, ao gerar incentivos para a maneira como usuários e plataformas optam por agir (Gasser; Schultz. 2015, p. 13-14).

Ao mesmo tempo, outro aspecto que influencia a estruturação de um sistema de governança das intermediárias *online* é a arena internacional, tendo em vista a transnacionalidade das plataformas digitais. As regras de funcionamento do sistema podem, assim, ser inspiradas por mecanismos internacionais, e podem ser resultado até de acordos bi ou multilaterais, como demonstram as negociações da Parceria Transpacífica (*Trans-Pacific Partnership Agreement*); ou regras internacionais de direitos humanos como a Declaração dos Direitos Humanos e o Pacto Internacional dos Direitos Civis e Políticos, o *Digital Services Act*

na União Europeia; ou, ainda, resultado de esforços conjuntos de diversas partes interessadas, como a iniciativa da *NETmundial* (Gasser; Schultz. 2015, p. 14-15).

Ainda, é necessário reconhecer as especificidades dos modelos de governança das intermediárias *online* de acordo com as experiências locais, tanto no que diz respeito às razões para sua estruturação, a forma como surge e as especificidades dos modelos de governança, no que diz respeito a estratégias, instituições, modalidades e efeitos gerados pela regulação. Ainda, não é possível haver uma boa compreensão dos sistemas de governança sem levar-se em consideração aspectos socioeconômicos, políticos e culturais (Gasser; Schultz. 2015, p. 16), conforme demonstrado pela importância de tais fatores na forma como se dá a comunicação social pelas mídias digitais, discutido no primeiro capítulo.

A despeito das especificidades, os modelos de regulamentação das plataformas digitais podem ser subdivididos em três categorias: aqueles em que não há normas especificamente voltadas à governança das intermediárias *online*; aqueles em que há normas específicas e distintas para as intermediárias *online*; e aqueles em que estão se estabelecendo normas específicas de governança das intermediárias. O impulso e a motivação para a regulamentação estão geralmente relacionados a aspectos específicos concernentes a cada experiência local, especialmente no que diz respeito ao regime de responsabilização das intermediárias *online*. Ao analisar o desempenho comparado dos diferentes sistemas de governança, portanto, destacam-se os incentivos criados pelos sistemas e a forma como tais incentivos agem em relação aos diferentes sujeitos que atuam na área. Os governos, nesse sentido, atuam em basicamente três funções no sistema de governança, anteriormente descritas: permissões, sopesamento e restrições (Gasser; Schultz. 2015, p. 16).

Assim, estratégias de regulamentação das plataformas digitais devem levar em consideração alguns aspectos da estrutura das tecnologias de comunicação e informação, além de aspectos socioeconômicos e políticos: 1) a questão do problema da definição do que é uma intermediária *online*; 2) as distinções entre as diferentes intermediárias *online*; 3) a estrutura do mecanismo de notificação e exclusão de conteúdos; e 4) os custos de *compliance* e aplicação das normas (Gasser; Schultz. 2015, p. 16). Para a adequada compreensão do tema, portanto, é necessário:

- a) Compreender a função e a economia das intermediárias *online*: por serem um fenômeno recente, as intermediárias *online* são um condutor e um reflexo das mudanças estruturais do ecossistema de informações. Grande parte do esforço de regulamentação das intermediárias *online* reside na necessidade de compreender qual é o papel exercido por tais empresas, uma vez que estas não são criadoras de conteúdo, nos termos da

legislação sobre a mídia tradicional, mas participam na disseminação dos conteúdos elaborados por terceiros. Dessa forma, definir os papéis das intermediárias *online* na sociedade contemporânea é fundamental para o estudo sobre a governança destas (Gasser; Schultz. 2015, p. 17);

- b) Considerar a importância da dimensão normativa nos sistemas de governança: a intersecção entre a economia digital e a responsabilidade das intermediárias *online* gera atritos entre os benefícios econômicos do desenvolvimento das intermediárias *online* e a proteção a direitos, especialmente os direitos humanos. É imperioso, desta forma, contrapor os benefícios econômicos à necessidade de regulação das intermediárias *online* (Gasser; Schultz. 2015, p. 17);
- c) Analisar todos os mecanismos regulatórios: as intermediárias *online* estão sujeitas à regulação de diversos mecanismos distintos, sejam estes diretos ou indiretos, *ex ante* ou *ex post*, ou por meio de normas coercivas ou não coercivas (*soft or hard obligations*). Os diferentes atores do sistema de governança são dotados de recursos diferentes, níveis distintos de *expertise* e abordagens distintas. Ter em mente os diferentes mecanismos atuantes em um sistema de governança das intermediárias *online* é fundamental para a adequada compreensão do sistema (Gasser; Schultz. 2015, p. 18).
- d) Considerar os custos da regulação das intermediárias: por tratar-se de uma atividade de alta complexidade e que exige *expertise*, os governos têm a tendência de direcionar as normas de governança às intermediárias *online*. Entretanto, tal solução pode gerar dificuldades como a regulação excessiva, conforme exposto acima, que resulta em impactos negativos a direitos fundamentais, à inovação e à economia digital. Ademais, há outros custos relativos à responsabilização das intermediárias *online* sobre o conteúdo postado por terceiros, como o risco de conferir demasiado poder para empresas que já são poderosas, para decidir sobre a divulgação de conteúdo (Gasser; Schultz. 2015, p. 18);
- e) Fortalecer mecanismos de aprendizagem entre os diferentes sistemas: uma vez que a regulamentação das intermediárias *online* gera questões comuns às diversas experiências locais, como no caso de estruturação de mecanismos de notificação e exclusão de conteúdos, e que as intermediárias *online* atuam de maneira transnacional, há um grande potencial no compartilhamento de experiências entre os países para o desenvolvimento e fortalecimento de sistemas de governança (Gasser; Schultz. 2015, p. 18).

2.3. *Fake news* e discurso de ódio: o que e como moderar? Um esboço

Ao longo do primeiro capítulo, buscamos compreender o fenômeno da disseminação de *fake news* e discurso de ódio por meio das tecnologias de informação e comunicação digitais. Compreendemos que há aspectos sociais, econômicos e tecnológicos que confluem para resultar na estrutura da comunicação digital que procuramos descrever. Nesse panorama, *fake news* e discurso de ódio são apenas alguns dos materiais utilizados como ferramenta de uma disputa da realidade fática por meio de mídias hiperpartidárias, pautadas na capacidade de identificar anseios e sentimentos dos indivíduos e direcionar os conteúdos de maneira individualizada, por meio do capitalismo de vigilância. Assim, é imperioso reconhecer que há outros conteúdos que não necessariamente contêm inverdades ou que incitam a animosidade em determinados grupos que, ainda assim, podem ter como resultado impactos negativos para a democracia, como a polarização política, por exemplo. Ou seja, ainda que houvesse uma forma de anular toda e qualquer disseminação de *fake news* e discurso de ódio pelas plataformas digitais, a perda da racionalidade discursiva, suplantada pela racionalidade digital e a psicopolítica dado-pulsional, ainda se imporia como corolário da comunicação digital.

Ao mesmo tempo, ao identificar a estrutura da economia do capitalismo de vigilância, foi possível observar que há diversos agentes envolvidos em uma campanha de disseminação de conteúdos na seara digital, desde a captação dos dados e a análise psicométrica, realizada por empresas especializadas; robôs que iniciam o processo de compartilhamento dos conteúdos; influenciadores que dão sentido e propagam aquela mensagem; até usuários das plataformas digitais que compartilham, de boa-fé e inseridos em câmaras de eco, conteúdo que entendem ser verídico. Trata-se, portanto, de um processo composto por várias etapas, que não pode ser compreendido apenas pelo seu resultado final, como na veiculação de uma peça ou um conjunto de peças digitais que sejam considerados falsos, com a intenção de gerar engano, sendo imprescindível entender quais são os agentes que financiam, encomendam e produzem tais materiais com a finalidade de, dotados dos dados dos usuários, que foram inadvertidamente compartilhados, sem ter a consciência de que seriam utilizados para influenciar na própria formação de vontade, elaborar os conteúdos que serão disseminados, utilizando-se da arquitetura das plataformas digitais.

Nessa cadeia produtiva da informação e comunicação digital, não é possível imputar aos próprios algoritmos a vontade de agir ou a culpabilidade, instrumentos jurídicos tradicionais. Entretanto, as empresas que desenvolvem e oferecem tais plataformas são, naturalmente, responsáveis pelos eventuais danos decorrentes destes serviços, não apenas

pautado na proteção da democracia e do interesse público, da construção e manutenção de uma esfera pública vibrante, mas também em função da própria realização do princípio constitucional da liberdade de expressão e da autodeterminação, uma vez que resta prejudicada a própria formação de vontade e autonomia na racionalidade digital.

Desta forma, *fake news* e discurso de ódio são sintomas de uma comunicação social e debate público adoecidos, calcados em estruturas sociopolíticas, econômicas e tecnológicas que incentivam a fragmentação da vida comum e a exacerbação do consumo e do individualismo, derivadas do capitalismo e do informacionalismo. Sem que se dissipem as causas, dificilmente superam-se as consequências.

Neste panorama, e na imprescindibilidade da regulamentação dos meios de comunicação e informação digitais, torna-se necessário discutir coletivamente, pela atuação dos diversos atores envolvidos na comunicação digital, notadamente a sociedade civil, as plataformas digitais, os governos e os usuários, estratégias complementares que visem incidir sobre as condutas ilegítimas do processo de comunicação digital, protegendo-se a liberdade de expressão tanto em sua acepção tradicional, quanto na nova interpretação derivada da racionalidade digital. Assim, a médio e longo prazo, é imprescindível que haja maior transparência sobre o funcionamento das plataformas digitais, tanto em relação aos usuários, que devem poder compreender melhor como os conteúdos são veiculados pelas mídias digitais, quanto a auditoria externa, que viabilize a avaliação quanto a eventuais abusos cometidos no mercado de comunicação.

Torna-se necessário, mediante os preocupantes dados sobre a disseminação de *fake news* e discurso de ódio pelas mídias digitais, influenciando inclusive em votações nas campanhas presidenciais estadunidense e brasileira, assim como no *Brexit*, que se instituem mecanismos de indisponibilização de conteúdos, ou estratégias alternativas como o etiquetamento ou *linking*, além da educação digital, enquanto medidas jurídicas-institucionais imediatas para minimizar os efeitos perniciosos da disseminação destes conteúdo.

Definir um conceito jurídico de *fake news*, que seja objetivo, entretanto, é objeto de amplo debate tanto no Brasil quanto no mundo de maneira geral, talvez jamais superável. Conforme discutimos anteriormente, há diversos tipos de conteúdo que contém falsidades, cujo objetivo não é gerar desinformação ou confusão do público, mas com objetivos diversos como fazer uma brincadeira, sátira ou uma crítica. Conteúdos dotados de uma opinião ou perspectiva sobre determinado fato não devem ser considerados *fake news*, uma vez que não buscam necessariamente distorcer os fatos, gerando desinformação, mas sim elaborar um determinado juízo sobre estes. Ademais, há conteúdos que, apesar de conter falsidades, são resultado de um

erro ou engano acidental, porém justificável. Há, ainda, o uso indevido do termo *fake news*, especialmente por personalidades políticas, com o objetivo de desacreditizar críticas de determinados grupos ou políticos ou como forma de deslegitimar a imprensa em geral, o que representa uma ameaça à liberdade de expressão e informação (Horbach, 2022, p. 192-193).

Entretanto, reafirmamos que o uso do termo *fake news* não deve ser descartado, ao menos na experiência brasileira, apenas pelos eventuais enganos ou pelo uso leviano de determinadas personalidades. De fato, o termo tem o condão de delimitar uma questão fundamental à democracia brasileira contemporânea, possibilitando o debate social, legislativo e judicial. Ao mesmo tempo, faz-se necessário compreender quais conteúdos de fato se constituem enquanto *fake news*, enquanto uma possibilidade de categoria jurídica, ainda que seja impossível definir objetivamente, de maneira a identificar um equilíbrio que privilegie, ao mesmo tempo, o princípio da liberdade de expressão e da liberdade de informação na sociedade contemporânea, enquanto garantias e pressupostos de um sistema democrático, estabelecendo balizas para a atuação das empresas de tecnologia, Estado, sociedade civil e usuários.

A ampla veiculação de *fake news* é ameaçadora à democracia, entre outros motivos, porque fragiliza a legitimidade de veículos de informação socialmente reconhecidos e autoridades e enfraquece a capacidade social de debate racional sobre fatos compartilhados. O que se deve buscar ao regulamentar a comunicação digital, portanto, não é limitar a veiculação de conteúdos que contém sátira, opiniões ou erros, o que implicaria em grave ofensa ao princípio da liberdade de expressão, mas, sim, identificar outros materiais com ampla veiculação, produzidos deliberadamente, e que pretendem se passar por fatos, mas que se tratam de falsidades (*fake*), com o objetivo de gerar desinformação e manipulação, buscando parecer legítimo, passando-se por notícia ou novidade (*news*). São categorias deste tipo de conteúdo: falsa conexão, em que as manchetes, imagens ou subtítulos implicam em uma falsa percepção da notícia, que é real; conteúdo enganoso, que é a distorção de um fato com o objetivo de gerar confusão; o contexto falso, em que uma notícia verdadeira é integrada com falsidades para gerar confusão; o conteúdo impostor, que atribui a determinada instituição, pessoa, entidade ou organização o material falso, buscando conferir confiabilidade a este; o conteúdo manipulado, cuja informação é elaborada de forma a confundir o público; e o conteúdo fabricado, que é completamente falso (Horbach, 2022, p. 192-193).

Em uma outra abordagem jurídica que poderia proporcionar um conceito mais restrito de *fake news*, tornando possível uma regulamentação sobre o tema, destaca-se que a disseminação de conteúdo falso pelas mídias digitais é um fenômeno específico da transição da comunicação tradicional para a comunicação digital, em que agentes que disseminam estes

conteúdos se aproveitam da credibilidade de empresas de jornalismo para produzir notícias fraudulentas ou falsificadas, passando-se por jornalismo, mas sem o processo de produção de material genuinamente jornalístico, motivo pelo qual ressalta-se o potencial de “enganosidade” deste tipo de conteúdo, ou seja, a capacidade de parecer ser verdadeiro (Maranhão; Campos, 2019, p. 322).

Tendo em vista, entretanto, que a comunicação digital se trata de um ambiente em constante e rápida transformação, o próprio conceito de *fake news* não deve ser engessado, de forma que sua compreensão, identificação e combate deve acompanhar as mutações promovidas neste ambiente comunicacional. Desta feita, e uma vez que as empresas de tecnologias, provedores de conteúdo e provedores de redes sociais na *internet* são dotados da especialidade técnica necessária ao tratamento do tema, a participação destes atores no sistema de governança que tenha como finalidade o combate à disseminação de *fake news* é imprescindível, do que decorre a instituição de sistemas de autorregulação. Todavia, não cabe ao particular, especialmente em um mercado com alta concentração de poder como o mercado de tecnologia, delimitar os parâmetros de atendimento ao interesse público na regulação de matéria tão sensível, de forma que cabe ao Estado, no papel de regulação, induzir e balizar o sistema de autorregulação adotado pelas empresas (Maranhão; Campos, 2019, p. 321-322).

Os mecanismos de controle de disseminação de conteúdos falsos devem ser dotados de contraditório e ampla defesa, de forma que, na tomada de decisão, deve ser oportunizado àquele que publicou o material a fundamentação e justificação da medida. Tal procedimento, tendo em vista a velocidade com que se dá a disseminação destes conteúdos e, consequentemente, a realização de seu potencial lesivo, deve ser célere, de forma a dar eficácia ao sistema. O atual sistema vigente no Brasil delega ao Judiciário a solução de todos os casos envolvendo a disseminação de conteúdos nas plataformas digitais. Ocorre que o Judiciário não é dotado de celeridade equivalente à velocidade com que se realizam os danos provenientes da disseminação de tais conteúdos, conforme ficará evidente pelos dados analisados no Capítulo 3 desta dissertação, e, tampouco, da *expertise* necessária ao tema. Além disso, pela natureza da solução judicial, ou seja, o caso-a-caso, não se promove uma solução sistemática ao problema, restando ilesos os agentes ou organizações que atuam na elaboração e disseminação massiva e articulada de *fake news* (Maranhão; Campos, 2019, p. 322-323).

Demonstra-se, portanto, a necessidade de se instituírem medidas que atuem sistematicamente sobre a estrutura da comunicação digital. Para proteger a liberdade de expressão em sua concepção tradicional, entretanto, medidas mais gravosas como a exclusão

de conteúdos e perfis das plataformas digitais, devem ser criteriosamente realizadas e, evidentemente, todas as medidas administrativas devem ser passíveis de revisão judicial.

A comunicação digital é marcada pelo aspecto da extraterritorialidade das plataformas digitais em relação aos diversos países, como por exemplo o Brasil, bem como pela transnacionalidade da própria comunicação, em que a confluência dos mundos virtual e real resulta na estruturação de uma espécie de Cosmópolis, ou seja, ações tomadas em locais fisicamente distantes se aproximam por meio da virtualidade. Essa característica implica em que a normatização da *internet* resulte em questões de legitimidade e territorialidade da aplicação das normas. Nesse sentido, se a *internet* conecta diversas localidades, a normatização ainda é majoritariamente nacional e, eventualmente, regional, de forma que se pode identificar a existência de múltiplas “*internets*”, resultado da intersecção entre a normatização local e a estrutura técnica em si (Macedo Júnior, 2020, p. 233-234).

A maneira como a comunicação digital se dá, com a participação massiva dos próprios usuários na produção de conteúdo, implica em duas características: a primeira é que informações que eram tratadas como privadas passaram a ser amplamente publicizadas, sendo, em verdade, o modelo de negócio das próprias plataformas, como discutido no primeiro capítulo. Por um lado, tal possibilidade democratizou o acesso aos cidadãos comuns ao público e, por outro, cedeu a empresas privadas, no caso as plataformas digitais, informações pormenorizadas sobre os indivíduos, possibilitando o uso dos seus dados para influenciá-los de maneira inadvertida. A segunda característica é que a mídia de massa tradicional, especialmente jornais impressos, rádios e TVs, além de funcionarem de maneira distinta das plataformas, desenvolveram, ao longo de décadas de debate e regulamentação, normas éticas de imprensa e jornalismo, estabelecendo uma espécie de cultura de compromisso com a verdade, que não se aplica à comunicação digital (Macedo Júnior, 2020, p. 234-235).

Mediante as tecnologias de informação e comunicação digital da sociedade algorítmica, há uma nova escola da regulação do discurso, pautada sobre uma nova interpretação do princípio da liberdade de expressão e os novos desafios à democracia. Dois aspectos que têm sido regulados ao redor do mundo são o direito ao esquecimento, especialmente desenvolvido pela Corte de Justiça da União Europeia, e a disseminação de *fake news*, problema que passou a ser reconhecido mediante as eleições presidenciais estadunidenses de 2016 (Balkin, 2018, p. 1172-1173).

A nova escola de regulação do discurso tem como objeto a própria estrutura da comunicação digital, abrangendo provedores de *internet*, serviços de *internet*, *hostings*, plataformas digitais e mecanismos de pesquisa, diferentemente da escola tradicional, que atua

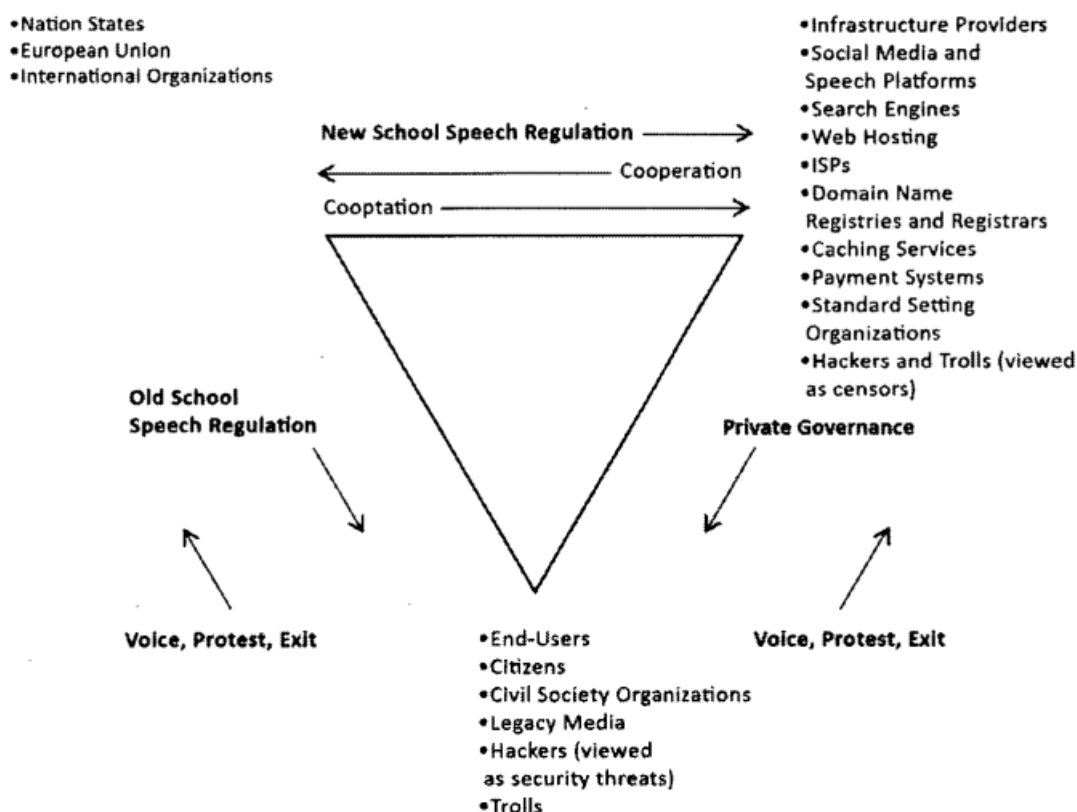
mediante a responsabilização dos publicadores de conteúdo, com medidas como responsabilização civil e criminal, prisão e liminares. Tendo em vista a dificuldade de regular os conteúdos veiculados digitalmente, esta nova estratégia se sustenta sobre a capacidade das grandes empresas de tecnologia em bloquear, filtrar e remover conteúdo, tendo em vista sua estrutura de vigilância e controle sobre os serviços que oferecem. Esse sistema é dotado de três características: 1) gera incentivos para que se exerça uma espécie de “censura colateral”; 2) envolve cooperação ou cooptação público/privada; 3) envolve governança privada dos conteúdos, especialmente realizada pelas redes sociais e mecanismos de busca (Balkin, 2018, p. 1173-1176).

A escola tradicional de liberdade de expressão e regulação do discurso, desenvolvida no século XX, é caracterizada por conter dois polos: o Estado regulador de um lado e os indivíduos e publicadores regulados de outro. Na nova escola, entretanto, trata-se de uma estrutura pluralista composta por: Estados e entidades supranacionais como a União Europeia, que utilizam ferramentas da escola tradicional e da nova escola, agindo como reguladores; as plataformas digitais, especialmente redes sociais e mecanismos de busca, que exercem governança por meio de seus termos de serviço, licenças dos usuários e normas internas, que podem assumir caráter transnacional e conter milhões de usuários em diversos países; e os usuários que utilizam as plataformas para se comunicarem, sendo objeto de vigilância e controle tanto público quanto privado (Balkin, 2018, p. 1184-1188).

As relações que se estabelecem pelo direito, o mercado, as normas e as questões técnicas e de arquitetura da *internet*, inclusive os próprios algoritmos, *softwares* e *hardwares*, resultam em uma nova estrutura de poder na sociedade contemporânea, disputada por uma miríade de atores distintos como “organizações internacionais, governos nacionais, companhias, engenheiros, meios de comunicação” e “campanhas de massa físicas ou virtuais” (Macedo Júnior, 2020, p. 234-235).

Balkin elabora um diagrama em que as relações entre os sujeitos envolvidas na estrutura da nova escola da liberdade de expressão estão representadas:

Diagram 1. The Pluralist Model of Speech Regulation



Balkin, 2018, p. 1189.

O novo sistema é pluralista, portanto há diversos sujeitos que se relacionam de maneiras particulares, entre elas: 1) diferentes organizações e plataformas que podem criar seu próprio rol de governança digital, atuando de maneiras e em níveis distintos; 2) há a tensão entre as mídias tradicionais e as mídias digitais: inicialmente o debate se concentrava em torno de conteúdos protegidos por direitos autorais, entretanto, cada vez mais, as mídias tradicionais exercem *lobby* para que as mídias digitais sejam responsabilizadas por discurso de ódio e outros conteúdos veiculados em seus serviços; 3) os usuários e as organizações da sociedade civil podem pressionar as plataformas tanto para moderarem o conteúdo veiculado quanto para que o controle seja menor; 4) a atuação de *hackers* que pressionam as plataformas; 5) os grupos de usuários podem usar técnicas para disciplinar, confundir, desencorajar e ameaçar outros usuários, como na atuação de *trolls* ou na distribuição massiva de conteúdos que buscam desacreditar mídias confiáveis e gerar confusão, ao dificultar o reconhecimento do que é verdadeiro ou falso (Balkin, 2018, p. 1189-1192).

No sistema de comunicação social pautado nas mídias de massa tradicionais, como a televisão, o jornal impresso e o rádio, a estrutura comunicacional era diversa e os meios de

comunicação exerciam o papel de fornecer as informações relevantes e construir uma determinada perspectiva de mundo, sem um direcionamento individual, mas sim produzindo material mais geral e acessível. A produção do conteúdo jornalístico nesse sistema é balizada por dois aspectos: as técnicas de checagem e padrões éticos próprios do trabalho jornalístico e a possibilidade de responsabilização civil e penal do meio de comunicação pelas informações veiculadas. Por meio destas duas balizas, as democracias liberais se estruturaram com a possibilidade de controle judicial e político da comunicação social, panorama que foi modificado pela comunicação digital (Maranhão; Campos, 2019, p. 324).

A autocomunicação de massa, propiciada pelas novas tecnologias de comunicação e informação digitais, carece das balizas características da comunicação social que se dá pelas mídias de massa tradicionais. Se a produção e circulação das informações se modificou, o alcance dessas foi ampliado, especialmente em razão da possibilidade de direcionamento de conteúdo para públicos específicos. A velocidade com que os conteúdos são veiculados tornam obsoletos mecanismos jurídicos que, antes, eram adequados, tornando ineficazes medidas como a decisão judicial de remoção do conteúdo ou o direito de resposta. Nesse novo panorama, as plataformas digitais exercem, através da coleta de dados dos usuários e direcionamento de conteúdos, papel central na comunicação social, de forma que devem ser o principal foco de uma estratégia de regulação adequada às novas condições tecnológicas e sociais, que seja transparente e possibilite o direito de defesa e que, ao mesmo tempo, estabeleça deveres de informação das plataformas frente ao Estado, em razão da especificidade e complexidade técnica do tema, de maneira a possibilitar o norteamento pelo interesse público (Maranhão; Campos, 2019, p. 324-326).

O papel exercido pelos agentes que influenciam no fluxo de informação e comunicação na sociedade digital deve coadunar com a própria estrutura tecnológica e sua posição de poder relativa. Assim, aventa-se sobre o papel que os sujeitos envolvidos na comunicação digital podem desempenhar em um sistema moderno e adequado às tecnologias contemporâneas:

1) Judiciário: tradicionalmente as Cortes lidam com o caso-a-caso para averiguar a legalidade de determinado conteúdo veiculado pelas mídias, uma vez que é muito difícil instituir por meio da lei parâmetros de sopesamento entre princípios constitucionais. Todavia, o largo volume de conteúdos veiculados pelas mídias digitais torna inviável que as Cortes avaliem todos os casos e deem uma resposta satisfatória. Ao invés disso, as Cortes poderiam averiguar se o procedimento administrativo realizado pela mídia digital respeita critérios do devido processo legal e possibilitou ao usuário, mediante notificação, uma adequada avaliação sobre a legalidade daquele conteúdo. Assim, seria criada uma espécie de autorregulação

regulada, em que o legislador institui normas básicas de procedimento para moderação de conteúdo e o Judiciário passa a analisar se os procedimentos adotados pela plataforma cumprem com tais requisitos (Hartmann, 2020, p. 6-7);

2) usuários: os espectadores das mídias de massa tradicionais exerciam um papel passivo, a despeito de não apenas aceitarem o conteúdo tal como veiculado, como discutido no primeiro capítulo, mas no que diz respeito à capacidade de produzir e disseminar conteúdo. As mídias digitais modificaram esse panorama, possibilitando que os usuários produzam e disseminem conteúdos, bem como interajam com os conteúdos de outros usuários. A capacidade dos usuários de averiguar quais conteúdos são irregulares e notificarem as plataformas sobre cada um desta vasta gama de materiais é muito maior que a mesma capacidade das Cortes. Aos usuários, assim, pode ser conferido o papel de averiguar, caso-a-caso, quais são os conteúdos veiculados que são prejudiciais, notificando as plataformas que, então, instaurariam o procedimento administrativo para verificar a validade do conteúdo. Assim, não seria necessário instituir previamente por via legislativa quais conteúdos em tese são potencialmente prejudiciais, mas essa identificação seria exercida pela própria comunidade, ressalvada a avaliação posterior realizada pela plataforma e a possibilidade de revisão judicial do procedimento adotado pela plataforma. Assim, se as plataformas não têm legitimidade para decidir sobre a legalidade de um conteúdo, e as Cortes não têm a estrutura para decidir cada caso de moderação de conteúdo, a comunidade em si, por meio das notificações dos usuários, exerceria uma espécie de influência na comunicação digital, ao participar do questionamento da validade/legalidade dos conteúdos veiculados, reinstituindo-se uma espécie de debate público (Hartmann, 2020, p. 7-8).

3) plataformas digitais: as plataformas não devem, assim, ser detentoras do poder de moderar os conteúdos em cada caso, mas empoderar os usuários para que o façam em comunidade. Assim, as plataformas devem contribuir ao instituir mecanismos de moderação com participação dos usuários, cujas diretrizes sejam pautadas legislativamente e o procedimento seja revisado pelo Judiciário. O papel das plataformas seria de identificar a correção das notificações enviadas pelas comunidades, levando em consideração aspectos como se há uma grande quantidade de usuários que apontaram irregularidade do conteúdo ou se os usuários que notificaram constituem um grupo coeso filiado a determinada perspectiva que aponta conteúdos de outros grupos como irregular apenas por discordar destes, entre outros aspectos.

A própria comunidade, então, seleciona os casos que serão analisados pelas plataformas, reduzindo o esforço que estas teriam de fazer para averiguar todo o conteúdo veiculado em seus

serviços. Incentivos para que os usuários produzam esse *feedback*, para que estes sejam elaborados por vários grupos de usuários distintos, evitando os efeitos negativos dos filtros-bolha e que sejam os mais pormenorizados possíveis, disponibilizando-se espaços para que o usuário descreva por que entende que aquele conteúdo é irregular, são pontos fundamentais para instituir um sistema funcional de notificação e exclusão de conteúdo. Ademais, é fundamental que as plataformas sejam transparentes quanto aos critérios e decisões adotadas, especialmente no que diz respeito a decisões tomadas por meio de algoritmos e inteligência artificial, tornando possível que os usuários e o Poder Público possam revisar os critérios adotados (Hartmann, 2020, p. 8-9).

Conforme veremos através dos dados do TSE e do *Facebook* analisados no último capítulo, as plataformas são capazes de fazer (e fazem) uma análise do impacto e alcance de determinado conteúdo, bem como da potencial irregularidade desde, diminuindo substancialmente a quantidade de conteúdos que necessitam de avaliação humana, tornando mais viável o processamento do volume de conteúdo veiculado diariamente.

4) o Poder Executivo: o conceito de liberdade de expressão tradicional e seus desdobramentos na comunicação por meio das mídias de massa, implica no impedimento do Executivo de exercer qualquer espécie de moderação de conteúdo. Entretanto, a nova estrutura da comunicação digital implica, como exposto anteriormente, em que a avaliação quanto a cada conteúdo propriamente ou a critérios de moderação de conteúdos sejam apenas parte dos mecanismos de combate à disseminação de conteúdos irregulares, com atuação acentuada dos usuários e do Judiciário. Por outro lado, o próprio procedimento a ser adotado para a moderação de conteúdos, cujo protagonista são as plataformas digitais, deve ser objeto de fiscalização da sociedade e do Poder Público. Em razão da complexidade envolvida nos procedimentos, especialmente no que diz respeito ao uso de inteligência artificial e algoritmos, e da necessidade econômica de que tais estruturas sejam protegidas por sigilo, o Poder Executivo poderia exercer, por meio de uma agência, com participação multissetorial, por exemplo, o papel de fiscalizar essas etapas do procedimento, capacitando-se tecnicamente para tanto (Hartmann, 2020, p. 9).

2.3.1. Reféns da governança privada: entre a cruz e a espada

Na decisão sobre o direito ao esquecimento na União Europeia, verifica-se que a responsabilização da plataforma é uma espécie de cooptação do Estado para que as empresas

privadas exerçam moderação de conteúdo, uma vez que as próprias estruturas das Cortes e de agências de governo são insuficientes para responder à grande quantidade de demandas pelo direito ao esquecimento, de forma que um agente público, no caso a União Europeia, coopta uma organização privada, no caso o *Google*, para exercer governança, pautada na crescente capacidade de vigilância das plataformas. Esse processo tem duas implicações distintas: as plataformas, empresas privadas, ganham o direito-dever de instituir os mecanismos de governança, podendo desenvolvê-los de acordo com seus próprios interesses, enquanto o Estado utiliza a estrutura das plataformas para exercer governança, o que pode gerar incentivos para que, posteriormente, os agentes públicos façam ainda mais exigências. Assim, ao mesmo tempo em que é exigido às plataformas que atuem em função do interesse público, as instituições públicas tornam-se cada vez mais dependentes das plataformas para exercerem governança, tornando-as indispensáveis (Balkin, 2018, p. 1206-1207).

Da mesma forma, o problema das *fake news*, mais relacionado às redes sociais, tem mobilizado estruturas da nova escola de regulação do discurso, uma vez que é muito dificultoso que o Estado persiga as pessoas e instituições que se utilizam destes expedientes, tendo em vista que muitos deles são anônimos, utilizam *bots* ou estão localizados fora do âmbito nacional. Assim, os Estados, bem como os usuários em comunidade, tendem a incentivar que as redes sociais instituem mecanismos de moderação de conteúdo, identificando materiais com *fake news* e seus produtores, bloqueando *links* ou provendo informações verdadeiras sobre o tema, bloqueando, filtrando e vigiando os conteúdos que são veiculados em seus serviços. Também neste caso, as plataformas são incentivadas a ampliar seus mecanismos de governança privada (Balkin, 2018, p. 1207-1209).

O assunto tem sido objeto de regulamentação em diversos países ao redor do mundo. A lei alemã, *NetzDG*, tem sido apontada como uma das pioneiras na instituição de mecanismos de regulamentação com o objetivo de enfrentar a disseminação de *fake news* e discurso de ódio nas plataformas digitais. A principal previsão do dispositivo legal é a de que as plataformas devem, mediante notificação dos usuários, apurar a legalidade do conteúdo postado por terceiros e, caso constatarem se tratar de material irregular, indisponibilizá-lo, sob pena de multa que pode ir a até cinquenta milhões de euros (Macedo Júnior, 2020, p. 240-242). Trata-se, portanto, da estratégia de responsabilização das plataformas digitais sobre o conteúdo postado por terceiros, mediante notificação dos usuários.

Tal mecanismo tem sido objeto de críticas, especialmente no que diz respeito à dificuldade de se identificar o que é *fake news* e se tal previsão cria uma espécie de “obrigação de dizer a verdade” às plataformas; ao risco de que conteúdos legais sejam indevidamente

excluídos, no afã das plataformas de evitar a responsabilização; à criação de regras punitivas ou criminais, as quais deveriam ser últimas medidas, ao invés de buscar soluções alternativas; e a dificuldade de uma empresa estrangeira determinar quais conteúdos devem ser considerados *fake news* ou discurso de ódio e, assim, ser excluído. Ademais, conforme estudamos anteriormente, o conceito tradicional de liberdade de expressão tem em seu núcleo a proibição de que o Estado decida o que pode ou não ser divulgado, de forma que não apenas permitir, mas exigir, que as plataformas digitais o façam, parece estar em dissonância com o princípio da liberdade de expressão. Há uma profunda dificuldade em se identificar o que constitui falsidade em si, uma vez que apesar de determinados materiais serem facilmente aferíveis, há uma vasta quantidade de conteúdos que ficam em uma zona cinzenta, entre o exagero da comunicação, uma falta de precisão justificável ou mesmo a necessidade de se entender o contexto de uma determinada mensagem (Macedo Júnior, 2020, p. 242-245).

As plataformas digitais, nesse ecossistema, ainda que não sejam dotadas da autoridade formal para criar normas, como os Estados, atuam, ao estabelecer suas plataformas, de maneira a ditar decisivamente a maneira como as informações e comunicações serão veiculadas. Na verdade, o poder das plataformas nesse sentido, o qual é objeto de pouca ou nenhuma limitação, responsabilidades e controles público e social, é até mesmo maior que o de entes públicos. Além disso, as plataformas não são dotadas de deveres democráticos característicos de Estados Democráticos de Direito, uma vez que são empresas privadas. A posição ocupada pelas mídias digitais difere das mídias tradicionais na medida em que o conteúdo veiculado é publicado por terceiros, de forma que a estrutura técnica e as normas atinentes às mídias tradicionais são completamente distintas. Porém, ainda que os conteúdos que são amplamente veiculados pelas mídias digitais sejam produzidos por terceiros, a disseminação em si é diretamente relacionada com a arquitetura das plataformas, que atuam de maneira a impulsionar um ou outro conteúdo, sendo utilizada, por vezes, para produzir campanhas deliberadas de desinformação em massa. Assim, a despeito da possibilidade de qualquer usuário publicar um conteúdo que seja amplamente veiculado, a realidade é que cada publicação tem uma grande chance de se perder em meio ao vasto volume de conteúdo disseminado diariamente, enquanto estratégias relativamente baratas podem ser adotadas para que determinados conteúdos alcancem o grande público (Macedo Júnior, 2020, p. 236-237).

A governança, nos novos sistemas de regulamentação da comunicação digital, é exercida por um ente privado, dotado de infraestrutura técnica, que influencia o comportamento dos usuários, possibilitada pelas leis de telecomunicação, os direitos de propriedade das empresas que detém a infraestrutura e dos contratos de uso com os usuários, pautada sobre a

capacidade de controle e vigilância das plataformas. As intermediárias desenvolvem mecanismos de governança privada que decidem qual conteúdo pode ou não ser postado, baseado no interesse dos próprios usuários e da empresa de impedir a disseminação de determinados conteúdos, como por exemplo discurso de ódio e pornografia, que podem resultar em exclusão do conteúdo ou do próprio usuário da comunidade digital. Assim, as empresas são incentivadas pelos próprios usuários, além dos governos, a exercer governança. Esses sistemas são especialmente bem desenvolvidos em redes sociais e mecanismos de busca, onde as plataformas digitais reconheceram que seus usuários constituem uma comunidade dotada de regras de comportamento, ou seja, governança, e que é exercida pela própria empresa, por isso privada. Assim, as plataformas já são, hoje, independentemente da ação estatal, reguladoras de conteúdo em seus serviços e os usuários são regulados (Balkin, 2018, p. 1182-1184).

Um dos principais entraves que as plataformas apontam a qualquer mecanismo de regulamentação das plataformas digitais diz respeito aos custos derivados das operações de controle, a serem assumidos majoritariamente pelas plataformas. Essa alegação, entretanto, não subsiste mediante os lucros aferidos pelas grandes empresas de tecnologia de informação e comunicação digital, que não condizem com a perspectiva de que alguns mecanismos de controle tornariam sua atividade economicamente inviável (sobre a qual ilegalmente lucram). Apesar da alegação de que moderar os conteúdos disseminados digitalmente seria muito custoso ou, até mesmo, impossível pelas plataformas, devido às limitações na contratação de pessoal ou na impossibilidade de se treinar algoritmos para que façam um trabalho adequado, as próprias plataformas têm histórico, durante o desenvolvimento de seus sistemas, de atuação conjunta entre algoritmos e seres humanos para correção de rota quanto ao direcionamento de conteúdo.

Durante os anos 2000, por exemplo, o *Google* percebeu a atividade de produtores de conteúdo que se utilizavam de determinados termos-chave populares para produzir material com pouca ou nenhuma pesquisa, mas que teria bastante acesso. No início dos anos 2010, o próprio *Google* admitiu ter um problema com esses produtores, que chamou de “fazendas de conteúdo”, atuando através da contratação de revisores humanos para colaborar com o algoritmo para adequar os resultados esperados pelo *design* da plataforma. Neste caso, a plataforma optou por realizar um juízo de valor, determinando quais conteúdos mereciam ter mais disseminação do que outros. O mesmo pode ser apontado sobre o combate à disseminação de *fake news* e discurso de ódio. Nesse sentido, as plataformas já exercem supervisão sobre o conteúdo que é veiculado em seus programas, a despeito de se fazerem passar por mecanismos isentos (Pasquale, 2017b, p. 13-14).

É importante destacar que medidas de combate à disseminação de conteúdos danosos, especialmente *fake news*, têm o potencial de prejudicar o exercício ao direito fundamental da liberdade de expressão, do livre desenvolvimento da personalidade individual, a manifestação artística e intelectual, a produção de conteúdo satírico, religioso, ficcional, literário e qualquer outra forma de manifestação cultural. Assim, mecanismos de regulação devem respeitar tais liberdades, sob pena de gerar, ao revés, no anseio de proteger a liberdade de expressão da produção de conteúdo enganoso, fragilização deste preceito fundamental. No que diz respeito especificamente à disseminação de conteúdos enganosos, a exclusão e retirada de conteúdo ou perfis é medida mais gravosa, que pode implicar em excessividade, de forma que a opção por estratégias diversas, como providenciar que, em publicações consideradas falsas, sejam incluídos *links* com informações verdadeiras, por exemplo, ou com o impulsionamento de conteúdos verdadeiros, em detrimento de conteúdos reputados falsos, podem ser importantes estratégias complementares a serem adotadas (Maranhão; Campos, 2019, p. 322).

Ao se estabelecer um sistema de responsabilização das plataformas digitais sobre o conteúdo postado por terceiros, sem balizas jurídicas que delimitem o terreno de atuação das plataformas digitais, cria o risco de gerar censura colateral, que ocorre quando a norma busca incentivar que um agente, no caso as plataformas digitais, ao ser responsabilizado pelo conteúdo disponibilizado em seus serviços, aja de maneira proativa para bloquear, restringir ou excluir conteúdo de um terceiro, o que pode gerar a tendência de as empresas, para evitarem a responsabilização, agirem de maneira excessivamente zelosa, implicando na exclusão, permanente ou provisória, de conteúdo que não é irregular. Por meio da exclusão administrativa dos conteúdos, exercida pelas plataformas, as garantias do devido processo legal exigidas para a responsabilização por determinado discurso são, em geral, suprimidas, de forma que o conteúdo moderado passa por uma espécie de censura prévia incompatível com as liberdades civis.

Os sistemas de governança que atuam por meio da regulação indireta buscam criar incentivos e desincentivos, de acordo com a pretensão do legislador. Desta feita, é de fundamental importância, no momento da elaboração da norma, que o legislador tenha em mente quais são os objetivos pretendidos e quais incentivos a norma estabelecerá para privilegiar determinados comportamentos. Alguns casos são sintomáticos de efeitos secundários da norma, não previstos ou pretendidos, que criam incentivos diferentes daqueles previstos pelo legislador, como o caso da Coreia do Sul em que a legislação menciona “ação temporária” da intermediária *online* na indisponibilização do conteúdo postado por terceiros, o que incentivou as plataformas a remover o conteúdo após a notificação, seja tal conteúdo legal

ou não, como uma forma de proteger-se da responsabilização. O caso da Índia é semelhante, tendo a Suprema Corte sido provocada a decidir sobre as normas que incentivam as plataformas a excluir quaisquer conteúdos, quando da notificação, independentemente do seu conteúdo (Gasser; Schultz, 2015, p. 7).

Assim, a objeção feita à adoção da responsabilização das plataformas digitais diz respeito ao arcabouço regulatório necessário para instituir um sistema eficiente e robusto, o qual poderia ser deturpado para atingir interesses diversos, como promover censura, conforme reportado em casos de países como a Tailândia, China, Rússia e Turquia; para promover determinados interesses particulares e de lobistas, como o de grandes produtores de petróleo, suprimindo manifestações de grupos de preservação ambiental; ou mesmo que o sistema de regulação possa ser capturado pelas próprias plataformas, por meio de indicações dos membros de agências ou comissões de regulação. Apesar de serem preocupações legítimas, há pouca evidência que aponte que países com esferas públicas democráticas mais estabelecidas tenham a tendência de adotar medidas similares a de países com menor tradição democrática, além de existirem mecanismos para evitar a captura de órgãos de regulação, como o impedimento de que os membros atuem para empresas da área por algum período após deixar o cargo, pagamento de altos salários, compatíveis com a função e mecanismos de autofinanciamento para evitar a interferência política (Pasquale, 2017b, p. 14-15).

Mediante as características da comunicação digital sobre a qual discurremos anteriormente, estabeleceu-se um panorama da questão da liberdade de expressão e regulação das mídias digitais, problema fundamental do século XXI. Identifica-se, neste sentido, as seguintes características derivadas da comunicação digital: 1) que as plataformas atuam como fiduciárias das informações dos usuários; 2) que existem duas escolas de regulação, a tradicional e a nova, que são pautadas em realidades e interpretações distintas do princípio da liberdade de expressão; 3) que há uma relação de cooperação e cooptação público-privada; e 4) que as plataformas exercem uma governança privada dos discursos. Tendo em vista tais características, é preciso discutir, à luz do princípio da liberdade de expressão, quais práticas de governança privada são desejáveis, como os Estados devem empregar mecanismos da nova escola de regulação de discurso e demais implicações deste cenário (Balkin, 2018, p. 1209-1210).

2.3.2. Em busca do meio termo

Nesse sentido, Maranhão e Campos elaboram uma proposta de regulamentação, que parte de uma perspectiva restritiva sobre o conceito de *fake news*, limitando-se a conteúdos que se passam por material jornalístico. Propõem, assim, que devem ser instituídos parâmetros para o que são chamados provedores de redes sociais na *internet*, como por exemplo com o estabelecimento de um mínimo de usuários no país, que permitem compartilhamento de conteúdo entre os usuários e que não realizam controle editorial (em sentido estrito), excluídos provedores de conteúdo que sejam empresas jornalísticas, (Maranhão; Campos, 2019, p. 330), como uma forma de delimitar as empresas privadas que serão objeto da autorregulação regulada, ao mesmo tempo em que destaca-se que tais normas devem ser aplicar prioritariamente a plataformas com amplo impacto na comunicação social digital.

No que diz respeito às regras de *compliance*, destacam que deve ser previsto um prazo, por exemplo de vinte e quatro horas, para que as plataformas, mediante notificação do ofendido ou de terceiros, promovam medida informativa sobre conteúdo reputado fraudulento, restritos a conteúdos que se passem por material jornalístico, não impedindo sua visualização ou audição, mas informando, classificando e ordenando notícias que privilegiam informações jornalísticas sobre o tema, como uma forma de conferir eficácia, celeridade e legitimidade à estrutura de governança.

Ademais, deverá ser oportunizado ao usuário defesa e acesso ao procedimento adotado pela plataforma, sendo possibilitado ainda recurso da decisão, bem como encaminhamento à autoridade administrativa a ser instituída, como por exemplo a Comissão Interministerial das Plataformas Digitais na Internet- CIPLADI- de relatórios bimestrais sobre os procedimentos, equipes de trabalho, critérios e ferramentas tecnológicas adotados, assim como dados sobre o número de procedimentos e medidas tomadas, dados sobre a origem dos conteúdos reputados falsos e os valores gastos com propaganda e impulsionamento e o número de compartilhamentos do conteúdo antes e após a medida informativa.

A autoridade administrativa deve contar com a participação de membros do governo, empresas jornalísticas e provedores de redes sociais (art. 7º), com competência para encaminhar parecer técnico ao Ministério Público ou Justiça Eleitoral no caso de identificação de indícios de crimes, apurar infrações e certificar as entidades de autorregulação, vedando-se que a entidade atue de maneira a interferir no conteúdo decisório, especialmente no que diz respeito ao sentido de notícia fraudulenta ou medida informativa, limitando-se a atuação à fiscalização do cumprimento do procedimento previsto no projeto (art. 9º) (Maranhão; Campos, 2019, p. 331-333), de forma a instituir parâmetros, bem como atuar de maneira sistemática contra a

estrutura de desinformação e discurso de ódio que atua nas mídias digitais. A omissão na tomada das medidas informativas ou na entrega de relatório bimestral implicará em multa.

Nesse sistema, as entidades de autorregulação, organizadas pelas empresas de comunicação digital, deverão ser compostas por representantes de diversos setores da sociedade civil, inclusive representantes dos consumidores, da imprensa escrita, blogs eletrônicos, jornalismo audiovisual, empresas provedoras de conexão e aplicativos de *internet*, entidades acadêmicas e organizações não-governamentais, além de uma comissão de jornalistas profissionais que verificará as notícias fraudulentas, devendo contar, ainda, como auditoria independente (Maranhão; Campos, 2019, p. 333-334).

É preciso tomar cuidado, ainda, ao se instituir mecanismos de regulação, com o efeito da circularidade do poder, por meio do qual entidades economicamente poderosas utilizam deste fato para angariar poder político e, no sentido inverso, utilizam o poder político para reforçar seu poder econômico, em um processo cíclico. Caso tal estrutura se institua, as medidas de regulamentação terão efeito contraproducente, gerando a corrosão das instituições democráticas (Pasquale, 2017b, p. 15).

Além disso, medidas de regulação devem ser cuidadosamente elaboradas e objeto de observação, uma vez que até mesmo mecanismos bem-intencionados podem piorar os efeitos dos processos que buscam amenizar. Por exemplo, medidas voltadas para diminuir os efeitos da formação de filtros-bolha ou câmaras-de-eco, como por exemplo a apresentação de conteúdos com opiniões diversas daquela que o usuário possui, podem acabar gerando o efeito inverso, uma vez que não há como saber qual será a reação do público à exposição ao material. Tomando-se por base uma sociedade hipotética subdividida em dois lados, esquerda e direita, em que a direita está mais disposta a mudar de opinião e a esquerda está menos disposta, a exposição das duas opiniões para todos os indivíduos implicaria em que alguns indivíduos identificados com a direita tivessem aceitabilidade quanto a opiniões de esquerda, enquanto o oposto não ocorreria, agravando a polarização e consolidando o poder político de determinado grupo social (Pasquale, 2017b, p. 16-17).

Tendo em vista o potencial de mecanismos de regulamentação das plataformas digitais serem deturpados, é preciso que a sua implementação seja cuidadosa e realizada apenas em democracias estáveis. A necessidade de profissionalização e regulação da criação, curadoria e disseminação de conteúdo *online*, entretanto, tem ficado cada vez mais evidente, uma vez que a perspectiva utópica de que o ambiente digital tem o poder democratizante de conferir a qualquer indivíduo o potencial de acessar a esfera pública e participar de um debate livre é cada vez mais suplantada pela realidade constatada da atuação de agentes e do uso de milhões de

dólares em investimentos para produzir desinformação, promover determinados discursos ou interferir ilegalmente em eleições. É importante ter em vista, portanto, o papel das plataformas enquanto curadores dos conteúdos disseminados, atuando enquanto reguladores da comunicação global, a despeito de alegarem o contrário, e as discussões sobre mecanismos de regulação que sopesem adequadamente interesses comerciais e o interesse público são fundamentais para o futuro da democracia (Pasquale, 2017b, p. 18-20).

É importante destacar que mecanismos de combate à disseminação de *fake news* e discurso de ódio devem ser objeto de amplo debate social e legislativo, bem como constante aprimoramento, para que superemos a atual opacidade do sistema de comunicação digital, que é pautado apenas em interesses privados do mercado de comunicação, em detrimento do interesse público, de forma que faz-se fundamental que sejam instituídos mecanismos com transparência, tendo em vista o balanceamento entre a liberdade de expressão e a necessidade de evitar os efeitos negativos da veiculação de *fake news* e discurso de ódio.

A governança exercida por plataformas digitais também tem ocasionado problemas procedimentais, como a supressão do devido processo, falta de notificação e transparência nas decisões tomadas pelo ente privado. Assim, embora os usuários concordem com as políticas de uso das plataformas e conseqüentemente a instituição de mecanismos de governança, para garantir a segurança da comunidade, a ausência de normas procedimentais adequadas gera graves prejuízos à liberdade de expressão. Decorre, portanto, que as plataformas, ao exercerem governança em conformidade com o princípio da liberdade de expressão, devem atender a critérios de transparência, notificação e justo processo; informar os motivos das decisões e de mudanças nas políticas; disponibilizar meios de os usuários demandarem mudanças nas políticas; e permitir, ainda que de forma mínima, que os usuários participem do processo de governança (Balkin, 2018, p. 1196-1198).

A insatisfação dos usuários, entendidos enquanto uma comunidade com capacidade de influenciar nas escolhas das plataformas digitais, pode gerar diversas reações, como a saída da plataforma ou a exigência por mudanças, acreditando na possibilidade de melhorias. Os usuários também atuam notificando as plataformas sobre os conteúdos que entende descumprirem as normas da comunidade, exercendo uma espécie de trabalho não remunerado no processo de moderação do conteúdo, influenciando a maneira como este funciona. Geralmente, os custos sociais de deixar uma plataforma digital implicam em que os usuários optem por esperar ou exigir mudanças, mantendo-se conectados. Entretanto, mesmo quando optam pela exclusão destas redes, os usuários podem estar sujeitos a que os dados que já haviam sido compartilhados sejam mantidos pelas plataformas e até mesmo comercializado para

terceiros, como estudamos anteriormente o que, de certa forma, implica em uma mitigação da possibilidade de deixar esta rede. Caso a comunidade perca a confiança em uma plataforma, esta deixa de ser utilizada massivamente, motivo pelo qual manter uma boa relação com os usuários é fundamental para as plataformas digitais. A maneira como as plataformas adotam as medidas exigidas pela comunidade, entretanto, não é democrática, uma vez que não implica em participação efetiva ou organizada da comunidade, de forma que as decisões são tomadas de maneira unilateral, pelas próprias plataformas, e sua implementação é burocrática e técnica, dificultando a compreensão de seu funcionamento (Balkin, 2018, p. 1198-1201).

Tendo em vista que delegar às plataformas a responsabilidade de, utilizando as mesmas técnicas derivadas da mídia tradicional, realizando a avaliação caso-a-caso dos conteúdos a serem removidos, decorre em uma dificuldade procedimental quanto ao vasto volume de conteúdo veiculado, bem como a questões de legitimidade de uma empresa privada exercer moderação de conteúdo, além da possibilidade de, tendo em vista tentar evitar a responsabilização, as empresas optarem por indisponibilizar conteúdos legais, é fundamental investir em novas ferramentas que se tornaram viáveis na comunicação digital que, todavia, eram inviáveis nas mídias tradicionais, para ajudar no combate à disseminação de conteúdos prejudiciais, tais como: 1) informar, por meio de “etiquetamento”⁷ do conteúdo, que este é falso, incorreto, ou discurso de ódio, por exemplo, sem necessariamente excluí-lo ou diminuir seu alcance; 2) mediante a notificação dos usuários e o apontamento reiterado de que determinado conteúdo é potencialmente prejudicial, pode haver a indisponibilização temporária ou permanente do conteúdo; 3) modular a maneira como o conteúdo é veiculado e seu alcance; 4) adicionar informações ao conteúdo. Todos esses mecanismos são possibilitados pela arquitetura das plataformas digitais e podem constituir um novo sistema de moderação de conteúdo, pautado em uma espécie de colaboração dos usuários para apontar conteúdos que são potencialmente prejudiciais. Assim, em contraposição ao sistema criado para a moderação de conteúdo nas mídias de massa tradicionais, pautado na dualidade entre um material poder ser veiculado ou não, um sistema pautado na comunicação digital pode flexibilizar essa lógica (Hartmann, 2020, p. 6).

Em 2004 um *site* que questionava a veracidade do holocausto contra o povo judeu estava entre os dez principais resultados da pesquisa do termo judeu no mecanismo de pesquisa do *Google*. Quando questionada, a empresa colocou um aviso sobre o funcionamento do mecanismo na sua página, alegando que não era responsável por este. Para tentar minimizar os

⁷ O etiquetamento pode ser realizado de diversas maneiras. Hartmann, 2020, p. 5, refere-se ao uso de *red flag*, ou seja, uma bandeira vermelha, que indique que aquele conteúdo é potencialmente problemático.

efeitos deste tipo de direcionamento, sem que, contudo, haja a necessidade de excluir conteúdo ou *link*, no caso de discurso de ódio explícito, uma possibilidade aventada por legisladores, especialmente na Europa, é a estratégia de *labelling*, ou seja, indicar que aquele conteúdo contém discurso de ódio ou desinformação, disponibilizando, na mesma página, *links* que direcionem para páginas com informações de instituições confiáveis sobre o tema. A estratégia já é adotada em relação aos mercados de comida e remédios nos EUA, por exemplo, não representando óbice ao princípio da liberdade de expressão. De fato, grande parte das plataformas digitais atuam em uma espécie de “mercado de informação”, no qual não estão oferecendo produtos e serviços diretamente, porém influenciam decisivamente, por meio do alcance que conferem a determinado conteúdo, quais empresas ou grupos obtém sucesso (Pasquale, 2017b, p. 7-8).

Assim, algumas estratégias alternativas no combate à disseminação de *fake news* e discurso de ódio, não necessariamente excludentes, mas complementares, para evitar o recurso às estratégias punitivistas e criminais, são: a checagem de fatos; mudanças nos algoritmos das plataformas digitais; organização de grupos especializados para oferecer respostas rápidas em períodos eleitorais; e a oposição de informações verdadeiras às informações falsas veiculadas (Macedo Júnior, 2020, p. 242-245). Conforme veremos no último capítulo, estas estratégias já foram colocadas em prática, no Brasil, durante as eleições de 2020 e 2022.

2.4. Fake news, discurso de ódio e o caso brasileiro

No Brasil, as *fake news* se tornaram especialmente relevantes no debate público durante as eleições gerais de 2018, especificamente no que diz respeito a disputa à Presidência da República. Na ocasião, a polarização entre o candidato de extrema-direita Jair Bolsonaro e o candidato de esquerda Fernando Haddad produziu e foi produzida pela formação de *clusters* (ou grupos/aglomerações) de usuários em torno de páginas consoantes com concepções pré-concebidas dos usuários. Assim, à esquerda, páginas ligadas ao feminismo, ao movimento negro e LGBTQIA+, à ONGs e grupos de direitos humanos e, à direita, páginas de liberalismo econômico e conservadorismo moral, bem como de políticos e personalidades identificadas com a respectiva posição no espectro político, constituíram grupos coesos no que diz respeito aos usuários que tinham acesso aos conteúdos veiculados digitalmente (Alves; Maciel, 2020, p. 155-156).

Combinada ao enviesamento dos conteúdos de acordo com as preferências dos usuários, houve a atuação de empresas que utilizam análise comportamental-psicológica e os dados dos usuários para formular os conteúdos a serem produzidos e disseminados. Assim, empresas como a *War Room* e a *CA-Ponte*, parceira da *Cambridge Analytica*, bem como empresas de disparo de conteúdo em massa no *WhatsApp*, representaram importante papel nas referidas eleições (Alves; Maciel, 2020, p. 156).

O impacto da disseminação de notícias falsas pode ser melhor compreendido observando-se os seguintes dados: das 123 checagens de fatos realizadas pelo site Congresso em Foco, 104 eram conteúdos prejudiciais ao candidato de esquerda Fernando Haddad; no caso do “Kit Gay”, que se trata de *fake news* que alegava que o candidato de esquerda pretendia distribuir materiais impróprios para escolas infantis, 85,2% dos eleitores de Bolsonaro receberam tal conteúdo, dos quais 83,7% acreditaram na veracidade, enquanto entre os eleitores de Haddad 61% receberam, dos quais apenas 10,5% acreditaram. Esses dados reforçam, por sua vez, tanto a atuação para maior disseminação do conteúdo para usuários com tendência a confiar nestes, quanto o viés de confirmação de acordo com as preconcepções dos usuários (Alves; Maciel, 2020, p. 156-157).

Ademais, o Brasil foi apontado como um dos países que mais acredita em conteúdos falsos disseminados digitalmente, alcançando cerca de 63% da população. Vários aspectos podem ser apontados para compreender o alto índice de confiança em conteúdos falsos no país, como o dado de que 7 em cada 10 brasileiros utiliza as plataformas digitais para se informar e, entre a população brasileira que recebe até três salários-mínimos, há prevalência do uso de smartphones como única fonte de acesso à internet. Assim, a existência de pacotes de *internet* móvel baratos com dados ilimitados para acesso a plataformas digitais, mas limites quanto a outros conteúdos, somada à desigualdade social e pobreza, favorece a prevalência das redes sociais como meio de informação, a despeito de sites de notícias ou mecanismos de busca, onde os usuários poderiam conferir com maior isenção e mais fontes a veracidade dos conteúdos (Alves; Maciel, 2020, p. 157-158).

No Brasil, portanto, as eleições de 2018 foram o momento em que, até então, houve maior atuação nas plataformas digitais para a divulgação de desinformação em massa. Com a proibição de financiamento de campanhas eleitorais por empresas, o apoio às campanhas migrou para outras searas, como a digital, em que foi financiada a divulgação de conteúdos com desinformação por meio de plataformas digitais.

A legislação tradicional para combater desinformação, como no caso de calúnia, difamação e injúria, demonstra-se insuficiente para confrontar o problema na *internet*. A Lei nº

9.504/1997, por exemplo, que enquadra ofensas na *internet* a candidatos, partidos ou coligações, foi elaborada tendo em vista as grandes corporações e, com as redes sociais, cujas principais características são a comunicação mais veloz e descentralizada, passou a ser ineficiente (Queiroz, 2022).

Pela caracterização do fenômeno das *fake news* e do discurso de ódio e sua disseminação por meio das plataformas digitais, demonstra-se a relevância de algum tipo de regulamentação sobre o fluxo de conteúdos digitais, respeitando-se a liberdade de expressão. Pela natureza do setor de tecnologia, trata-se de desafio transnacional, implicando em, ao menos, atuação das próprias plataformas na moderação de conteúdos e elaboração de legislação em cada país (Dourado, 2020, p. 109).

Desde 2016, com as eleições estadunidenses e o Brexit, as plataformas digitais têm sido pressionadas a agir nesse sentido. A União Europeia elabora um documento, denominado Código de Práticas sobre Desinformação, em que indica ações a serem tomadas pelas plataformas digitais, indicando 5 eixos de atuação:

“1. “Interrupção de receitas de publicidade de contas e sites que espalham informação falsa”; 2. “Tornar mais transparentes a publicidade política e a publicidade baseada em questões”; 3. “Enfrentar a questão das contas falsas e dos robôs *online*”; 4. “Capacitar os consumidores para relatar desinformação e recorrer a diferentes fontes de notícias, ao mesmo tempo que melhora a visibilidade e a facilidade de encontrar conteúdos confiáveis”; 5. Capacitar a comunidade científica para monitorar informação falsa *online* por meio do acesso aos dados das plataformas compatível com a política de privacidade adotada” (Dourado, 2020, p. 110).

Seria ingênuo considerar que o poder econômico e financeiro não usaria os meios de comunicação digital para a consecução de seus interesses. A disseminação de conteúdos falsos e discurso de ódio nas plataformas digitais tem em seu cerne o financiamento de propaganda e impulsionamento destes conteúdos, o que fica evidente pelo escândalo da *Cambridge Analytica*, que minerou dados e produziu informação hipersegmentada para influenciar o debate público. Por esse motivo o primeiro eixo de atuação proposto pelo documento é a interrupção de receitas de publicidade para sites que desinformam (Dourado, 2020, p. 110-111).

Enfrentar a questão da análise de *big data*, ou seja, de dados massivos dos usuários, é fundamental para a proteção do sistema democrático, uma vez que a produção segmentada de narrativas desloca o debate político da esfera pública para ser entregue sob demanda aos usuários das plataformas digitais. Algumas plataformas, como o *Twitter* e o *Google*, impediram o impulsionamento de conteúdos políticos, tendo em vista as eleições estadunidenses de 2020,

em busca de evitar o uso dos dados para este fim, enquanto outras, como o *Facebook*, continuam permitindo a prática (Dourado, 2020, p. 111).

No Brasil, o Marco Civil da Internet prevê a possibilidade de exclusão de conteúdo apenas após notificação judicial específica, que permite a localização inequívoca do conteúdo a ser indisponibilizado, e que o ofendido deve, em regra, apontar o perfil do ofensor. A regra é demasiadamente branda, implicando em grande fluxo de *fake news* e discurso de ódio no país. Estima-se que 12 milhões de contas difundem *fake news* sobre política no Brasil e que este material alcança praticamente todos os usuários de plataformas digitais (Dourado, 2020, p. 117-118).

Mediante o impacto das informações falsas veiculadas durante as eleições de 2018, diversas medidas passaram a ser adotadas para tentar minimizar a sua disseminação e conferir maior capacidade de compreensão sobre a falsidade do conteúdo entre os usuários das plataformas digitais. Já nas eleições de 2018, o Tribunal Superior Eleitoral estabeleceu o Conselho Consultivo sobre Internet e Eleições, que teve participação da Agência Brasileira de Inteligência, sociedade civil e exército; lançou o site “Esclarecimento sobre informações falsas”; organizou seminários para discutir o tema, em parceria com a União Europeia, com participação de empresas do setor como *Google*, *Facebook* e *WhatsApp*, especialistas do FBI, da Polícia Federal, da Organização dos Estados Americanos (OEA), do Poder Judiciário e da Sociedade Civil. No mesmo sentido, empresas como *Facebook* e *WhatsApp* tomaram medidas como parcerias com agências de verificação, como a Agência Lupa e Aos Fatos e o Projeto Comprova, que visa informar e educar para a comunicação digital (Alves; Maciel, 2020, p. 158-159).

Em dezembro de 2017 o TSE instaurou o Conselho Consultivo sobre Internet e Eleições e publicou a Resolução TSE nº 23.551 que destacou que a liberdade de expressão somente poderia ser limitada por ofensa à honra de terceiros e conteúdos manifestamente inverídicos. Em 4 de maio de 2018 o Conselho de Comunicação Social da Congresso Nacional elaborou relatório sobre as leis que tratavam sobre o tema das *fake news*, indicando que toda legislação deveria repelir qualquer forma de censura e, ainda em maio de 2018, foi instituída Frente Parlamentar Mista cujo objetivo era aprovação uma legislação antes das eleições daquele ano. No início de junho de 2018 o TSE convocou os partidos e empresas de tecnologia a assinar termos de compromisso, cujo um dos objetivos era combater as *fake news*. Em 7 de junho o Tribunal determinou a exclusão de conteúdo publicado pela página Partido Anti-PT no *Facebook*, com 1,7 milhões de seguidores, por se tratar de *fake news* contra a candidata Marina Silva (Dourado, 2020, p. 118-119).

Neste período do meio do ano de 2018 a questão estava em voga no país. Luiz Fux, presidente do TSE à época, chegou a afirmar que a Justiça Eleitoral exerceria poder de polícia contra as *fake news*; que a Agência Brasileira de Inteligência, o Exército e a Política Federal seriam chamados a atuar de maneira preventiva; e que, no limite, o uso de *fake news* na campanha eleitoral poderia implicar em anulação das eleições, com fundamento no art. 222 do Código Eleitoral. Diversas outras medidas foram tomadas ao longo do período pré-eleitoral, como recomendação do Ministério dos Direitos Humanos, exclusão de conteúdos por ordem da Justiça Eleitoral e um memorando de entendimento, assinado por plataformas digitais *Facebook* e *Google*, para prevenção da disseminação de notícias falsas (Dourado, 2020, p. 119-120).

Apesar das medidas descritas e outras tomadas, as eleições de 2018 no Brasil foram consideradas pela Organização dos Estados Americanos amplamente manipulada por meio de *fake news*. Foi constatado, por exemplo, o envio de mensagens anti-PT em até 300 mil grupos favoráveis ao candidato Jair Bolsonaro. Pesquisa indica que afirmações falsas, como a de que o candidato Fernando Haddad havia criado o kit-gay e defendido o incesto em um livro, eram tratadas como verdadeiras por 36% e 15% dos eleitores, respectivamente. Após as eleições, ainda, o *WhatsApp* anunciou o bloqueio temporário de contas que tiveram comportamento inadequado, inclusive de Flávio Bolsonaro, filho do candidato à presidência vencedor em 2018 (Dourado, 2020, p. 121-122).

Outras estratégias têm sido elaboradas pelas plataformas. O *Facebook*, por exemplo, efetivou parcerias com agências de verificação de fatos independentes, inserindo alertas sobre a inverdade das publicações. Anunciou, também, que irá indisponibilizar provisoriamente conteúdos nos quais acredita-se haver falsidade, enquanto aguarda a verificação pelas agências. Atuando também mediante a notificação dos usuários, a empresa tem atuado no sentido de tentar coibir infratores reincidentes, retirando direitos de publicidade, por exemplo. Ademais, informa que trabalha na criação de inteligência artificial para automatizar a exclusão de *fake news*, a qual atuaria na exclusão de conteúdos que já foram identificados enquanto falsos pelas agências de verificação, mas que continuam sendo compartilhados. Tendo em vista a proteção às eleições estadunidenses de 2020 a plataforma ainda informou que desarticulou 50 redes que atuavam por meio de contas falsas para influenciar o pleito, estabeleceu mecanismos para proteger contas e representantes políticos e suas equipes contra ações *hackers*, e informou que passará a informar a organização responsável por páginas com grande quantidade de seguidores, além de informar quando a página é vinculada ao Estado, ainda que parcialmente (Dourado, 2020, p. 112-114).

Reconhecendo a insuficiência dos instrumentos atualmente utilizados, o Tribunal Superior Eleitoral tem adotado diversas medidas voltadas ao combate à desinformação nas eleições de 2022, em especial por meio do Programa Permanente de Enfrentamento à Desinformação da Justiça Eleitoral, criado pela Portaria TSE nº 510/2021. Uma das principais ações do Tribunal foi a convocação das plataformas digitais para a assinatura de acordos para atuarem em conjunto no combate a desinformações, moderando conteúdos, oferecendo sistemas de checagem e limitando o repasse de mensagens (Queiroz, 2022).

Ademais, o Congresso Nacional está discutindo, em texto aprovado pelo Senado Federal em 2020, mas sem aprovação da Câmara dos Deputados, o Projeto de Lei das *Fake News*, ou Lei Brasileira de Liberdade, Responsabilidade e Transparência na *Internet*, principal projeto sobre a regulamentação da comunicação digital. No bojo do sistema proposto neste projeto de lei estão quatro estratégias regulatórias: a) a principalização, por meio da qual o legislador se exime de especificar condutas ilícitas, gerando indeterminação; b) cognitivização, criando um sistema pautado em conhecimentos especializados que, no caso de tecnologias de informação, são altamente complexos e dominados por empresas globais; c) periferização, no mesmo sentido do sistema jurídico, o sistema político busca delegar a definição de conceitos para a periferia do sistema político, onde estão acadêmicos, *experts*, movimentos sociais e grupos de interesse; d) procedimentalização, por meio da qual se institucionalizam fontes de direito não estatais e autorregulatórias (Amato; Saba; Loschiavo Leme de Barros, 2022, p. 553-555).

No Projeto de Lei das *Fake News* encontra-se a estratégia regulatória da procedimentalização. No art. 31 do referido Projeto de Lei, há a instituição da chamada autorregulação regulada, por meio da qual as intermediárias *online* poderiam criar instituição de autorregulação, voltada à transparência e à responsabilidade do uso da *internet*, mediante certificação do Conselho de Transparência e Responsabilidade na Internet, a ser criado. Tal método corresponde ao chamado direito reflexivo, programas relacionais ou procedimentalização, ou seja, por meio do qual o ente público cria parâmetros procedimentais para a tomada de decisão da iniciativa privada (Amato; Saba; Loschiavo Leme de Barros, 2022, p. 556).

Para além da mera instituição de princípios, a instituição da autorregulação regulada é um caminho promissor para a questão das *fake news*. O contato entre o direito estatal e o direito autorregulado poderia propiciar melhor troca de informações entre os dois sistemas, implicando em melhor tradução dos critérios técnicos para critérios jurídicos, colaborando com a dificuldade de determinação de conceitos presente nos instrumentos jurídicos, nesse caso derivado da impossibilidade de se engessar um conceito de *fake news*, por exemplo. Esse

processo coincide com a busca das intermediárias *online* pela legitimidade para moderar os conteúdos, tomando para si a responsabilidade de tratar dos abusos do uso das plataformas digitais de sua propriedade, ao mesmo tempo em que o direito estatal atua enquanto baliza reguladora (Amato; Saba; Loschiavo Leme de Barros, 2022, p. 556-558), para evitar os abusos das plataformas.

O trabalho de moderação dos conteúdos que são veiculados nas plataformas digitais não é novidade, sendo exercido por diversas empresas de acordo com seus termos e condições, aos quais o usuário adere quando cria um perfil na plataforma. Assim, empresas como *Facebook* e *Twitter* atuam na exclusão de conteúdos considerados irregulares na plataforma, como discurso de ódio e violência, suicídio e pornografia. O *Twitter* informa, ainda, que só atua mediante notificação dos usuários, ou seja, pautando-se em mecanismo de notificação e exclusão de conteúdo, e que há preferência pela exclusão de contas associadas à disseminação de conteúdo contrário a seus termos. Outras plataformas, como o *WhatsApp*, não exercem qualquer tipo de moderação nas mensagens trocadas entre os usuários, ainda que em grupos públicos, sob a alegação de que a criptografia ponta-a-ponta impede que até mesmo a plataforma tenha acesso aos conteúdos compartilhados entre os usuários (Dourado, 2020, p. 111-112).

A moderação de conteúdos, entretanto, traz à luz o princípio da liberdade de expressão, uma vez que determinados conteúdos podem ser indisponibilizados. Para conferir legitimidade à atividade de moderação das plataformas, deve haver, no mínimo, entre outros critérios, transparência quanto à forma de análise de exclusão de conteúdos, de forma que todos os usuários possam questionar a exclusão. Inclusive, se há uso de algoritmo para a moderação de conteúdos, entende-se que este deve ser publicizado, ou ao menos os critérios utilizados, de forma a permitir a verificação pelos usuários e demais interessados (Dourado, 2020, p. 114).

É importante destacar que a atuação do legislador está frequentemente pressionada por escândalos recentes que influenciam a opinião pública: na tramitação do Marco Civil da Internet era o Caso Snowden; da Lei Geral de Proteção de Dados Pessoais, o caso da *Cambridge Analytica* e a entrada em vigor da *General Data Protection Regulation* da União Europeia; e o Projeto da Lei das *Fake News*, o caso de disparo massivo de desinformação nas eleições de 2018 e 2020 (Amato; Saba; Loschiavo Leme de Barros, 2022, p. 555).

Alguns países têm desenvolvido legislação para combater a disseminação de discurso de ódio e *fake news* nos meios digitais. A Alemanha aprovou a previsão de multas às plataformas digitais, que podem alcançar 50 milhões de euros, caso estas permitam a disseminação destes conteúdos; a França previu a possibilidade de exclusão imediata de *fake news* mediante determinação judicial; a Malásia criminalizou a distribuição de *fake news*,

prevendo penas de prisão de até 6 anos e multa de mais de 100 mil dólares; Singapura conferiu aos ministros do governo a competência para decidir sobre conteúdos considerados falsos, permitindo multas de mais de 700 mil dólares às plataformas digitais pelo descumprimento de decisões estatais e até 10 anos de prisão para sujeitos identificados; o Vietnã também conferiu ao governo o poder de determinar a exclusão de conteúdos que gerem danos à segurança nacional e ordem social; na Rússia, o governo pode até mesmo restringir o tráfego da *internet*, filtrando conteúdo; nos EUA as principais discussões sobre o tema são acerca da influência de países estrangeiros em assuntos internos. Verifica-se, portanto, que em alguma medida os países têm sopesado o direito à liberdade de expressão, com suas significações e distinções entre as diversas experiências, com a necessidade de enfrentar a disseminação de *fake news* e discurso de ódio nas plataformas digitais (Dourado, 2020, p. 114-117).

A Lei alemã, chamada *Network Enforcement Act*, ou *Net-zDG* entrou em vigor em janeiro de 2018 e regulamenta a disseminação de conteúdos que violem o código penal alemão em plataformas com mais de dois milhões de usuários, estabelecendo limites de 24 horas ou uma semana, a depender da complexidade do caso, para a exclusão de conteúdo ilícito, sob pena de multas às plataformas digitais de até 50 milhões de euros. A lei estabelece obrigações de transparência às plataformas, como informar os usuários sobre conteúdos que forem indisponibilizados e a responsabilidade de armazenar os conteúdos por dez semanas, para produção de provas. O curto prazo para análise dos conteúdos tem sido apontado como uma das principais críticas ao sistema, uma vez que poderia implicar em que as plataformas, para evitar a possibilidade de serem responsabilizadas, poderiam excluir conteúdo potencialmente legítimo (Alves; Maciel, 2020, p. 153-154).

Já a lei malaia criminalizou a publicação de *fake news*, definindo o conceito enquanto qualquer notícia, informação, dados ou relatório parcial ou totalmente falsos, publicados de maneira digital ou física por órgãos de imprensa, civis, nacionais ou estrangeiros, com multas de até 122 mil dólares e 6 anos de prisão. As principais críticas quanto a este sistema são de duas naturezas: quanto à efetividade do sistema criminal para lidar com o tema e o uso enquanto meio de censura, tendo em vista o contexto político em que a norma foi aprovada. Importante ressaltar que há relatos do uso da norma como forma de censura por parte do governo malaio (Alves; Maciel, 2020, p. 154-155).

À exemplo das normativas internacionais, pode-se observar dois graves problemas na criminalização da disseminação de *fake news*: o primeiro que, por meio do sistema criminal, não se adotam medidas estruturais para lidar com o problema da desinformação *online*, mas apenas criam-se tipos penais vagos que poderiam, inclusive, imputar a qualquer cidadão o

cometimento de crime pelo mero compartilhamento de notícia que acreditava ser verdadeira; o segundo grave problema constatado em projetos no sentido da criminalização da disseminação de *fake news* é que estes podem ter como consequência graves prejuízos à liberdade de expressão, por meio da tomada de decisões arbitrárias na identificação de quais conteúdos são falsos e quem são os sujeitos ativos (Alves; Maciel, 2020, p. 161-162). Assim, seja pela possibilidade de desvirtuamento para a censura, quanto pela ineficácia, o sistema de criminalização parece ser incompatível com o Estado Democrático de Direito.

Já quanto ao sistema de responsabilização das plataformas digitais sobre o conteúdo postado por terceiros, cujo principal expoente é o alemão, por meio da *Net-zDG*, as principais preocupações residem em dois pontos distintos: primeiro que o curto prazo para a tomada de decisão da exclusão de conteúdo pode resultar em indisponibilização de conteúdos lícitos, tendo em vista a intenção das plataformas de não serem responsabilizadas. Os prazos da lei alemã, por exemplo, de 24 horas ou até uma semana, a depender da complexidade do caso, são considerados pelo Relator Especial das Nações Unidas para a Proteção da Liberdade de Expressão, David Kaye, exíguos, e os critérios, vagos e imprecisos. A segunda preocupação apontada é a ampliação do poder das plataformas digitais para determinar quais conteúdos devem ser excluídos, ou seja, exercer censura colateral, tendo em vista que estas empresas já são muito poderosas e constituem verdadeiros monopólios. O *Google*, por exemplo, é dono do *YouTube*, *Android*, *Waze*, *Chrome* e o próprio mecanismo de pesquisa homônimo, enquanto a *Meta* é dona do *Facebook*, *WhatsApp*, *Instagram* e *Messenger* (Alves; Maciel, 2020, p. 162-163).

Desta feita, um sistema de responsabilização das plataformas digitais privilegia o tratamento sistemático da questão das *fake news*, entretanto deve ser dotado de mecanismos de transparência, governança e aperfeiçoamento perenes, até mesmo em virtude da possibilidade de reavaliação judicial da decisão administrativa de exclusão ou não de conteúdos por parte das plataformas digitais. Ao mesmo tempo, uma definição clara e objetiva do conceito de *fake news* é imprescindível para possibilitar a tomada de decisão caso-a-caso, tanto por parte das empresas quanto do Judiciário, por mais que seu engessamento seja inviável. Medidas mais restritivas devem ser excepcionais e a participação dos usuários e transparência das plataformas é indispensável.

É preciso ter em mente que os impactos da desinformação na sociedade democrática são amplos e que, seja qual for a abordagem escolhida, não abrangerá todos os aspectos desse complexo fenômeno. Assim, medidas complementares e aprimoramentos constantes são a tônica de políticas que visem melhorar a comunicação digital e promover a democracia,

associado à participação de todos os entes envolvidos no complexo emaranhado da comunicação digital, quais sejam a sociedade civil, os Estados, os setores econômico, acadêmico e técnico (Alves; Maciel, 2020, p. 165-166).

No Brasil, o Marco Civil da Internet protege as plataformas de responsabilização pelos conteúdos postados por terceiros, ressalvadas ordens judiciais para indisponibilização de conteúdo. Cabe ao Judiciário, assim, provocado pelas partes, decidir caso a caso se determinado conteúdo constitui *fake news* ou discurso de ódio, determinando ou não sua remoção. A jurisprudência recente dos Tribunais não permite encontrar unicidade, de forma que em temas semelhantes, há decisões divergentes. Durante as eleições de 2018, por exemplo, o Tribunal Superior Eleitoral decidiu sobre conteúdos postados em desfavor dos então candidatos à presidência Marina Silva e Ciro Gomes. No primeiro caso, o Tribunal determinou a exclusão de materiais que associavam a candidata a esquemas de corrupção, enquanto no segundo, indeferiu o pedido, sob o fundamento de que para se constituir enquanto *fake news* a inveracidade do material deve ser conferível de plano. O mesmo Tribunal também criou o Conselho Consultivo sobre Internet e Eleições para instituir mecanismos de combate à disseminação de *fake news* durante o período eleitoral que adotou medidas mais restritivas, da mesma forma que há projetos de lei em tramitação que propõem criminalizar o compartilhamento de conteúdo falso. Assim, a importância do uso das mídias digitais para influenciar o debate público tem suscitado severas iniciativas de regulamentação, que podem, potencialmente, ser prejudiciais à liberdade de expressão (Macedo Júnior, 2020, p. 242-244).

Nesse sentido, o Tribunal Superior Eleitoral atuou ao longo das eleições de 2018, mediante provocação, na proteção à integridade do processo eleitoral, julgando casos de conteúdos que ofenderam a honra, imagem, intimidade e que continham desinformação, com fulcro no art. 57-D da Lei nº 9.504/1997, chamada Lei Eleitoral. Apesar de a Constituição Federal permitir que se estabeleçam limites à liberdade de expressão, estes devem ser cuidadosamente identificados, para interferir o mínimo possível no debate democrático, como prevê o art. 33 da Resolução nº 23.551/2017. Desta maneira, o peticionante, ao alegar que determinado conteúdo se trata de *fake news*, deve apontar especificamente sobre qual/quais materiais requer a revisão judicial, indicando inclusive a *URL*, sob pena de nulidade. Durante as eleições de 2018, foram propostos 448 processos que versavam sobre propaganda eleitoral, dos quais 10% eram sobre *fake news*, julgados por três Ministros do Tribunal Superior Eleitoral, Ministro Luis Felipe Salomão, Ministro Sergio Banhos e Ministro Carlos Horbach. O primeiro adotou uma visão mais abrangente do princípio da liberdade de expressão, garantindo apenas o direito de resposta nas plataformas digitais, enquanto os dois últimos determinaram, além do

próprio direito de resposta, remoção de conteúdo com informações notadamente falsas e com ampla veiculação, de forma a ter potencial para afetar o equilíbrio eleitoral. (Horbach, 2022, p. 199-200).

Assim, com base na Lei Eleitoral e no Código Eleitoral, o Tribunal Superior Eleitoral atuou em conformidade com a Constituição Federal e a interpretação do princípio da liberdade de expressão na tradição jurídica brasileira, determinando medidas de combate à disseminação de *fake news* e discurso de ódio, entre elas a remoção de conteúdo, a proibição de acesso às aplicações e a responsabilização por promover ou se beneficiar de campanhas de desinformação, buscando respeitar, ainda, o princípio da mínima intervenção, tendo em vista a proteção ao princípio da liberdade de expressão (Horbach, 2022, p. 200-201).

É importante acrescentar, ainda, que o legislador brasileiro instituiu, por meio do §1º do art. 19 e do art. 21 do Marco Civil da Internet, que preveem que as plataformas são responsabilizadas pelo conteúdo postado por terceiros, mediante notificação, no que violar direitos autorais ou conteúdos com cenas de nudez ou atos sexuais de caráter privado, o chamado *revenge porn*, limitações à liberdade de expressão para a comunicação digital. Tais exceções, ao serem interpretadas conjuntamente com o caput do art. 19, que dispõe sobre a isenção de responsabilidade das plataformas digitais sobre o conteúdo postado por terceiros, com o objetivo expresso de “assegurar a liberdade de expressão e impedir a censura”, demonstram, especificamente no caso de conteúdos veiculados por meio de mídias digitais, a possibilidade de limitação ao princípio da liberdade de expressão, em razão de outros interesses e garantias, inclusive privados. Ou seja, a legislação brasileira já permite, e mais, exige, sob pena de responsabilização, a atuação de uma empresa privada na moderação de conteúdo sem que, entretanto, haja ofensa ao princípio da liberdade de expressão.

A regulação da comunicação digital adquire especial relevância mediante o impacto que a circulação de determinados conteúdos nas novas tecnologias de comunicação e informação digital tem na esfera pública, ou seja, na forma como veiculam a comunicação política e social dentro das democracias liberais pós-modernas. Uma das principais dificuldades quanto ao tema, entretanto, deriva da velocidade com que se promovem modificações no meio da comunicação digital. Os sistemas de governança da comunicação digital, portanto, devem ter como ponto de partida o reconhecimento do impacto causado pelas tecnologias e da especialidade técnica afeita ao tema.

A autorregulação regulada, enquanto estratégia normativa, busca convergir as características 1) da autorregulação, que tem por mérito reconhecer que a complexidade técnica do funcionamento das plataformas digitais torna tais agentes fundamentais para a estruturação

de um sistema funcional, possibilitando uma constante revisão de conceitos, que coaduna com o ambiente tecnológico, mas que, por outro lado, tem como revés o fato de que o particular, não necessariamente, irá fundamentar suas decisões em função de interesses e valores públicos e 2) da regulação por terceiros, no caso o Estado, que apesar de não ser dotado do conhecimento técnico característico do ambiente tecnológico, tem o condão de exercer coerção ao particular em função do interesse público (Maranhão; Campos, 2019, p. 323-324).

Tendo em vista o fundamento e compatibilidade constitucional, as estruturas de autorregulação regulada não deveriam, necessariamente, ser impostas pelo Poder Público, mas sim que reconhecidas como meio de *compliance* com a legislação vigente, mediante opção das empresas privadas, consagrando, a um só tempo, os mandamentos constitucionais da liberdade de associação e da iniciativa empresarial e a atuação do Estado enquanto regulador da atividade econômica, este último com fulcro no art. 174 da Constituição Federal de 1988. Essas estruturas de *compliance* parecem ser um mecanismo mais adequado ao combate à disseminação de conteúdos falsos, uma vez que tornariam possível a identificação de organizações que atuam de maneira irregular na produção e disseminação destes conteúdos, ou seja, na busca de soluções estruturais, o que tem se mostrado inviável por meio da solução judicial caso-a-caso vigente sob o Marco Civil da Internet (Maranhão; Campos, 2019, p. 326-327).

Desta forma, conclui-se que uma possibilidade de regulamentação que se impõe no presente momento, de reconhecimento da estrutura da comunicação digital, é a adoção da autorregulação regulada, que encontra fundamentação constitucional nos dispositivos mencionados, especificamente no art. 220, §3º, II, ao exigir-se que as plataformas, portanto empresas privadas, privilegiem a veiculação de conteúdos informativos corretos, em detrimento de desinformação, uma vez que tal estrutura, apesar de ser um “embaraço” à veiculação de informação, tem o condão de informar os indivíduos para que possam, adequadamente, se defenderem de conteúdo potencialmente nocivo, nos exatos termos da Constituição Federal, com a ressalva de que medidas mais restritivas devem ser excepcionais, tendo em vista o potencial lesivo à liberdade de expressão (Maranhão; Campos, 2019, p. 329-330).

CAPÍTULO 3- MAPEAMENTO DE ARRANJO-JURÍDICO INSTITUCIONAL: A REGULAMENTAÇÃO DA COMUNICAÇÃO DIGITAL NO BRASIL A PARTIR DO MARCO CIVIL DA INTERNET

Ao longo do segundo capítulo, desenvolvemos um raciocínio jurídico que, através do reconhecimento da estrutura socioeconômica que pauta o mercado das mídias digitais, demonstra que a possibilidade de regulamentação da comunicação digital não obsta, por si só, a efetivação do princípio da liberdade de expressão e que, ao contrário, ao vigiar incessantemente e direcionar conteúdo a partir da análise psicométrica, as chamadas *big techs* realizam uma nova forma de ataque à liberdade de expressão, que impacta a autonomia e autodeterminação dos indivíduos, convertidos em usuários, de maneira que, nesse ponto, a regulamentação se mostra não apenas legal e constitucionalmente justificada, como social e democraticamente desejável.

Identificamos, assim, que no final dos anos 2010 houve um amplo reconhecimento, especialmente por parte de Estados e Organizações Internacionais, mas também por parte da sociedade civil, academia e das próprias plataformas digitais, de que alguma forma de regulamentação é imprescindível. De fato, verificamos que as próprias plataformas realizam governança privada dos conteúdos postados por terceiros. O recurso à bibliografia jurídica que discorre sobre as experiências de regulamentação adotadas mundo afora, especialmente no que diz respeito a medidas similares àquelas que vem sendo tomadas ou aventadas no Brasil, deixou claro que não se trata de tarefa das mais fáceis estabelecer um sistema de moderação de conteúdos que ao mesmo tempo diminua a disseminação de conteúdos perniciosos, como *fake news* e discurso de ódio, e esteja em restrito cumprimento ao princípio constitucional da liberdade de expressão.

3.1. Algumas palavras sobre o método

Neste terceiro capítulo, recorreremos à aplicação sistemática do método de mapeamento de arranjos jurídico-institucionais, calcado na revisão bibliográfica realizada nos capítulos anteriores, por meio do qual através da observação do plano mesoinstitucional buscamos identificar de que maneira a regulamentação da comunicação digital se dá no Brasil, tendo como marco temporal inicial o Marco Civil da Internet, identificando os pontos de inflexão que redirecionaram o arranjo; a participação dos diversos atores envolvidos no processo, entre eles o governo, a sociedade civil, o judiciário, os usuários e as plataformas digitais; a motivação e

os efeitos esperados dos arranjos adotados; e de que forma o arranjo jurídico-institucional impactou na moderação de conteúdos nas mídias digitais (ao gerar incentivos para a atuação do particular, especialmente).

O método de mapeamento de arranjos jurídico-institucionais parte do reconhecimento de que as instituições moldam comportamentos e regulam expectativas, sendo fontes de modelos morais e cognitivos. O neoinstitucionalismo destaca a assimetria existente no acesso ao poder entre os grupos sociais que determina os arranjos institucionais decorrentes. Ademais, reconhece-se que o momento da tomada de decisão normativa é um ponto de inflexão, uma vez que as alternativas que deixaram de ser tomadas perdem força frente à opção institucional tomada. Nesse sentido, as conjunturas críticas são caracterizadas por definir os contornos da evolução institucional subsequente de determinada política. A vantagem do método de mapeamento de arranjo jurídico-institucional é o esclarecimento das relações entre os atores e interesses envolvidos no processo decisório, através de uma análise jurídica do quadro normativo. O estudo da trajetória do quadro normativo possibilita identificar a vida de uma política, com seus discursos, argumentos, dimensões legislativas e regulamentares e as transformações, inflexões e evoluções desta política (Bucci; Gaspardo, 2024).

A Teoria do Estado, campo em que se insere esta dissertação, enquanto Ciência Jurídica, tem o condão de identificar “alternativas de desenhos institucionais emergentes” e “propostas de engenharia institucional”, partindo-se tanto de análise de estudos empíricos provenientes, em especial, da Ciência Política, quando de modelos teóricos da democracia, oriundos da Teoria Política (Bucci; Gaspardo, 2024). O método do mapeamento de arranjos-jurídico institucionais, que se situa na seara da Teoria do Estado e, portanto, de base jurídica, busca apreender determinado arranjo a partir de uma abordagem das relações entre aspectos jurídicos com outros aspectos, especialmente políticos. Tem-se, enquanto fundamentação, o neoinstitucionalismo histórico, corrente que busca observar a intermediação entre a estrutura e a ação social, no espaço entre teorias que priorizam as estruturas e sistemas sociais, como o marxismo e o funcionalismo, e aquelas que focam nos indivíduos e processos de decisão individual, como o behaviorismo. Uma das principais dificuldades metodológicas da pesquisa em direito, que é dotado de linguagem e articulação interna próprias, é o uso de fontes teóricas exógenas à área de acordo com uma racionalidade organizada. Impõe-se, entretanto, como se depreende do desenvolvimento desta dissertação, que a pesquisa jurídica dialogue com fontes diversas. O conceito de arranjos jurídico-institucionais abarca o conjunto de normas de diferentes naturezas, hierarquia e emanados por autoridades diversas, que regulamentam determinado

tema. O estudo dos arranjos permite compreender a racionalidade do conjunto normativo a partir de uma abordagem jurídica (Bucci; Gaspardo, 2024).

Os arranjos jurídico-institucionais são do plano mesoinstitucional, de forma que se busca observar a “ação governamental agregada em unidades maiores, volta à escala ampla das políticas públicas”, entre o plano macroinstitucional, que busca observar o jogo e as forças políticas visando o governo e o plano microinstitucional, que observa as relações intersubjetivas e a ação atomizada. O termo arranjo compreende a integração de conjuntos dinâmicos determinados com base em um sentido compartilhado por atores sociais, normas, processos e instituições jurídicas ou problemas jurídicos interconectados, cuja multidimensionalidade e multidisciplinaridade impõe que leve-se em consideração aspectos extrajurídicos para compreender a dinâmica, e dotado de flexibilidade formal, tendo em vista a complexidade jurídica específica que decorre da miríade de normas e polos de competência que regulamentam determinado tema. O termo institucional se refere ao aspecto sistemático de determinado quadro normativo, o nexo de unidade e sentido que une os elementos que compõem o arranjo, buscando estabilidade no tempo e a ideia diretriz, que orienta este conjunto à mudança, de acordo com determinada perspectiva. Ademais, denota o quadro organizativo da ação, com base no direito, e a despersonalização do arranjo, uma vez que este deixa de depender de seu instituidor (Bucci; Gaspardo, 2024).

Já o termo jurídico diz respeito à “conformação e operacionalização do quadro organizativo de ação pelo direito” e abrange as dimensões objetiva, com definição de atribuição e competências dos agentes envolvidos, discriminação de autoridades e tipologia das decisões previstas, tradução dos interesses privados de todos os sujeitos afetados pela política, como consumidores e empresas; subjetiva, com a definição de direitos, deveres, obrigações e faculdades derivados das normas, estruturas e procedimentos jurídicos; processual, através do regramento da participação social na efetivação da política; legitimação democrática; estabilização política; material, enquanto realização de objetivos constitucionais e legais; formal, que abrange a documentação em normas jurídicas, atos e medidas com base no sistema jurídico; preenchimento jurídico, que abrange normas infralegais tais como portarias, decretos e regulamentos, além de procedimentos e rotinas (Bucci; Gaspardo, 2024).

O conceito de arranjo institucional é correntemente utilizado no Campo de Públicas e na Ciência Política. Neste primeiro, denota o conjunto de regras, mecanismos e processos que definem a maneira como se coordenam atores e interesses na implementação de uma política pública. Por se tratar de um conceito proveniente de uma ciência extrajurídica, o elemento normativo é relegado a segundo plano, e a institucionalidade é reconhecida apenas como regras

gerais que estabelecem o funcionamento dos sistemas, chamado ambiente institucional, e regras específicas que regulam relações econômicas, políticas ou sociais particulares, chamados arranjo institucionais. Assim, a abordagem da Teoria do Estado tem o condão de conferir um adensamento na análise jurídica dos arranjos institucionais, valendo-se da interdisciplinaridade (Bucci; Gaspardo, 2024). No direito, a noção de arquitetura constitucional, desenho institucional e instituições dialoga com o conceito de arranjos jurídico-institucionais, uma vez que reconhece os elementos normativos enquanto força autônoma, que se relaciona, porém não é determinado, apenas por fatores extrajurídicos (Bucci; Gaspardo, 2024).

A Teoria do Estado é dotada de dois elementos centrais que a distingue e dá especificidade frente a outras ciências: o objeto, que é a institucionalização jurídica do poder por meio de normas jurídicas socialmente reconhecidas, tais como constituições, leis, decretos, decisões judiciais, etc., que vedam ou estimulam comportamentos, regem o exercício do poder e estabelece procedimentos para a mediação de conflitos políticos; e o método, que é a formulação de classificações e conceitos unificadores, ou tipos empíricos. Todavia, a globalização e a interdependência, mediante a complexidade e diversidade de arranjos jurídico-institucionais contemporâneos, sugere o alargamento da ideia de “conhecimento de síntese”, para colocar como enfoque o aspecto jurídico-institucional, em diálogo com fatores exógenos. Assim, o método de mapeamento institucional, inspirado em Roberto Mangabeira Unger, parte do pressuposto de que a institucionalidade vigente é um dos arranjos possíveis, perscrutando-se, a partir deste reconhecimento, alternativas de reformas institucionais frente aos desafios apresentados ao Estado contemporâneo, a partir de um diagnóstico sobre os problemas do Estado, sobre os quais tais alternativas são formuladas, e quais são os valores e interesses que os orientam. Desta feita, observam-se tanto questões de arquitetura institucional, ou seja, dos valores, quanto de engenharia institucional, ou as forças políticas, econômicas, o momento histórico etc. (Bucci; Gaspardo, 2024).

O mapeamento institucional, assim, parte do experimentalismo democrático, tendo como pressupostos teóricos: a necessidade de superar a prostração intelectual e política frente a alternativas institucionais àquelas vigentes, mediante um certo fetichismo institucional e estrutural; a possibilidade de realizar-se reformas institucionais, descartando-se a necessidade de alterações globais e revolucionárias; a necessidade de partir de circunstâncias históricas, entendidas enquanto condicionantes e não impedimentos absolutos, para gerar ideias programáticas realistas; os modelos teóricos-normativos não devem prescindir da análise empírica do exercício institucional; propostas de experimentação democrática não são neutras; a necessidade de combinar mobilização social e participação institucionalizada; as instituições

devem ser desenhadas tendo em vista os sujeitos reais e suas limitações. Ou seja, não se desdenha do conhecimento teórico dogmático, mas faz-se o exercício de observar as instituições de fato frente aos limites que se impõem, exógenos e endógenos. Assim, através da contraposição entre a forma institucional racional e experimental, evidencia-se a distinção entre os objetivos pretendidos e os efeitos efetivamente decorrentes de determinado contexto histórico (Bucci; Gaspardo, 2024).

O recurso ao método do mapeamento de arranjo jurídico-institucional permitiu, nesta dissertação, através da análise do plano mesoinstitucional, compreender nuances e relações entre os atores envolvidos no processo de um tema que, por natureza, é multiparticipativo (com atuação de governos, sociedade civil, *experts* em comunicação digital e tecnologia, academia e usuários). Nesse sentido, foi possível observar características de um movimento convergente entre plataformas e reguladores (em função da pressão dos segundos sobre os primeiros) no sentido de exercer moderação da comunicação digital e, especialmente, da disseminação de conteúdos pelas mídias digitais, que parte de um reconhecimento quanto à estrutura das plataformas por parte de governos e *experts*, e dos impactos que a comunicação digital produziu nas democracias recentemente. É possível observar, também, que a atuação institucional, tanto de governos nacionais quanto regionais, especialmente da União Europeia, e a velocidade com que o debate tem avançado, modificam constantemente a postura das plataformas ao exercer moderação de conteúdo.

O método de mapeamento institucional realiza uma abordagem crítica das instituições vigentes, porém busca se afastar de teorias normativas ideais, pautados em um ideal de justiça, formuladas pela Filosofia Política, de forma que parte, ao contrário, de uma abordagem sobre as injustiças derivadas das instituições reais, ou seja, parte da *clinical institutional theory*, proposta por Wiens, que propõe realizar uma identificação de falhas nas instituições vigentes, em um processo descritivo, e os problemas que a provocam, proporcionando um diagnóstico. Identifica-se, assim, que se deve suprir uma certa carência sobre problemas de engenharia institucional, havendo uma superexposição de questões de arquitetura institucional, de forma que as soluções que se propõem a partir do mapeamento institucional são voltadas a problemas existentes, de forma temporária, contingente e experimental, e não para superar as distâncias entre modelos ideais e a realidade (Bucci; Gaspardo, 2024).

A partir da aplicação do método, torna-se possível produzir estudos sobre arranjos específicos, sem que se perca de vista aspectos políticos sobre a institucionalidade, partindo-se da análise de normas jurídicas. Assim, sob uma perspectiva jurídica, torna-se viável analisar as estruturas, processos e valores de determinado arranjo institucional em sua relação com o

contexto histórico em que se insere. Tem-se como resultado uma esquematização explicativa da evolução temporal e das forças políticas que incidem sobre aquele arranjo, ou seja, uma análise sobre as instituições em movimento (Bucci; Gaspardo, 2024).

O método de mapeamento de arranjos jurídico-institucionais neste trabalho se deu por meio da aplicação sistemática das seguintes etapas: 1) periodização; 2) seleção e descrição dos arranjos jurídico-institucionais; 3) documentação jurídica; 4) contexto e atores políticos; 5) análise das motivações, efeitos esperados e obtidos; 6) análise combinada dos tópicos. O método comporta, ainda, outras duas etapas complementares, que consistem na análise das instituições em movimento e em derivações prescritivas do mapeamento, que não foram realizados no presente estudo.

3.2. Periodização

A primeira etapa, de periodização, se trata do reconhecimento de um recorte temporal justificado e os pontos de inflexão, ou conjunturas críticas, que mudaram a direção do arranjo estudado. Para tanto, pode-se apoiar na bibliografia da Ciência Política ou outras disciplinas pertinentes. No caso em estudo, observa-se a responsabilização das plataformas digitais sobre o conteúdo postado por terceiros, enquanto uma forma de combate à disseminação de *fake news* e discurso de ódio.

3.2.1. O Caso Snowden e o Marco Civil da Internet

O projeto de lei do Marco Civil da Internet, conforme estudaremos adiante, foi resultado de uma proposta multiparticipativa, que surgiu da parceria entre o Governo e a academia, consolidado e proposto ainda no ano de 2011. O projeto, entretanto, não foi votado até o ano de 2013, quando as discussões foram retomadas em função da descoberta do Caso Snowden.

No parecer proferido em plenário do PL nº 2.126/2011, que posteriormente foi transformado no Marco Civil da Internet, o relator Alessandro Molon (à época do PT/RJ), destaca que após longas discussões sobre o projeto, a votação chegou a ser pautada no Plenário da Câmara dos Deputados, mas deixou de ser votado ao longo do ano de 2012. O Caso Snowden foi determinante para mudar o cenário, acelerando o trâmite do projeto de lei, mediante requerimento de urgência realizado pela Presidência da República, autora do projeto, aprovado em 12 de setembro de 2013. Segundo o parecer, ao revelar-se que o governo estadunidense

monitorava indivíduos e governos estrangeiros, inclusive o governo brasileiro e a então Presidente Dilma Roussef, bem como outros governos, como da Alemanha e da França, preocupações quanto à soberania nacional e a privacidade dos indivíduos foram alçadas a um patamar prioritário. Em discurso de abertura da Assembleia Geral da Organização das Nações Unidas, a Presidente Dilma defendeu que a *internet* deve ser livre, aberta, democrática, descentralizada e que a governança deve ser exercida no formato *multistakeholder*, ou seja, multiparticipativo, com participação do Governo, empresas, sociedade civil e academia (BRASIL, 2014a, p. 21-22).

Ainda, em parecer da Comissão de Ciência, Tecnologia, Inovação, Comunicação e Informática do Senado Federal ao Projeto de Lei do Marco Civil da Internet, sob relatoria do Senador Zezé Perrella (à época PDT/MG), destaca-se que a *internet* é um ambiente que, diferentemente das mídias de massa tradicionais, possibilita que os próprios usuários publiquem conteúdos, de maneira a se constituir enquanto um ambiente de liberdade e, também, de abusos. Os abusos mencionados expressamente no relatório dizem respeito ao cerceamento ao livre pensamento e o monitoramento de comunicações privadas, empresariais e governamentais. Assim, é mencionado expressamente, outra vez, o Caso Snowden e a investigação, realizada pela Agência Nacional de Segurança estadunidense, sobre cidadãos brasileiros, inclusive a Presidente Dilma (Brasil, 2014e, p. 4).

O parecer do relator na Comissão Especial da Câmara dos Deputados destinada a discutir o PL nº 2.126/2011 foi lido em plenário na data de 12 de fevereiro de 2014, tendo sido oferecido texto Substitutivo àquele apresentado inicialmente pelo Executivo Federal, em que foram discutidas 34 emendas parlamentares. Neste documento, que será analisado posteriormente, constam as principais discussões sobre o conteúdo do que veio a ser aprovado enquanto a Lei nº 12.965/2014, o Marco Civil da Internet, com as motivações e efeitos esperados pelo legislador. Outras alterações, majoritariamente de redação, decorreram de emendas parlamentares em plenário, de números 35 a 69, que constam da Subemenda Substitutiva Global às Emendas de Plenário ao Projeto de Lei nº 2.126, de 2011, o qual também será analisado posteriormente, e que foi aprovado na Câmara dos Deputados em 25 de março de 2014.

Após ser aprovado na Câmara dos Deputados, o PL nº 2.126/2011 foi recebido no Senado Federal em 26 de março de 2014, como Projeto de Lei da Câmara nº 21, de 2014, encaminhado às Comissões de Ciência, Tecnologia, Inovação, Comunicação e Informática; de Meio Ambiente, Defesa do Consumidor e Fiscalização e Controle; e de Constituição, Justiça e Cidadania. O prazo para proposição de emendas no Senado Federal foi bastante exíguo, em

função do regime de urgência, sendo de apenas cinco dias úteis, nos termos do art. 122, II, “b” c/c 375, I, do Regimento Interno do Senado Federal, tendo sido apresentadas 41 emendas e, posteriormente, é apresentada mais uma emenda. Foram realizadas duas audiências públicas. Em votação no plenário do Senado Federal em 22 de abril de 2014, portanto menos de um mês após o encaminhamento do PL ao Senado Federal, foi aprovado o Marco Civil da Internet, tendo sido o texto enviado pela Câmara aprovado na integralidade, alterado pelas emendas nº 2 e 3, que promovem apenas pequenas alterações de redação. A Lei nº 12.965 é, então, sancionada em 24 de abril de 2014.

3.2.2. O escândalo da Cambridge Analytica e o Programa Permanente de Enfrentamento à Desinformação

Já nas eleições gerais brasileiras de 2018, a preocupação com a ampla disseminação de *fake news* durante o período eleitoral justificou a instituição, por parte do Tribunal Superior Eleitoral, do Conselho Consultivo sobre Internet e Eleições, com participação da Justiça Eleitoral, Ministério Público, Ministério da Defesa, Ministério da Justiça, Polícia Federal, Ministério da Ciência e Tecnologia, entre outros, bem como da sociedade civil. As atribuições do órgão eram, de acordo com o art. 2º da Portaria nº 949/2017 do TSE (Brasil, 2017):

- “I - Desenvolver pesquisas e estudos sobre as regras eleitorais e a influência da Internet nas eleições, em especial o risco das *fake news* e o uso de robôs na disseminação das informações;
- II - Opinar sobre as matérias que lhe sejam submetidas pela Presidência do TSE;
- III - propor ações e metas voltadas ao aperfeiçoamento das normas.” (Brasil, 2017).

A partir dos estudos elaborados pelo próprio Tribunal Superior Eleitoral durante as eleições de 2018, averiguou-se a necessidade de estabelecer medidas para o combate à desinformação nas eleições subsequentes, tanto nas eleições municipais de 2020 quanto nas eleições gerais de 2022, sendo, então, criado o Programa de Enfrentamento à Desinformação com Foco nas Eleições 2020, por meio da Portaria nº 663 de 30 de agosto de 2019.

O escopo do Programa é muito mais amplo que o do Conselho Consultivo sobre Internet e Eleições, o qual foi responsável por produzir relatório que fundamentou o novo programa, resultando em propostas organizadas por grupo de trabalho instituído pela Portaria do TSE nº 382/2019. Através do Programa, houve a ampliação da interação entre o Tribunal Superior Eleitoral, as plataformas digitais e a sociedade civil, além da tomada de ações concretas para combater a disseminação de desinformação durante o período eleitoral, o que está documentado

em Plano Estratégico e Relatório de Ações e Resultados do Programa. Destacam-se, enquanto eventos citados no Plano Estratégico (Brasil, 2020a), tanto os resultados da atuação do Conselho Consultivo sobre Internet e Eleições em relação às eleições de 2018, onde constatou-se a veiculação metódica e deliberada em massa de conteúdo com desinformação, quanto os impactos causados na saúde pública pela disseminação deste tipo de conteúdo durante a pandemia de Covid-19, demonstrando os efeitos perniciosos que tais materiais podem causar, o que reforçou a percepção de necessidade de ações afirmativas no combate à disseminação de *fake news* (Brasil, 2021, p. 31-32).

O Relatório de Ações e Resultados do Programa de 2020 destaca o cenário de disseminação massiva de desinformação durante o pleito eleitoral de 2018, bem como o caso das eleições estadunidenses em 2016 e o *Brexit*, em 2020, em que se tornou clara a atuação de agentes com interesses em obter ganhos, sejam políticos ou econômicos, com a disseminação sistemática e organizada de falsidades, colocando em dúvida até mesmo o próprio processo eleitoral (Brasil, 2021, p. 10). Da mesma maneira, o Programa Permanente de Enfrentamento à Desinformação, em plano estratégico para as eleições gerais de 2022, destaca que as campanhas de desinformação e ódio por meio da comunicação digital vêm afetando o próprio pleito eleitoral, buscando atacar a credibilidade do sistema, gerando desconfiças e buscando, através de alegações infundadas de fraude, deslegitimar as eleições. O plano destaca, assim, que esse processo ocorreu não apenas nas eleições estadunidenses de 2016 e 2020, mas também nas eleições brasileiras em 2018 e 2020, colombianas de 2018, nas eleições presidenciais peruanas e legislativas no México em 2021, entre outros casos mundo afora (Brasil, 2022a, p. 10).

Se à época do debate legislativo e da aprovação do Marco Civil da Internet o Caso Snowden era frequentemente apontado enquanto um ponto de inflexão na compreensão sobre a capacidade de governos acessarem e processarem dados de usuários e outros governos, o escândalo da *Cambridge Analytica*, que envolveu a eleição presidencial estadunidense de 2016 e a campanha do *Brexit*, bem como a atuação de empresas como a *War Room* e a *CA-Ponte*, parceira da *Cambridge Analytica* no Brasil, que influenciou as eleições presidenciais brasileiras de 2018, são eventos que constituem um segundo ponto de inflexão, ao demonstrarem a capacidade de empresas que atuam com o processamento de dados dos usuários para influenciar decisões políticas e eleitorais.

3.3. Seleção e descrição de Arranjos Jurídico-Institucionais

A segunda etapa é a seleção e descrição dos arranjos jurídico-institucionais, destacando o objeto a ser estudado de seu entorno. Em algumas pesquisas este recorte é mais facilitado, como na abordagem de direito e políticas públicas, em que o próprio objeto já é bem delimitado, utilizando-se do conceito de ação governamental. Em pesquisas de maior abrangência, como sobre a crise da democracia, pode-se adotar uma abordagem dos eixos jurídico-institucionais estruturantes de determinada estrutura macroinstitucional, o que possibilita a análise independente de cada um, pautado no conceito de ação política que engloba a ação legislativa, partidária, de organizações internacionais e agentes privados etc. e a gestão governamental (Bucci; Gaspardo, 2024).

No caso desta pesquisa, o recorte é o estabelecimento de medidas jurídico-institucionais voltadas ao combate à disseminação de *fake news* e discurso de ódio. Uma vez que se trata de um tema com atuação multissetorial, do Estado, plataformas, academia, sociedade civil e usuários, sua delimitação pode ser difícil. Entretanto, partindo do documento que inaugura o arranjo jurídico-institucional observado, o Marco Civil da Internet, observa-se que a previsão de isenção de responsabilidade das plataformas digitais sobre o conteúdo postado por terceiros, presente no art. 19 e seguintes, estabelece um parâmetro para as ações e políticas desenvolvidas posteriormente, em um contexto de profunda transformação da maneira como a matéria é entendida.

Nesse sentido, se até o momento o Congresso Nacional não promoveu reforma legislativa nesse aspecto, tampouco o Supremo Tribunal Federal decidiu sobre a constitucionalidade ou não deste dispositivo legal, os temas vêm sendo discutidos tanto em um quanto em outro. Nesse ínterim, o combate à disseminação de *fake news* e discurso de ódio nas plataformas digitais foi modificado por outras disposições oriundas tanto das plataformas, ao exercer governança privada dos discursos, quanto do Estado, especialmente por meio da atuação do Tribunal Superior Eleitoral durante as eleições de 2020 e 2022. Sem que tenha ocorrido mudança no quadro jurídico instituído pelo Marco Civil da Internet, ou seja, ainda vigente a isenção de responsabilidade das plataformas digitais sobre o conteúdo postado por terceiros, a atuação do TSE, conforme verificaremos, representa uma importante alteração no tratamento dado à matéria no Brasil.

3.4. Documentação jurídica

A terceira etapa é a descrição das características do arranjo a ser estudado, destacando-se sua base normativa, inclusive o processo legislativo, o problema que se quer enfrentar, o desenho jurídico-institucional e a composição política e social que o apoia. É importante destacar que não se trata de mera descrição legal cronológica, mas sim da compreensão do quadro normativo que dá materialidade institucional ao arranjo (Bucci; Gaspardo, 2024).

Tendo como marco inicial o Marco Civil da Internet, enquanto documento legislativo que inaugura uma regulamentação específica para as tecnologias de comunicação e informação digitais no Brasil, com vigência relativamente recente, tendo sido sancionado no ano de 2014, o qual isenta de responsabilidade civil as plataformas digitais sobre o conteúdo postado por terceiros, foram identificados pontos de inflexão na atuação das plataformas que, por meio da ação institucional, geraram incentivos/desincentivos, tendo como referencial o arcabouço normativo centrado no Marco Civil da Internet, para o combate à desinformação ou ao discurso de ódio, seja por meio da remoção de conteúdos, seja por meios alternativos como a checagem de fatos ou divulgação prioritária de materiais de órgãos oficiais. Nesse sentido, serão analisados: o trâmite do Marco Civil da Internet; as Resoluções do Tribunal Superior Eleitoral, concernentes à eleição geral de 2022 e às eleições municipais de 2020, bem como os memorandos de entendimento firmados com as plataformas digitais *Facebook*.

Conforme exposto nos primeiros capítulos, a responsabilização das plataformas digitais sobre o conteúdo postado por terceiros, ou sua isenção, gera incentivos para que atores privados, no caso as plataformas digitais, exerçam moderação de conteúdo quanto a determinados materiais, que, por exemplo, incentivem a violência, contenham discurso de ódio ou *fake news*. Até a adoção do Marco Civil da Internet, a jurisprudência brasileira se subdividia em três correntes quanto à possibilidade de responsabilização: uma que determinava a responsabilidade objetiva, outra subjetiva e uma terceira que isentava as plataformas de responsabilidade. Já sob a égide do Marco Civil da Internet, e mediante o impacto de eventos descritos na fase de periodização, como a influência da análise psicológica-comportamental e o *microtargeting* em eleições, por exemplo, foram tomadas outras medidas específicas, notadamente aquelas circunscritas ao período eleitoral, no caso das ações do Tribunal Superior Eleitoral, que modificaram a abordagem das plataformas, em medida a ser observada neste tópico, ainda que sem ter revogado a isenção de responsabilidade disposta no art. 19 do Marco Civil da Internet.

Ademais, tendo em vista que as plataformas exercem governança privada do discurso, conforme discutido no segundo capítulo, faz-se mister observar os Termos e Condições de Uso das plataformas *Facebook* e *Instagram*, objeto da presente pesquisa, enquanto uma relação contratual ou quase contratual entre os usuários e a plataforma. Além de cumprir com o objetivo

de compreender de que maneira as plataformas exercem a autorregulação, observar por um lado as normas públicas que regulam a comunicação digital e por outro os contratos privados, elaborados de maneira unilateral pelas próprias plataformas, permitirá investigar a hipótese de que as medidas tomadas pelo poder público -e mesmo aquelas que não se sujeitaram ao trâmite legislativo- geram determinados incentivos aos agentes privados envolvidos na comunicação digital, avaliando-se a efetividade, legitimidade e legalidade do sistema de governança resultante, bem como o respeito às balizas do princípio constitucional da liberdade de expressão.

Quanto ao Marco Civil da Internet, foi analisado o texto legal, especialmente no que diz respeito à Seção 3- Da Responsabilidade por Danos Decorrentes de Conteúdo Gerado por Terceiros do Capítulo III- Da Provisão de Conexão e Aplicações de Internet, bem como as normas gerais constantes do Capítulo I Disposições Preliminares e do Capítulo II Dos Direitos e Garantias dos Usuários. Foi analisado o trâmite legislativo por meio dos sites da Câmara dos Deputados (link: <https://www.camara.leg.br/proposicoesWeb/fichadetramitacao?idProposicao=517255>) e do Senado Federal (<https://www25.senado.leg.br/web/atividade/materias/-/materia/116682>) e, da análise qualitativa dos andamentos, foi realizada análise sobre os seguintes documentos: 1) Avulso Inicial do Projeto de Lei nº 2.126/2011 de Autoria da Presidência da República; 2) Parecer em Plenário da Câmara dos Deputados, sob relatoria do Deputado Alessandro Molon (à época PT/RJ); 3) Subemenda substitutiva global às emendas de Plenário da Câmara dos Deputados, sob autoria do Deputado Alessandro Molon (à época PT/RJ); 4) encaminhamento ao Senado Federal do PLC nº 21/2014, com informação do regime de urgência; 5) Parecer da Comissão de Constituição, Justiça e Cidadania do Senado Federal, sob relatoria do Senador Vital do Rêgo (MDB/PB); e 6) Parecer da Comissão de Ciência, Tecnologia, Inovação, Comunicação e Informática do Senado Federal, de autoria do Senador Zezé Perrella (à época PDT/MG)⁸.

Quanto às eleições municipais de 2020 foram analisados 1) Resolução-TSE nº 23.610, de 18 de dezembro de 2019, sobre propaganda eleitoral, utilização e geração do horário gratuito e condutas ilícitas em campanha eleitoral; 2) o Plano Estratégico do Programa de Enfrentamento à Desinformação com Foco nas Eleições 2020; 3) o Relatório de Ações e

⁸ O PLC nº 21/2014 foi encaminhado, ainda, à Comissão de Meio Ambiente, Defesa do Consumidor e Fiscalização e Controle, sob relatoria do Senador Ricardo Ferraço (à época do PMDB/ES). O relator proferir Parecer em Plenário do Senado Federal, em favor da aprovação do PLC nº 21/2014, nos termos do Relatório da Comissão de Ciência, Tecnologia, Inovação, Comunicação e Informática do Senado Federal.

Resultados do Programa de Enfrentamento à Desinformação com Foco nas Eleições 2020; 4)) o Memorando de Entendimento-TSE nº 40/2020 entre o Tribunal Superior Eleitoral e o *Google*, sobre o mecanismo de busca homônimo e a plataforma de vídeos *YouTube*; 5 o Memorando de Entendimento-TSE nº 41/2020 entre o Tribunal Superior Eleitoral e o *Facebook*, sobre a rede social homônima e a rede social *Instagram*.

Quanto às eleições municipais de 2022 foram analisados: 1) Resolução-TSE nº 23.671 , de 14 de dezembro de 2021, sobre propaganda eleitoral, utilização e geração do horário gratuito e condutas ilícitas em campanha eleitoral, que atualizou a Resolução-TSE nº 23.610, de 18 de dezembro de 2019; 2) Resolução-TSE nº 23.650, de 9 de setembro de 2021, que institui a Política Geral de Privacidade e Proteção de Dados Pessoais no Âmbito da Justiça Eleitoral; 3) Lei de Crimes contra o Estado Democrático de Direito, instituída pela Lei nº 14.197 , de 1º de setembro de 2021; 4) o Plano Estratégico do Programa Permanente de Enfrentamento à Desinformação no âmbito da Justiça Eleitoral para as Eleições Gerais de 2022; 5) o Relatório de Gestão 2022 do Tribunal Superior Eleitoral; 6) o Memorando de Entendimento-TSE nº 1/2022 entre o Tribunal Superior Eleitoral e o *Google*, sobre o mecanismo de busca homônimo, plataforma de vídeos *YouTube* e demais serviços da empresa; 7) o Memorando de Entendimento-TSE nº 03/2022 entre o Tribunal Superior Eleitoral e *Facebook*, sobre a rede social homônima e a rede social *Instagram*. Ressalte-se que não foi possível analisar o relatório de ações e resultados do Programa Permanente para as Eleições Gerais de 2022, em função deste não ter sido publicado até a presente data.

Quanto às políticas e relatórios produzidos pelo *Facebook* e *Instagram* foram analisados: 1) os Termos e Políticas de Uso das plataformas; 2) os dados do relatório de transparência das plataformas; 3) os dados do Comitê de Supervisão do *Facebook*; e a Política Corporativa de Direitos Humanos da *Meta*, detentora das duas plataformas.

3.5. Contexto e atores políticos

A quarta etapa é a identificação do contexto e das forças políticas e sociais que atuam sobre aquele quadro institucional, entre eles atores políticos, agentes privados como empresas de tecnologia, organizações da sociedade civil etc. (Bucci; Gaspardo, 2024).

3.5.1. O Marco Civil da Internet

O Marco Civil da Internet se originou do Projeto de Lei nº 2.126/2011, de autoria do Executivo Federal, o qual foi proposto em 24 de agosto de 2011 por meio da Mensagem n. 326/2011. Inicialmente o PL foi encaminhado, em 29 de agosto de 2011, pela Mesa Diretora da Câmara dos Deputados, às Comissões de Defesa do Consumidor; Ciência e Tecnologia; Comunicação e Informática; e Constituição e Justiça e de Cidadania. No dia 26 de setembro de 2011, entretanto, por força do art. 34, inciso II do Regimento Interno da Câmara dos Deputados, determinou-se a criação de Comissão Especial para discutir o PL, a qual foi efetivamente criada em 26 de outubro de 2011, sendo designado relator o Deputado Alessandro Molon (à época PT/RJ).

Conforme discutido anteriormente, a regulamentação da comunicação digital envolve a participação de múltiplos atores e interesses, entre eles o Poder Público, as plataformas digitais, a sociedade civil organizada, especialistas e acadêmicos e, ainda, dos usuários. Desde o início, o PL nº 2.126/2011, que se tornou o Marco Civil da Internet, contou com a participação de múltiplos setores, tendo sido iniciativa do Poder Executivo Federal que, por meio de uma parceria entre a Secretaria de Assuntos Legislativos do Ministério da Justiça- SAJ/MJ e o Centro de Tecnologia e Sociedade da Escola de Direito da Fundação Getúlio Vargas do Rio de Janeiro, elaborou a proposta legislativa. Nesta iniciativa houve participação de instituições públicas e da sociedade civil, bem como dos próprios usuários, por meio de uma ferramenta digital na plataforma Cultura Digital, gerida pelo Ministério da Cultura e pela Rede Nacional de Ensino e Pesquisa, bem como manifestações feitas em plataformas como o *Twitter* (Brasil, 2011, p. 9-10).

A participação popular foi constituída, nesta fase “pré-legislativa”, por duas etapas: a primeira em torno dos princípios a serem incluídos no Marco Civil da Internet, quais sejam a liberdade de expressão, privacidade, neutralidade da rede, direito de acesso, responsabilidade das intermediárias e promoção da inovação, sendo, ao final, redigido um texto de lei preliminar, de autoria dos líderes do projeto; a segunda fase tinha como foco os artigos de um projeto de lei preliminar, sendo, ao final, elaborado o texto do PL nº 2.126/2011 (Lemos; Steibel; Affonso de Souza, 2015, p. 204-205).

Durante o trâmite do Projeto de Lei nº 2.126/2011, que se tornou a Lei nº 12.965/2014, o Marco Civil da Internet, conforme se afere do parecer proferido em plenário pelo Relator Deputado Alessandro Molon (à época PT/RJ), foram promovidas sete audiências públicas e seminários nas cidades de Brasília, Porto Alegre, Curitiba, Salvador, São Paulo e Rio de Janeiro, com a participação de 62 palestrantes, representantes dos setores Público, acadêmico, sociedade civil organizada, plataformas digitais, jornalismo e telecomunicação. Os temas

debatidos eram amplos, tendo em vista o escopo do Marco Civil da Internet, envolvendo: os direitos dos usuários, com dois seminários; responsabilidade civil de terceiros, com cinco seminários; atuação do Poder Público; guarda de logs e privacidade dos usuários; liberdade de expressão e o potencial de inovação da internet; neutralidade da rede e o potencial de inovação, com dois seminários; e neutralidade da rede e governança da internet. Ademais, foram disponibilizados canais de participação popular por meio da *internet*, com sugestões incorporadas ao relatório do PL nº 2.126/2011, no site e-Democracia da Câmara dos Deputados, além de manifestações por meio de *hashtags* em plataformas digitais como o *Twitter*. O site e-Democracia recebeu 45 mil visitas na página sobre o Marco Civil da Internet, além de 200 postagens e 140 propostas de alteração do PL. Houve, ainda, o envio de manifestações por escrito de outras cinquenta entidades (Brasil, 2014a, p. 9-21).

As contribuições recebidas durante o processo participativo eram feitas mediante identificação dos participantes, tendo em vista conferir maior transparência e identificar os posicionamentos dos setores envolvidos na normativa, possibilitando a crítica e escrutínio público, facilitando as negociações e reduzindo a assimetria de informações. Neste processo, determinados setores tiveram de renunciar a alguns pontos, em prol de um consenso em torno de um documento legislativo politicamente viável. Nesse sentido, o Ministério da Justiça teve de mudar de posição quanto aos registros de acesso, os fornecedores de acesso à internet e companhias de telecomunicação sobre a neutralidade da rede e a sociedade civil sobre a previsão de remoção de conteúdo de pornografia de vingança (Lemos; Steibel; Affonso de Souza, 2015, p. 203-205).

Lemos, Steibel e Affonso de Souza (2015, p. 206-208) analisaram, a partir das discussões do Marco Civil da Internet, os posicionamentos dos diversos setores envolvidos no processo, identificando-os de acordo com as seguintes categorias: Telecomunicações; Sociedade Civil; Companhias de Internet Globais; Companhias de Internet Brasileiras; Televisão e Rádio; Executivo Federal; Juristas (*law enforcement layers*) e Polícia Federal. Os temas de maior conflito durante o processo deliberativo foram divididos pelos autores nas seguintes categorias: neutralidade da rede; alta proteção à privacidade; “*safe harbor*” para o discurso, ou seja, isenção de responsabilidade das plataformas sobre o conteúdo postado por terceiros; retenção de dados; localização forçada de dados; “*safe harbor*” para direitos autorais e direitos conexos; e remoção de conteúdo por pornografia de vingança, sem necessidade de decisão judicial.

No que diz respeito a esta pesquisa, cumpre-nos compreender, de acordo com as categorias propostas pelos autores, qual o posicionamento dos atores políticos envolvidos na

aprovação do Marco Civil da Internet sobre os temas da responsabilidade das plataformas sobre o discurso, os direitos autorais e direitos conexos e pornografia de vingança. É importante destacar, inicialmente, que não houve consenso entre todos os atores políticos envolvidos no processo em nenhum dos três temas.

De acordo com a pesquisa, apoiaram a proposta de isenção de responsabilidade das plataformas digitais sobre o discurso, ou seja, os conteúdos postados por terceiros: a sociedade civil, as companhias de internet brasileiras e globais e o setor de transmissões (rádio e televisão). Tanto o setor de telecomunicações quanto o Executivo mantiveram-se neutros, enquanto advogados e a Polícia Federal se posicionaram contra a proposta que foi, ao final, aprovada.

Quanto à isenção de responsabilidade das plataformas digitais sobre conteúdos protegidos por direitos autorais e direitos conexos, foram a favor a sociedade civil e as companhias de *internet* globais, mantendo-se neutras, outra vez, as telecomunicações e o Executivo Federal e posicionando-se contra as companhias de internet brasileiras, o setor de transmissões e advogados e Polícia Federal, tendo sido a proposta ao final rejeitada.

Já quanto à remoção de conteúdo por pornografia de vingança, as telecomunicações, companhias de internet brasileiras e globais e o setor de transmissão se mantiveram neutros, enquanto a sociedade civil era contra e o Executivo, advogados e Polícia Federal se posicionaram a favor, sendo a proposta ao final aprovada.

A despeito da participação popular para a confecção do Marco Civil da Internet, é importante ressaltar que o apoio e os esforços da Presidência da República, tanto na iniciativa do projeto de lei durante o Governo Lula, em 2007, instrumentalizado especialmente pelo Ministério da Justiça e pelo então Ministro da Cultura Gilberto Gil, com o investimento e incentivo ao Fórum de Cultura Digital e, posteriormente, durante o trâmite do Marco Civil da Internet, mediante a divulgação do Caso Snowden em julho de 2013, com o consequente pedido de urgência da Presidente Dilma, foram indispensáveis para a aprovação do marco legal (Lemos; Steibel; Affonso de Souza, 2015, p. 203-205).

É importante ressaltar, ainda, a participação de dezenas de parlamentares na proposição e discussão do Marco Civil da Internet, de forma que ao PL nº 2.126/2011 foram anexados outros 43 projetos de lei, propostos entre 2000 e 2013, oriundos tanto da Câmara dos Deputados quanto do Senado Federal quanto do Executivo Federal, que versavam sobre a matéria (BRASIL, 2014a, p. 3-9). A própria origem do texto inicial do PL nº 2.126 é participativa, tendo em vista que partiu de um projeto conjunto entre a Secretaria de Assuntos Legislativos do Ministério da Justiça e da Escola de Direito do Rio de Janeiro da Fundação Getúlio Vargas, que

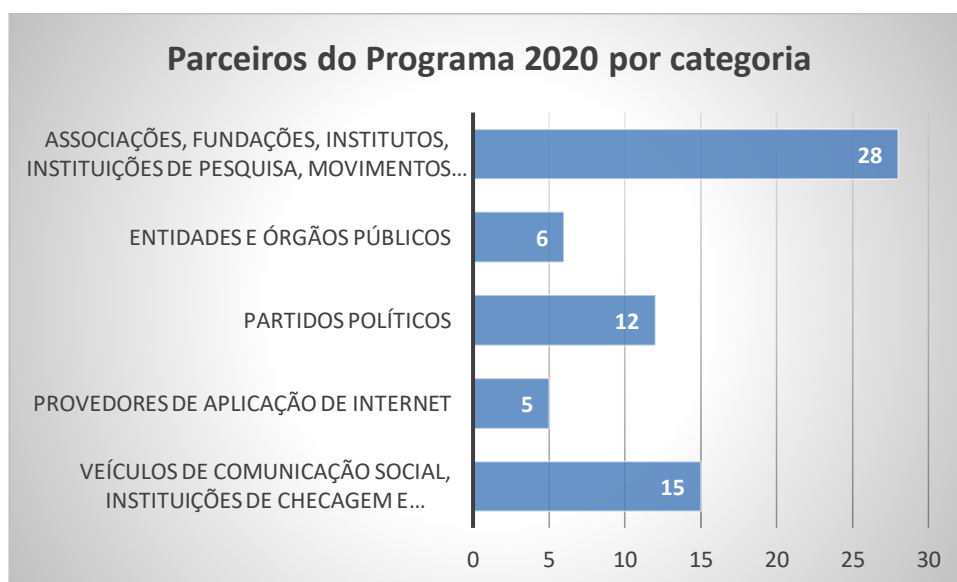
disponibilizaram um anteprojeto em um *site*, para receber contribuições populares, em outubro de 2009. Ao final, e após mais de 160 mil visitas ao site e mais de 2300 contribuições, apresentou-se o avulso inicial do PL nº 2.126, por meio da Mensagem nº 326/2011.

3.5.2. O Programa de Enfrentamento à Disseminação de Desinformação com Foco nas eleições de 2020

O projeto que deu origem ao Marco Civil da Internet contou com a iniciativa do Ministério da Justiça, portanto um órgão público do Executivo Federal, em parceria com uma instituição acadêmica e, portanto, da sociedade civil, com *expertise* na área, qual seja a Fundação Getúlio Vargas. Através do processo legislativo do Marco Civil da Internet se estabeleceu um amplo debate, envolvendo diversos setores, sobre regulamentação da comunicação digital. Esse ambiente integrado pode ser observado pela ampla participação de instituições, públicas ou privadas, mobilizadas pelo Tribunal Superior Eleitoral, durante os maiores programas de enfrentamento à disseminação de desinformação levados a efeito no Brasil, durante as eleições de 2020 e 2022. É importante observar que pela especificidade e complexidade do tema, bem como mediante sua relevância no debate público, há uma ampla gama de atores ativos, bem-organizados e capacitados para discutir o tema no país.

Nesse sentido, durante o Programa de Enfrentamento à Disseminação de Desinformação com Foco nas eleições de 2020, é possível analisar, a partir da lista de parceiros do programa, que o grupo de parceiros multissetorial, que consta no Plano Estratégico do Programa, contava, inicialmente, com 55 instituições, posteriormente com a adesão de mais 11 (Brasil, 2021, p. 8), somando o total de 66 instituições públicas ou privadas, que se enquadraram em cinco categorias: veículos de comunicação social, instituições de checagem e associações representativas de imprensa; provedores de aplicação de *internet*, incluindo redes sociais, serviços de mensageria e mecanismos de busca; partidos políticos; entidades e órgãos públicos; e “associações, fundações, institutos, instituições de pesquisa, movimentos ou grupos, com *expertise* relacionada às temáticas de desinformação, liberdade de expressão, tecnologia, democracia, eleições, direito eleitoral e direitos humanos, e notório reconhecimento público na sua área de atuação” (Brasil, 2020a, p. 14), os quais foram escolhidos por meio de critérios técnicos e científicos, inclusive por meio de busca ativa do TSE sobre parceiros importantes para promover as estratégias do Programa, e que se comprometeram por meio de termos de adesão ao Programa, acordos de cooperação e memorandos de entendimento (Brasil, 2020a, p. 14).

Posteriormente será analisada a natureza da participação de cada um dos atores envolvidos no Programa, especialmente no que diz respeito às ações de resposta à disseminação de desinformação tomadas pelo Programa em parceria com os provedores de aplicação de *internet*. Porém, é relevante destacar que, de fato, o Programa de 2020 contou com a múltipla participação de atores envolvidos no processo de comunicação digital, contando com 28 instituições da sociedade civil com *expertise* em temáticas de desinformação, liberdade de expressão, tecnologia, democracia, eleições, direito eleitoral e direitos humanos, inclusive com a participação da Fundação Getúlio Vargas, atuante desde a elaboração do Marco Civil da Internet, bem como de seu principal parceiro naquele projeto, o Ministério de Justiça e Segurança Pública, entre os 6 órgãos públicos envolvidos no Programa de Enfrentamento à Desinformação. Além disso, é de se destacar a presença de nove associações representativas de imprensa, entre elas do rádio, televisão, mídia impressa, telecomunicações e serviços de telefonia celular, três veículos de imprensa e três instituições de checagem (Boatos.org, Agência Lupa e Agência aos Fatos), além de cinco dos provedores de aplicação de *internet* com mais usuários no Brasil à época, quais sejam *Twitter*, *WhatsApp*, *Facebook (Instagram)*, *Google (YouTube)* e *ByteDance (TikTok)*.



3.5.3. O Programa Permanente de Enfrentamento à Desinformação no âmbito da Justiça Eleitoral

Para as ações voltadas às eleições gerais de 2022, o Programa Permanente manteve e ampliou a rede de conexões e parcerias com instituições que já haviam aderido ao Programa de

2020. No Planejamento, o Programa contava com 72 instituições parceiras, incluindo desde empresas de tecnologia, veículos de imprensa, instituições da sociedade civil etc., em categorias muito semelhantes às daquelas do Programa de 2020, descritas acima. É importante destacar que todos os parceiros do Programa de 2020 foram mantidos para o Programa Permanente, havendo a inclusão de mais quatro associações, fundações, institutos, instituições de pesquisa, movimentos ou grupos com *expertise* na área, uma empresa de tecnologia de software, a Microsoft, e um provedor de aplicação de internet, o Joyo Tecnologia Brasil LTDA (*Kwai*). Houve, ainda, a inclusão de uma nova categoria, qual seja “empresas de tecnologia que possam contribuir com ações na área de ciberinteligência”, que apesar de já terem participado desde 2020, foram destacadas no Programa de 2022 (Brasil, 2022a, p. 31-33).

Apesar de no relatório de ações e estratégias para as eleições de 2020 o Programa Permanente de Enfrentamento à Disseminação já contar com 72 parceiros, esta lista foi ampliada durante o processo para 162 instituições⁹, o que representou um aumento de 125% em apenas um ano, inclusive com a parceria com o provedor de aplicativo de *internet Telegram*, o qual inicialmente havia relutado em estabelecer parceria, tendo sido inédita aquela que efetivou com o TSE (Brasil, 2023, p. 54).

A ampla participação de diversos atores no contexto do combate à desinformação é frequentemente apontada enquanto fator fundamental para sua eficácia e legitimidade. Cada setor desse sistema multiparticipativo desempenha um papel distinto, de acordo com suas aptidões e características. O Plano Estratégico para as Eleições de 2022 (Brasil, 2022a), descreve as principais atividades desenvolvidas por cada um dos setores no combate à desinformação para as Eleições de 2022, conferindo materialidade e exemplificando a estrutura de um sistema multiparticipativo de combate à disseminação de desinformação.

Nesse sentido, destaca que a coordenação com plataformas digitais tem diversas frentes, entre elas a capacitação dos servidores da Justiça Eleitoral sobre o funcionamento das plataformas, tornando mais claros o funcionamento das plataformas; a utilização de ferramentas específicas de cada serviço para aumentar o alcance de informações verídicas sobre o processo eleitoral; e, especialmente, o combate à desinformação, tendo em vista a necessidade de celeridade, com uma “autorregulação estruturada”, nos moldes do procedimento adotado pelo TSE para lidar com ocorrências de desinformação, buscando garantir uniformidade e

⁹ A lista atualizada de parceiros do Programa Permanente de Enfrentamento à Desinformação ainda não havia sido publicada até a data de conclusão desta pesquisa, razão pela qual deixou de se elaborar análise quanto aos novos parceiros do programa.

transparência ao processo, inclusive com indicação de produção de relatórios de *feedback* por parte das plataformas quanto às ações tomadas (Brasil, 2022a, p. 33).

A coordenação com a sociedade civil se destaca pela cooperação nas ações de monitoramento de conteúdos desinformativos, bem como pela constante debate com especialistas de setores acadêmicos, científicos e jornalísticos; o setor de imprensa e instituições de checagem atuam para garantir a liberdade de expressão e que a avaliação quanto aos conteúdos seja elaborada de acordo com independência editorial e autonomia, guiada por critérios de transparência, apartidarismo, ética jornalística, consistência na checagem nos últimos dois anos e política de correção clara e comunicada ao público. Já a coordenação com órgãos públicos visa medidas de investigação e repressão à disseminação sistemática de desinformação em um momento posterior, portanto após análise prévia quanto à veracidade e ao potencial de dano do conteúdo, de maneira que a atuação conjunta com o Ministério Público Eleitoral e a Polícia Federal visa responder com agilidade e eficiência a hipóteses de crime, tais como ameaça, violência, danos às urnas, entre outros (Brasil, 2022a, p. 34-35).

Por fim, as interlocuções com partidos políticos e outros órgãos institucionais e Tribunais teve como enfoque a formação, capacitação e incentivo por parte do Tribunal Superior Eleitoral, mediante a experiência e esforço nos períodos eleitorais, para a tomada de ações que busquem desincentivar a desinformação (Brasil, 2022a, p. 35).

3.6. Análise das motivações, efeitos esperados e efeitos obtidos

A quinta etapa, de análise das motivações, efeitos esperados e obtidos, parte da análise das alterações jurídicas e institucionais, e tem como principal fonte os processos legislativos, além de bibliografia de áreas como a Ciência Política ou relatórios oficiais e notícias, por exemplo, sendo possível utilizar como técnica auxiliar a avaliação legislativa (Bucci; Gaspardo, 2024).

3.6.1. O Marco Civil da Internet

O escopo do Marco Civil da Internet é mais amplo que o recorte desta pesquisa, de forma que dar-se-á enfoque especificamente às previsões que dizem respeito à responsabilização das plataformas digitais sobre o conteúdo postado por terceiros e eventuais mecanismos e incentivos para que as plataformas exerçam governança, especificamente dos

artigos 19 e seguintes do Marco Civil da Internet. Todavia, os princípios, fundamentos, direitos e garantias dos usuários também são importantes para compreender o quadro normativo deste recorte de pesquisa, de forma que não se olvidará de uma abordagem sobre estes temas.

É importante destacar que o projeto não se apresenta enquanto solução definitiva para todas as questões relativas às tecnologias de comunicação e informação digitais, mas sim como um norteador para a atuação tanto do Poder Público, quanto das intermediárias, setor privado, sociedade civil e usuários. Assim, trata-se de uma “proposta legislativa transversal e convergente [que] possibilitará um posicionamento futuro mais adequado sobre outros importantes temas relacionados à internet que ainda carecem de harmonização” citando-se temas como “proteção de dados pessoais, o comércio eletrônico, os crimes cibernéticos, o direito autoral, a governança da internet e a regulação da atividade de centros públicos de acesso à internet” (Brasil, 2011, p. 11). Em artigo publicado por Lemos, Stiebel e Affonso de Souza (2015, p. 205), que atuaram no processo participativo de elaboração do Marco Civil, os autores destacam a norma enquanto uma série de princípios legais e direitos com o objetivo de fundamentar legislações posteriores sobre a *internet* no Brasil.

Na justificativa do texto proposto pelo Executivo Federal do PL 2.126/2011, que veio a se tornar o Marco Civil da Internet, destacou-se que o Poder Judiciário, ao se deparar com questões de responsabilização das plataformas digitais, proferiu decisões conflitantes e que poderiam, no limite, implicar em riscos à liberdade de expressão e privacidade de toda a sociedade (Brasil, 2011, p. 8).

Uma das principais motivações que fundamentam a regulamentação da *internet* no Brasil, recorrentemente mencionada durante o trâmite do Marco Civil da Internet, diz respeito à proteção ao princípio da liberdade de expressão que é destacado como princípio que rege o uso da internet no Brasil (art. 3º, I) e para o pleno exercício do direito de acesso à internet (art. 8º, *caput*), desde o PL original proposto pelo Executivo (Brasil, 2011), sendo ainda incluído no *caput* do art. 2º pelo relator do PL na Câmara dos Deputados, Deputado Alessandro Molon (então PT/RJ) (Brasil, 2011, p. 26-27).

Assim, a proteção à liberdade de expressão e a privacidade constam repetidamente enquanto motivação e objetivo pretendido do Marco Civil da Internet, recorrentemente expresso no relatório do substitutivo apresentado em plenário, como se extrai do seguinte trecho:

“Buscamos, com esse Substitutivo, preservar ao máximo o caráter principiológico e de promoção da Internet como um espaço de cidadania que o Marco Civil da Internet traz em sua origem. Esperamos, assim, respeitar a natureza da Internet como um espaço de liberdade, descentralizado, distribuído, que transcende as fronteiras nacionais e propicia contato direto entre as pessoas de todo o mundo, mesmo quando

separadas por culturas diferentes ou governos distintos e, em alguns casos, antagônicos.

O Marco Civil busca garantir que a Internet continue funcionando como uma rede cibernética de comunicação e controle, de acordo com os princípios basilares de quem a criou. A comunicação interativa depende do controle técnico dos protocolos, razão pela qual na rede há rastros digitais daquilo que todos fazem. Nesse sentido, é muito mais fácil monitorar o que os usuários fazem na internet do que no mundo real, e é justamente o monitoramento indevido que deve ser veementemente coibido, sob pena de tolher a liberdade de expressão e a privacidade.” (Brasil, 2014a, p. 49).

O segundo parágrafo deixa claro, ainda, a preocupação com a utilização dos dados dos usuários para tolher a liberdade de expressão e a privacidade, na esteira do quanto foi revelado pelo Caso Snowden. Desta feita, naquele momento, o uso dos dados como ferramenta de controle pelos governos guiou a proposição do Marco Civil da Internet, que são apontados, ao lado de grupos econômicos, como responsáveis por ações que visam restringir o tráfego, eliminar conteúdos, censurar informações e restringir a liberdade de expressão na *internet*. É nesse sentido que são criados limites à responsabilização das plataformas digitais, como uma forma de impedir o monitoramento e vigilância e que se punam “excessivamente jornalistas e blogueiros” por meio de uma “autocensura gerada pela ameaça de punição” (Brasil, 2014a, p. 50).

Propôs-se, ainda, no Substitutivo apresentado ao Plenário (Brasil, 2014a, p. 38-39), tendo em vista o caráter transnacional das principais plataformas digitais, regras voltadas a garantir o *enforcement* da lei brasileira. Assim, foi sugerida a inclusão de dois artigos, 11 e 12, que previam, respectivamente, que no que diz respeito a operações de coleta, armazenamento, guarda e tratamento de dados pessoais e comunicações de brasileiros, aplica-se a legislação brasileira e a possibilidade de que o governo brasileiro, mediante decreto, institua a obrigatoriedade de que os dados de brasileiros sob a guarda de grandes provedores de aplicações, sejam localizados em bancos de dados localizados em território brasileiro, prevendo-se, ainda, no art. 13, sanções pelo descumprimento.

No texto original do PL nº 2.126, a previsão sobre a isenção de responsabilidade das plataformas digitais sobre o conteúdo postado por terceiros se encontrava nos arts. 14 a 16, porém a inclusão de outros artigos no substitutivo apresentado em plenário levou a disciplina aos arts. 19 a 21 (Brasil, 2014a, p. 43-44). Na justificativa do texto original, de autoria do Executivo, destaca-se que a opção pela responsabilização subjetiva e a isenção de responsabilidade das plataformas sobre o conteúdo postado por terceiros, tem como finalidade proteger a ampla liberdade de produção de conteúdo pelos usuários, sem necessidade de aprovação prévia por intermediários, tendo-se como objeto os **usos legítimos**, protegendo-se a

privacidade e a liberdade de expressão, adotando-se o princípio da presunção de inocência e identificando os casos de “abusos como eventos excepcionais” (Brasil, 2011, p. 10).

O relatório da Comissão de Ciência, Tecnologia, Inovação, Comunicação e Informática do Senado Federal, de autoria do Senador Zezé Perrella (à época PDT/MG), ainda destaca que a opção por isentar as plataformas de responsabilização pelo conteúdo postado por terceiros tem fundamento em um princípio de governança e uso da rede, estabelecido pelo Comitê Gestor da Internet, denominado inimizabilidade da rede, segundo o qual “o combate a ilícitos na rede deve atingir os responsáveis finais e não os meios de acesso e transporte, sempre preservando os princípios maiores da defesa da liberdade, da privacidade e do respeito aos direitos humanos” (Brasil, 2014e, p. 11).

No relatório do substitutivo, a opção pela isenção dos provedores de conexão, que viria a ser o art. 18 do Marco Civil da Internet, se justifica tanto para garantir o potencial de inovação da rede, quanto para “proteger os diversos intermediários responsáveis apenas pela transmissão e roteamento de conteúdos, reconhecendo que a responsabilidade por eventuais infrações por danos decorrentes de conteúdo gerado por terceiros cabe àqueles que a cometeram, e não àqueles que mantêm a infraestrutura necessária para o trânsito de informações na internet” (Brasil, 2014a, p. 43-44).

Já quanto à isenção de responsabilidade do provedor de aplicações, objeto desta pesquisa, o relatório justifica a previsão no tópico “Fim da Censura Privada: responsabilidade civil por danos gerados por terceiros” (Brasil, 2014a, p. 44-45). O art. 15 do texto original era mais enxuto, estabelecendo que o provedor de aplicações somente seria responsabilizado no caso de, após decisão judicial específica de indisponibilização de conteúdo, deixasse de atender à ordem, além de um parágrafo único em que se dispõe que a decisão deve indicar de maneira clara e específica o conteúdo infringente. Desde o texto inicial, já se previa que eventuais disposições legais em contrário, como no caso da proteção da criança e do adolescente no Estatuto da Criança e do Adolescente, Lei nº 8.069/1990, quanto à pornografia infantil, não se aplica a isenção de responsabilidade das plataformas, que, apenas mediante simples notificação sobre a existência do conteúdo deveria indisponibilizá-lo.

No art. 20 do substitutivo, que posteriormente se tornou o art. 19 do Marco Civil da Internet, o legislador fez questão de pormenorizar a matéria. Foi incluído expressamente no *caput* deste artigo o objetivo de proteger a liberdade de expressão e impedir a censura, mantendo-se a internet como um “espaço de livre e plena expressão”. O relatório ainda indica que, através da previsão da identificação do conteúdo específico a ser indisponibilizado, busca-se evitar a indisponibilização completa de serviços ou páginas, como *blogs* ou portais de

notícias. Foi acrescentado, ainda, o §2º, que restringe a isenção de responsabilidade em casos que envolvam direitos autorais e direitos conexos, o que foi solicitado por entidades da sociedade civil, o setor privado e o Governo, especialmente o Ministério da Cultura que trabalhava, à mesma época da tramitação do Marco Civil da Internet, na reforma da Lei de Direitos Autorais (Brasil, 2014a, p. 44-45). Acrescenta-se, assim, expressamente, a previsão de que quanto a direitos autorais e direitos conexos, a mera notificação, feita por qualquer usuário, gera a responsabilização da plataforma por aquele conteúdo. Desta forma, o Marco Civil da Internet cria um incentivo para que as plataformas ajam, sem intervenção do Poder Público e especialmente do Judiciário, para indisponibilizar este conteúdo.

A Emenda Parlamentar da Câmara dos Deputados nº 51, de autoria do Deputado Domingos Sávio, líder da minoria, propôs a inclusão da possibilidade de, mediante notificação extrajudicial, o provedor de aplicações de *internet* excluir conteúdo que possa configurar crime de calúnia, difamação ou injúria, o que foi rejeitado pelo Relator, sob a fundamentação de que cabe ao Poder Judiciário decidir sobre a exclusão de conteúdo, e não a um ente privado, uma vez que este último é dotado de interesses econômicos e políticos e poderia, em tese, realizar uma espécie de censura (Brasil, 2012b, p. 72).

A Emenda ao PLC nº 21/2014 do Senado Federal nº 36, proposta pelo Senador Aloysio Nunes Ferreira (PSDB) propunha acrescentar ao rol de conteúdos que poderiam ser excluídos mediante notificação extrajudicial conversas privadas de cunho sexual indevidamente divulgadas e outros materiais que violem a dignidade da pessoa humana. O parecer da Comissão de Ciência, Tecnologia, Inovação, Comunicação e Informática do Senado Federal, entretanto, foi pela rejeição da emenda, sob a fundamentação de que a exclusão de conteúdo sem determinação judicial trata-se de medida excepcional, que não seria justificada em outros casos além daqueles já previstos (Brasil, 2014e, p. 31).

Outra proposta que poderia ter conferido maior transparência, celeridade e devido processo, além de gerar mais obrigações às plataformas digitais no que diz respeito ao conteúdo postado por terceiros, consta da Emenda ao PLC nº 21/2014 do Senado Federal nº 41, de autoria do Senador Aloysio Nunes Ferreira (PSDB), que previa a inclusão de dispositivo para a “solicitação particular de indisponibilização de conteúdo gerado por terceiro”, garantindo ao usuário das plataformas digitais o recebimento de um protocolo da solicitação, um prazo para resposta, além de uma resposta por escrito, informando as ações que foram tomadas ou as razões de não acatar o pedido (Brasil, 2014c). Entretanto, a emenda foi rejeitada em parecer da Comissão de Constituição e Justiça do Senado Federal, de relatoria do Senador Vital do Rêgo (MDB/PB) sob a fundamentação de que tal previsão já estava incluída no inciso XIII do art. 7º

do Marco Civil da Internet, ao invocar-se a aplicação da proteção ao consumidor, e, consequentemente, impor-se a responsabilização civil dos provedores de aplicação pelos danos causados a terceiros (Brasil, 2014d), bem como pelo parecer da Comissão de Ciência, Tecnologia, Inovação, Comunicação e Informática do Senado Federal, que indicou a rejeição da emenda por entender que a lei já garantia o acesso à informação e a aplicação de normas e proteção ao consumidor (Brasil, 2014e, 33).

Ainda, é importante destacar que o Marco Civil da Internet não criou distinções entre os serviços oferecidos por provedores de aplicações de *internet*, optando por uma disciplina ampla e abrangente, de forma a conferir o mesmo tratamento a mecanismos de busca, redes sociais, *blogs* e *wikis*, enfim, quaisquer provedores de aplicação que atuem no Brasil. Optou-se, também, por não regular quaisquer aspectos criminais nessa oportunidade, restringindo-se aos aspectos civis. Portanto, a responsabilidade das intermediárias pelo conteúdo postado por terceiros, discutida nessa dissertação, é de natureza civil.

Ante o reconhecimento da necessidade de celeridade nas decisões que versem sobre indisponibilização de conteúdo, mediante a velocidade com que materiais postados nas mídias digitais são veiculados, o substitutivo acrescenta os §§ 3º e 4º ao art. 20 (art. 19, §§ 3º e 4º do Marco Civil da Internet), que dispõe que no caso de danos relacionados à honra, reputação e direitos da personalidade pode-se aplicar o rito sumaríssimo dos juizados especiais, bem como pode ser concedida liminar para a indisponibilização do conteúdo.

O texto inicial do PL nº 2.126/2011 previa no art. 16 que a plataforma deveria, ao indisponibilizar conteúdo mediante ordem judicial, informar ao usuário que publicou tal conteúdo do cumprimento da decisão judicial. No substitutivo apresentado em plenário, a preocupação com a transparência foi ampliada, inclusive tendo em vista as novas possibilidades de exclusão de conteúdo mediante notificação simples que sobrevieram ao texto original. Assim, a comunicação quanto aos motivos e informações sobre a indisponibilização do conteúdo passam a figurar como elemento necessário à consecução da ampla defesa e do contraditório. Acrescentou-se, ainda, a previsão de que, quando solicitado pelo usuário, o conteúdo excluído deveria ser substituído pela motivação ou ordem judicial que determinou a exclusão do conteúdo, conferindo publicidade à decisão. A motivação para tal previsão é informar os usuários, difundir o conhecimento sobre o que pode ou não ser postado na *internet* e tornar “público eventuais abusos na retirada indevida de conteúdo, de forma a trazer maior equilíbrio, transparência e liberdade de expressão” (Brasil, 2014a, p. 46-47).

A última disposição sobre a responsabilização das plataformas sobre o conteúdo postado por terceiros foi a inclusão do art. 22 (art. 21 do Marco Civil da Internet), que prevê a

responsabilidade subsidiária dos provedores de aplicações de internet no caso de divulgação de imagens, vídeos ou outros materiais que contenham nudez ou atos sexuais de caráter privado, cuja disseminação não tenha sido aprovada pelos participantes. Assim, exige-se que as plataformas atuem de maneira diligente para indisponibilizar o conteúdo, sob pena de serem civilmente responsabilizadas. A inclusão de tal dispositivo se dá mediante a repercussão casos de suicídio de duas jovens após a divulgação de materiais desta natureza (Brasil, 2014a, p. 17). Trata-se, desta forma, sem prejuízo à correção da medida, de uma previsão casuística, incluída no texto legal em razão da coincidência de, durante o trâmite da norma, terem ocorrido e sido amplamente divulgados esses casos.

Assim, da análise do trâmite legislativo do Marco Civil da Internet, especialmente quanto ao Capítulo III- Da provisão de conexão e de aplicações de internet, Seção III- Da Responsabilidade por Danos Decorrentes do Conteúdo Postado por Terceiros, que no texto original, proposto pelo Executivo Federal, a norma de isenção de responsabilidade dos provedores de aplicação de *internet* era abrangente, restringindo-se apenas no caso de disposição legal em contrário, como no supramencionado caso de divulgação de material pornográfico infantil, com previsão no ECA. Já no substitutivo apresentado em plenário há exceções incluídas diretamente ao texto do Marco Civil da Internet, que implicam na obrigação das plataformas de, mediante notificação simples, realizada por qualquer indivíduo, atuar diligentemente para a indisponibilização de conteúdo postado por terceiros, sob pena de ser civilmente responsabilizada pela desídia. Assim, quanto a conteúdo protegido por direitos autorais ou direitos conexos e a divulgação de materiais pornográficos sem autorização, a isenção de responsabilidade é afastada.

É importante ressaltar que, durante as discussões e o trâmite legislativo do Marco Civil da Internet, desde o início do processo em 2007 até a sanção em 2014, a principal preocupação quanto à liberdade de expressão na *internet* estava associada à vigilância estatal, preocupação da escola tradicional da liberdade de expressão, potencializada após o desvelamento do Caso Snowden. Até aquele momento, a disseminação de *fake news* e discurso de ódio de maneira sistemática e organizada por meio das plataformas digitais não era prioridade do legislador, o que se modificou mediante o caso das eleições presidenciais estadunidenses de 2016 e a votação do *Brexit*. No Brasil, as eleições de 2018 foram um ponto de inflexão, na medida em que as plataformas digitais foram utilizadas para disseminar *fake news* e discurso de ódio durante a campanha eleitoral. Mediante a inexistência de legislação específica para enfrentar a questão, e tendo em vista a proteção à integridade do processo eleitoral brasileiro, o Tribunal Superior Eleitoral atuou ostensivamente, durante as eleições de 2020 e 2022.

3.6.2. O Programa de Enfrentamento à Disseminação de Desinformação com Foco nas eleições de 2020

Inicialmente, é importante destacar que o Programa de Enfrentamento à Disseminação de Desinformação do Tribunal Superior Eleitoral está voltado, entre outros objetivos, a instituir mecanismos de combate à disseminação de desinformação na *internet*, por meio de acordos e ações conjuntas com as principais plataformas digitais em operação no Brasil e organizações da sociedade civil, especificamente no que diz respeito a conteúdos voltados a desacreditar o próprio processo eleitoral ou a Justiça Eleitoral, não tendo em seu escopo, portanto, conteúdo que não atinja diretamente as eleições propriamente ditas. Nesse sentido, o objeto do Programa é o combate à desinformação sobre o processo eleitoral, abarcando desde o registro de candidaturas até a diplomação dos eleitos; o funcionamento das urnas eletrônicas; a Justiça Eleitoral; a segurança sanitária, tendo em vista o contexto da pandemia de Covid-19; e outros atos relativos às eleições (Brasil, 2020a, p. 7).

O marco normativo do Programa de 2020 inclui a Convenção Americana sobre Direitos Humanos (Pacto de São José da Costa Rica); a Constituição Federal de 1988; o Código Eleitoral; a Lei das Eleições; a Resolução do TSE nº 23.610/2019 sobre propaganda eleitoral, utilização e geração do horário gratuito e condutas ilícitas em campanha eleitoral, que instituiu a obrigatoriedade de os candidatos, partidos e coligações verificarem a fidedignidade das informações veiculadas em campanha, sob pena de direito de resposta (art. 9º) e a vedação do disparo em massa de conteúdos (art. 28, IV, a e b e 34) (Brasil, 2020a, p. 26-27); o Marco Civil da Internet; a Lei Geral de Proteção de Dados Pessoais e a Lei nº 13.188/2015, sobre direito de resposta ou retificação do ofendido em matéria divulgada, publicada ou transmitida por veículo de comunicação social (Brasil, 2020a, p. 7).

O Programa de 2020 conta com seis eixos de atuação, contemplando medidas de curto, médio e longo prazo, dispostos nos incisos do art. 3º da Portaria do Tribunal Superior Eleitoral nº 663/2019, quais sejam:

- “I - Organização interna: visa à integração e coordenação entre os níveis e áreas que compõem a estrutura organizacional da Justiça Eleitoral e à definição das respectivas atribuições na execução de ações de elaboração e divulgação de contrainformação a práticas de desinformação;
- II - Alfabetização midiática e informacional: visa à capacitação voltada à identificação e checagem de práticas de desinformação e à compreensão sobre o processo eleitoral, em especial quanto ao funcionamento e segurança das urnas eletrônicas e a situações passíveis de conduzir à anulação de votos e eleições;

III - **contenção à desinformação**: visa à instituição e ao aperfeiçoamento de medidas concretas voltadas a desestimular práticas de desinformação;

IV - **Identificação e checagem de desinformação**: visa à instituição e ao aperfeiçoamento de métodos de identificação de possíveis práticas de desinformação e de sua checagem;

V - Aperfeiçoamento do ordenamento jurídico: visa à revisão e ao aperfeiçoamento do ordenamento jurídico existente sobre a desinformação; e

VI - Aperfeiçoamento de recursos tecnológicos: visa ao desenvolvimento e ao aperfeiçoamento de recursos de tecnologia da informação e das comunicações idôneos, à identificação de práticas de desinformação e à divulgação das respectivas contrainformações.” (Brasil, 2019).

O Programa de Enfrentamento à Disseminação de Desinformação do TSE é pautado em uma reinterpretação dos princípios constitucionais da liberdade de expressão, imprensa e informação à luz da nova realidade comunicacional, ao lembrar que a posição de preferência conferida a tais princípios na Constituição Federal de 1988 se dá em função da consecução da própria democracia, de forma que se reconhece que um “amplo fluxo de informações” e um “debate público robusto e irrestrito” são necessários para a adequada “tomada de decisões da coletividade e para o autogoverno democrático”. Assim, através do reconhecimento de que a disseminação articulada de desinformação resulta em prejuízos ao debate público e à liberdade de informação e expressão, “inundando” as redes de conteúdos falsos, dispõe-se a, conjuntamente com as plataformas digitais e a sociedade civil, atuar de maneira a promover medidas para mitigar os efeitos perniciosos da disseminação destes materiais (Brasil, 2020a, p. 6).

A complexidade do fenômeno também é apontada pelo Programa, de maneira que se atenta à necessidade de interlocução entre diversos atores que participam da comunicação digital em um processo de permanente atualização, possibilitando a apreensão do fenômeno da desinformação em massa nas mídias digitais através de áreas de especialização distintas. Assim, a formação de câmaras-de-eco ou filtros-bolha e a capacidade de viralização de conteúdos falsos, criando-se um ambiente de desconfiança, inclusive sobre o próprio processo eleitoral, possibilitando a disseminação de *deep fakes*, comportamentos inautênticos coordenados, discursos de ódio e teorias conspiratórias, por meio de novas estruturas de comunicação, é que são os alvos da atuação sistêmica do TSE durante o pleito eleitoral de 2020 (Brasil, 2021, p. 10).

Essa reinterpretação, entretanto, não substitui a preocupação com a tradicional concepção de liberdade de expressão, mirando, portanto, em abusos e tratando-os, assim como no Marco Civil da Internet, enquanto exceções. O cuidado para evitar que o combate à desinformação possa prejudicar a liberdade de expressão, por meio da restrição de conteúdos, desincentivando cidadãos e veículos de se manifestar, chamado de “*chilling effect*”, ou efeito

resfriador do discurso, é um dos nortes do Programa, que se centrou em estratégias multissetoriais e “não regulatórias”, pautadas na: pluralidade informativa e criação de redes de confiança para produção e distribuição de conteúdo de qualidade; ação educacional, capacitando os usuários para identificar desinformação; e no foco em comportamentos, reprimindo comportamentos inautênticos de perfis digitais, “uso de propaganda computacional” e ações coordenadas de propagação de desinformação. Por meio dessa estratégia, o TSE se propôs a atuar no controle do comportamento digital, primeiramente, e apenas secundariamente do conteúdo (Brasil, 2020a, p. 6-7).

Entre as referências teóricas para a execução do Programa de 2020, são citados sete documentos de instituições que estudam a relação entre comunicação digital, eleições e desinformação, quais sejam: 1) Guia para garantir a liberdade de expressão contra a desinformação deliberada em contextos eleitorais (2019), da Relatoria Especial para Liberdade de Expressão da Comissão Interamericana de Direitos Humanos (CIDH) com contribuições do Departamento de Cooperação e Observação Eleitoral e do Departamento de Direito Internacional da Secretaria Geral da Organização dos Estados Americanos (OEA); 2) Declaração Conjunta sobre Liberdade de Expressão e Eleições na Era Digital (2020), do Relator Especial das Nações Unidas (ONU) para a Liberdade de Opinião e Expressão, Representante para a Liberdade dos Meios de Comunicação da Organização para a Segurança e Cooperação na Europa (OSCE) e Relator Especial para a Liberdade de Expressão da Organização dos Estados Americanos (OEA); 3) Relatório da Missão de Observação Eleitoral nas Eleições Gerais 2018 no Brasil (2019) da OEA; 4) Relatório: Abordagem Multidimensional da Desinformação (2018) da Comissão Europeia, Grupo de Especialista de Alto Nível em Fake News e Desinformação Online; 5) Relatório Internet e Eleições no Brasil: Diagnósticos e Recomendações (2019) do InternetLab; 6) White Paper: Recomendações Sistêmicas para combater a desinformação nas Eleições do Brasil (2018) do Instituto Tecnologia & Equidade; e 7) Plano de Ação Contra a Desinformação (2018) da Comissão Europeia e Alto Representante da União para Relações Exteriores e Política de Segurança (Brasil, 2020a, p. 7-11).

Todos os documentos apontados como referenciais teóricos para o Programa de 2020 destacam a necessidade de o combate à desinformação ser realizado de maneira multissetorial, especialmente levando-se em consideração a participação de Governos, sociedade civil e das intermediárias, notadamente as plataformas digitais e, especificamente, as redes sociais. Além disso, apontam a necessidade de empoderamento e educação digital dos usuários para que possam realizar a identificação de conteúdo que contenha desinformação, além de reforçar o

papel da imprensa e do jornalismo no estabelecimento de um ecossistema de mídia diverso e sustentável, possibilitando a pluralidade de ideias e fontes.

O documento elaborado pela Comissão Interamericana de Direitos Humanos (CIDH) com contribuições do Departamento de Cooperação e Observação Eleitoral e do Departamento de Direito Internacional da Secretaria Geral da Organização dos Estados Americanos (OEA), aponta que a responsabilização das intermediárias sobre o conteúdo postado por terceiros pode ter enquanto efeito colateral a censura privada, ou seja, a busca das plataformas de, para evitar serem responsabilizadas, excluírem preventivamente conteúdo que poderia, ao final, ser considerado regular. Nesse sentido, e para proteger a liberdade de expressão, defende que não deve ser elaboradas novas leis que possam implicar em prejuízos ao princípio constitucional, mas sim utilizar-se o arcabouço normativo existente para fundamentar estratégias de combate à disseminação. O documento opta, portanto, por projetos de cooperação entre as plataformas e o Poder Público, com participação multissetorial, que seja transparente e não suprima o debate público (Brasil, 2020a, p. 8).

Já o documento elaborado pela ONU e a OEA destaca que é imprescindível que acordos entre plataformas e Poder Público, sejam estes formais ou não, sejam transparentes. Enquanto formas ativas de combate à desinformação, o documento é refratário à criminalização da divulgação de desinformação, defendendo o uso de estratégias alternativas à exclusão de conteúdos, como a promoção de instâncias independentes de verificação de informações e campanhas de educação pública. Entretanto, no que diz respeito à disseminação de conteúdo que busque criar obstáculos à votação, como informações falsas sobre a data e horário da votação que possa confundir os eleitores, defendem a adoção de leis claras e proporcionais que proíbam a veiculação de tal conteúdo (Brasil, 2020a, p. 8-9).

O relatório elaborado por especialistas da Comissão Europeia, por sua vez, opta por estratégias complementares e multissetoriais que implicam em alterações na forma como funcionam e estão estruturadas as plataformas digitais. Ou seja, ao fundamentar as estratégias propostas em cinco pilares, quais sejam o aumento da transparência, especialmente sobre como o conteúdo é disseminado nas mídias digitais, as fontes de financiamento, as fontes de notícia *online* e processos jornalísticos e a checagem de fatos; alfabetização midiática e informacional dos usuários; empoderamento de jornalistas e usuários para enfrentar a desinformação; promoção à diversidade e sustentabilidade do ecossistema de notícias; e promoção de pesquisas continuadas sobre o impacto da desinformação, o documento opta, ao não promover moderação de conteúdo propriamente dita, como por meio de mecanismos de notificação e exclusão de

conteúdo, por exemplo, por estratégias colaterais que modificam a forma como os conteúdos são disseminados digitalmente (Brasil, 2020a, p. 10).

É importante ressaltar que estes quatro documentos foram elaborados por instituições internacionais e que, à exceção do Relatório da Missão de Observação Eleitoral nas Eleições Gerais 2018 no Brasil (2019) da OEA, não foram elaborados especificamente para o caso brasileiro. Já o Relatório elaborado pela Associação INTERNETLAB de Pesquisa em Direito e Tecnologia, é um documento de uma instituição brasileira voltado especificamente às eleições brasileiras, pautado nos diagnósticos elaborados sobre as eleições de 2018. Assim, este relatório está mais voltado às estratégias de regulamentação, levando-se em consideração o direcionamento de conteúdos realizado pelas plataformas digitais e o *marketing* digital, bem como está alinhado às estratégias que já são adotadas pela legislação eleitoral brasileira em relação a outros meios de comunicação (Brasil, 2020a, p. 10-11).

Desta forma, sugere-se a adoção de regulação específica para o *marketing* digital e impulsionamento de conteúdo e a contratação de influenciadores digitais, bem como a criação de um guia de boas práticas para estes últimos; aprimoramento do sistema de prestação de contas eleitorais, incluindo-se os gastos com *marketing* digital e sobre os bancos de dados dos usuários de plataformas utilizados para envio e direcionamento de propaganda; aprimoramento de canais de denúncia e fiscalização de práticas eleitorais ilícitas, como *spam* eleitoral e discurso de ódio; e, além da moderação de conteúdo, priorizar as estratégias em torno de comportamento digital inautêntico e uso de identidades falsas. Tratam-se, portanto, de estratégias regulatórias que, ao mesmo tempo em que priorizam a participação multissetorial e colaboração entre os atores envolvidos no processo, reafirmam a priorização do interesse público no processo eleitoral, promovendo a fiscalização sobre o funcionamento da comunicação digital, bem como a atuação efetiva para o combate ao que chama de comportamentos abusivos (Brasil, 2020a, p. 10-11).

O Programa de 2020 conta, então, para sua execução e gestão, com um Grupo Gestor, composto por membros do TSE e TREs, e um Comitê Estratégico de Combate à Desinformação- Eleições 2020, com participação de membros do TSE. Cada um dos TREs estabeleceu comitês locais, responsáveis pela interlocução com o TSE, inclusive para fins de reportar casos de desinformação. O Programa foi coordenado, portanto, pelo Grupo Gestor e levado a efeito a cargo da Equipe Executora, em parceria e comunicação com os TREs, com o Comitê Estratégico e com os parceiros (Brasil, 2020a, p. 12-13).

Conforme exposto anteriormente, o Programa de 2020 contou com seis eixos de atuação, subdivididos em: organização interna; alfabetização midiática e informacional;

contenção da desinformação; identificação e checagem de desinformação; aperfeiçoamento do ordenamento jurídico; aperfeiçoamento de recursos tecnológicos (Brasil, 2020a, p. 15).

O eixo 3, sobre contenção de desinformação, foi voltado à adoção de “medidas concretas para detectar, desestimular e conter práticas de desinformação, por meio do direcionamento de recursos tecnológicos” e “atitudes preventivas e repressivas contra comportamentos inautênticos, uso de propaganda computacional, uso de identidade falsa e demais fraudes no acesso à informação dos cidadãos” e outras “ações coordenadas” para a propagação de desinformação. Inicialmente, durante a gestão da Ministra Rosa Weber no TSE, este eixo era voltado à atuação de instituições públicas no combate à desinformação, incluindo a atuação de centro integrado de comando e controle, composto por membros da própria Justiça Eleitoral e outras instituições relacionadas com a realização das eleições tais como as polícias judiciárias, o Ministério Público, as Forças Armadas, etc., e um projeto piloto de monitoramento composto pelo TSE e a Agência Brasileira de Inteligência (Abin), que faz parte do Gabinete de Segurança Institucional. Sob a gestão do Ministro Luís Roberto Barroso, ampliou-se o escopo deste eixo para incluir novas estratégias de combate à disseminação de desinformação com atuação conjunta com atores não-governamentais, notadamente provedores de aplicação de *internet*, entidades da sociedade civil especializadas em desinformação, laboratórios forenses e órgãos de inteligência (Brasil, 2020a, p. 22-23).

Nesse sentido, foram adotadas três medidas cujo objetivo era combater a disseminação de desinformação:

1) engajar as plataformas digitais no enfrentamento a redes de desinformação e comportamento inautêntico, cujo público-alvo era a sociedade civil e os atores envolvidos eram os provedores de aplicação de *internet*, em parceria com o TSE. Por meio desta medida, o TSE promoveu a interlocução com as próprias plataformas digitais, buscando estabelecer parcerias e adesão ao Programa, firmando memorandos de entendimento para estabelecer medidas concretas de combate à desinformação, bem como realizando reuniões periódicas com as plataformas parceiras para discutir possibilidades de atuação conjunta no combate à desinformação (Brasil, 2020a, p. 24);

2) criação de rede de monitoramento de práticas de desinformação, com participação do TSE, TREs e Zonas Eleitorais, Gabinete de Segurança Institucional da Presidência da República (GSI e Abin), Ministério da Justiça (Centro Integrado de Comando e Controle – Polícias Judiciárias, Justiça Eleitoral, Ministério Público, Forças Armadas), Diretoria de Inteligência da Polícia Federal – DIP, laboratórios forenses, instituições de pesquisa, organizações da sociedade civil, provedores de aplicação de *internet* e cidadãos, com o objetivo de: monitorar

as redes coordenadas de disseminação de desinformação e estabelecimento de canais de comunicação para informar sobre mensagens suspeitas; contratação de empresa especializada em monitoramento de redes; implantar Centro Integrado de Comando e Controle para identificar prática de desinformação, estudar possibilidades de medidas de contrainformação e agilizar a resposta da polícia judiciária sobre possíveis crimes; estabelecer canais de denúncia nas plataformas digitais aos usuários e justiça eleitoral em casos de suspeita de disparo em massa e desinformação (Brasil, 2020a, p. 24). Essa medida foi complementada por outra, constante do eixo 6, sobre aperfeiçoamento de recursos tecnológicos, em que se buscou criar um canal, dos usuários para o TSE, para envio de denúncias de disparo em massa (Brasil, 2020a, p. 29-31);

E 3) estabelecimento de rede para compartilhamento de informações verídicas sobre o processo eleitoral (#euvotosemfake), cujo objetivo era produzir e divulgar mensagens com informações verídicas sobre o processo eleitoral, em parceria com as plataformas, parceiros do Programa, partidos políticos e candidatos, inclusive com a criação de um *chatbot* no *WhatsApp* para envio de informação ao eleitorado (Brasil, 2020a, p. 24-25).

Ainda, para minorar o impacto da disseminação de desinformação durante as eleições de 2020, o Programa contou com o eixo 4- identificação e checagem de desinformação, elaborado na esteira de estratégia semelhante adotada já nas eleições gerais de 2018, em que constituiu-se grupo voltado especificamente à checagem de fatos e divulgação de informações verídicas, naquela oportunidade publicadas por meio da página Esclarecimento Sobre Informações Falsas Veiculadas nas Eleições de 2018 e nas redes sociais do TSE. Para o Programa de 2020, foi estabelecida a Coalizão para Checagem- Eleições 2020, com a participação de instituições independentes de checagem de fatos, veículos de mídia e imprensa, provedores de aplicação de internet e TSE e TREs. O TSE e os TREs foram responsáveis, nesse eixo, por identificar e informar às agências de checagem de fatos sobre conteúdos que estavam sendo veiculados e que poderiam conter desinformação para que então os parceiros pudessem realizar a checagem independente. Após ser realizada a checagem, esta era disponibilizada por meio da página no site do TSE “Fato ou Boato”, a qual poderia, inclusive, ser acessada sem cobrança de tráfego de dados, ou seja, ampliando-se o alcance para quaisquer usuários de *internet* (Brasil, 2020a, p. 25-27).

Foram realizados 29 acordos prevendo ações e medidas concretas para o combate à desinformação nas Eleições 2020, que envolveram o combate à desinformação em três frentes: com informação, com capacitação e com foco em controle de comportamentos inautênticos. Nesse sentido, as principais ações de combate à desinformação foram a Coalizão para

Checagem- Eleições 2020, que produziu 274 matérias com verificação de conteúdos, por meio da participação de mais de cem jornalistas e servidores da Justiça Eleitoral; o *chatbot* do *WhatsApp* para informações sobre as eleições, que teve mais de 18 milhões de interações com mais de 1 milhão de usuários; a Página Fato ou Boato, que teve mais de 13 milhões de visualizações entre 27 de outubro e 30 de dezembro de 2020 e 400 milhões de impressões no *YouTube*; a campanha de TV e rádio “Se for *Fake News*, não Transmita”, que alcançou cerca de 130 milhões de brasileiros; o Programa Minuto de Checagem, com 4,5 milhões de visualizações no *YouTube*; o canal de denúncias de disparo em massa, que recebeu 5.229 denúncias, resultando em 1.042 contas banidas e mais de 360 mil contas banidas pelo *WhatsApp*; e o monitoramento das redes sociais, com 752 denúncias de desinformação (Brasil, 2021, p. 7-9).

As medidas tomadas pelo Programa se guiaram pelas dificuldades encontradas no combate à disseminação de *fake news* e discurso de ódio sob a égide do Marco Civil da Internet, em que a única exigência, em atendimento ao interesse público, para que seja exercida alguma espécie de ação por parte das plataformas digitais, diz respeito à ordem judicial de remoção de conteúdo, à exceção dos casos de pornografia de vingança e direitos autorais, o que dificulta ou, até mesmo, impossibilita, a tomada de medidas sistêmicas e céleres.

Assim, as diretrizes que nortearam a atuação do Programa de 2020 são a descentralização e multisetorialidade, com participação de diversas áreas de especialização e parceiros; a agilidade, tendo em vista o impacto negativo e a rápida viralização de conteúdos na seara digital; o viés inovador e conectado; eficiência na comunicação interna e externa; e adaptabilidade. Desta forma, o TSE buscou combinar estratégias de disseminação de informação confiável, capacitação e uso da tecnologia para enfrentamento à disseminação de desinformação, formalizar parcerias que garantiram alcance e eficácia às medidas e criar soluções tecnológicas (Brasil, 2021, p. 14).

Um dos principais elementos que diferenciam a comunicação digital das denominadas mídias de massa tradicionais é o fato de que o conteúdo disseminado digitalmente é majoritariamente coproduzido pelos próprios usuários, ao mesmo tempo em que é veiculado de acordo com a atuação das plataformas no direcionamento de conteúdo. Alinhando a nova estrutura comunicacional com o princípio da liberdade de expressão, o Programa de 2020 realizou acordos com algumas das principais plataformas digitais em operação no Brasil, acessadas diariamente por milhões de usuários, notadamente o *Facebook* (*Facebook*, *Instagram* e *WhatsApp*), *Google* (*Google* e *YouTube*) e *TikTok*, com o objetivo de, por meio da veiculação de informação confiável e acessível, modificar a própria arquitetura das plataformas, tendo em

vista o interesse público, para priorizar a disseminação de conteúdo informativo, de maneira clara e facilmente identificável, diretamente nas páginas dos serviços oferecidos pelas intermediárias.

No combate à desinformação por meio de comportamentos inautênticos, o TSE promoveu uma parceria com a plataforma *WhatsApp*, gerida pelo *Facebook*, para, com fulcro no art. 57-J da Lei nº 9.504/1997 c/c os arts. 28, IV, a e b e art. 34 da Resolução-TSE nº 23.610/2019, coibir a distribuição em massa de conteúdos, o que foi apontado enquanto um dos maiores meios de veiculação de falsidades ao longo das eleições de 2018. As denúncias eram realizadas por meio de formulário e incluíam informações sobre o usuário que havia enviado a mensagem, identificado por um número de telefone, bem como informações úteis para a verificação do conteúdo, como a data em que a mensagem foi recebida e o conteúdo do material (Brasil, 2021, p. 65-67).

Por meio do acordo foram processadas 5.229 denúncias de usuários que poderiam estar realizando disparo em massa. Após uma análise prévia do TSE (que descartou 207 denúncias por não estarem relacionadas à eleição), as denúncias eram encaminhadas para a plataforma, de maneira extrajudicial, o que iniciava um procedimento de verificação de descumprimento aos termos de uso da plataforma. Dos 5.022 procedimentos instaurados pelo *WhatsApp*, 3.527 diziam respeito a contas válidas (ou seja, excluídos contatos duplicados ou não existentes), dos quais quase 30%, ou 1.042, foram banidas em razão do disparo em massa de mensagens relacionadas às eleições. Além disso, o *WhatsApp* banuiu autonomamente mais de 360 mil contas no Brasil durante o período eleitoral, entre setembro e novembro de 2020, pela mesma razão (Brasil, 2021, p. 65-67).

Além desta parceria diretamente com o *WhatsApp*, o TSE estabeleceu um procedimento para identificação e encaminhamento de práticas de desinformação e comportamento inautêntico que envolveu o processamento de “apontamentos”, nome dado aos casos de suspeita das práticas irregulares. O processo se iniciava por meio de uma denúncia de qualquer pessoa que tivesse acesso a mensagem, publicação ou notícia suspeita, por meio da Ouvidoria do TSE, ou pelos pontos focais de combate à desinformação dos TREs, bem como por quaisquer dos participantes de uma rede de monitoramento, composta por instituições da sociedade civil que celebraram memorandos de entendimento e termos de adesão com o TSE para compartilhar informações, conhecimentos e apontamentos de casos de desinformação, quais sejam a *Safernet* Brasil, o Instituto Igarapé, o ITS Rio, a FGV/DAPP e o DFR *Lab* (*Atlantic Council*), bem como as agências de checagem que fizeram parte da Coalizão para Checagem- Eleições de 2020, quais sejam AFP Checamos, Agência Lupa, Aos Fatos, Boatos.org, E-farsas, Estadão Verifica,

Fato ou Fake, Projeto Comprova e Uol Confere; além de duas ferramentas tecnológicas de monitoramento de redes sociais, quais sejam a *Buzzsumo* e o *Knewin Monitoring*. Nesta primeira etapa, eram identificados os casos em que houvesse suspeita de práticas de desinformação ou comportamento inautêntico (Brasil, 2021, p. 67-69).

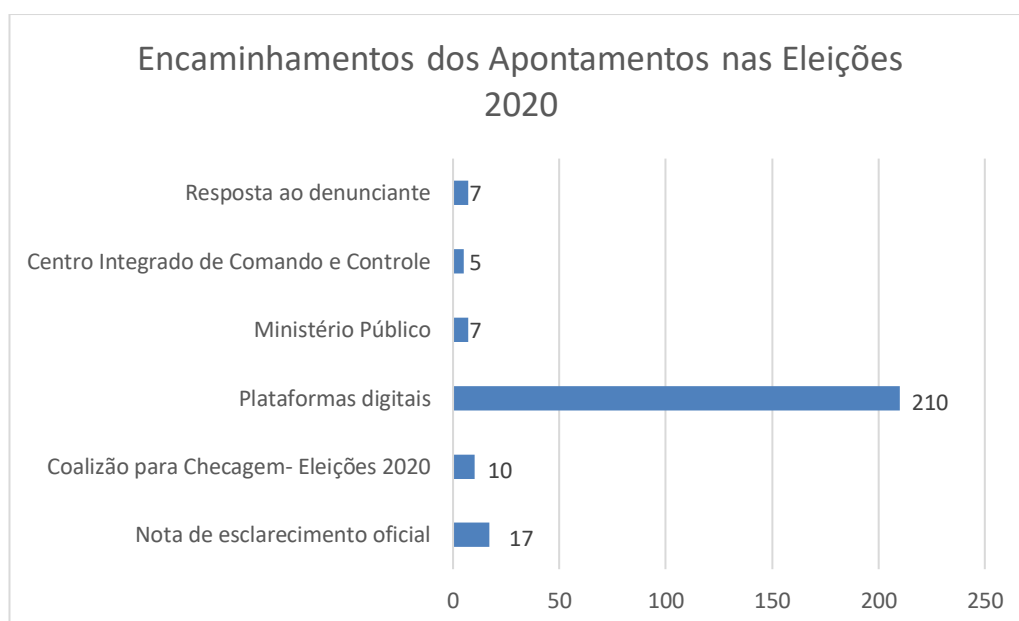
As ferramentas tecnológicas de monitoramento foram importantes no processo de identificação dos conteúdos veiculados digitalmente. O *Buzzsumo* permite analisar os temas e conteúdos veiculados por meio das plataformas *YouTube* e *Twitter*, através de pesquisa por palavras-chave, permitindo observar em tempo real de que maneira a desinformação é veiculada, realizando inclusive comparativos entre diferentes períodos, bem como quais temas eram mais comentados em um determinado momento, além de monitorar o desempenho das páginas oficiais da Justiça Eleitoral. No mesmo sentido, o *Knewin Monitoring* acompanha em tempo real tendências e estratégias de desinformação nas plataformas *Facebook*, *Twitter*, *YouTube*, *Instagram*, *Google*, fóruns e sites (Brasil, 2021, p. 72-73).

É importante reforçar que esse monitoramento foi circunscrito a questões relativas à integridade do Processo Eleitoral e da Justiça Eleitoral propriamente ditas, excluídos conteúdos sobre candidatos, partidos políticos ou coligações, bem como o exercício legítimo do direito de crítica, ainda que “contundente” e “ácida”. Era objeto de verificação, portanto:

“desinformação sobre: o Processo Eleitoral em suas diferentes fases, desde o registro de candidaturas até a diplomação dos eleitos; o funcionamento da urna eletrônica e demais aspectos do processo eletrônico de votação; a Justiça Eleitoral, seus membros, servidores e demais colaboradores; a segurança sanitária do Processo Eleitoral no contexto da pandemia da Covid-19; e outros atos relativos à organização e realização das eleições.” (Brasil, 2021, p. 69).

Os apontamentos eram, então, catalogados em um banco de dados e encaminhados ao Comitê Banco de Dados, uma subseção do Comitê Estratégico de Combate à Desinformação do TSE, composto por oito membros do TSE, que realizavam o encaminhamento do apontamento, de acordo com a avaliação do caso, direcionando para um ou mais dos seguintes procedimentos: 1) elaboração de nota de esclarecimento oficial do TSE; 2) verificação da Coalizão para Checagem- Eleições 2020; **3) tomada de ação por parte das plataformas digitais parceiras do Programa**; 4) apuração de ilícitos pelo Ministério Público Eleitoral; 5) apuração de crimes pela Polícia Federal; 6) apuração de crimes pelo Centro Integrado de Comando e Controle Nacional do Ministério da Justiça e Polícia Federal; 7) resposta ao denunciante; 8) averiguação de ilícito no dia da votação pelo TRE; e 9) arquivamento (Brasil, 2021, p. 69-72).

Entre os possíveis encaminhamentos dos apontamentos realizados ao longo do período eleitoral, houve uma ampla prevalência entre os casos arquivados, totalizando 506 apontamentos, que representaram pouco mais de dois terços (67%) dos 752 totais, seja por não ser conteúdo passível de checagem (272), por estarem fora do escopo do Programa (90), ou por ausência de repercussão (144). Entre os casos que não foram arquivados, entretanto, 85% dos casos foram encaminhados para a tomada de ação por parte das plataformas digitais parceiras do Programa, sendo 210 dos 246 apontamentos que geraram resposta (Brasil, 2021, p. 75). No gráfico a seguir, estão representados os apontamentos durante as eleições 2020, de acordo com o encaminhamento dado pelo Comitê Banco de Dados.



Para o encaminhamento às plataformas digitais parceiras do Programa, eram considerados os critérios e indicadores de

“(i) existência de checagem do mesmo conteúdo ou de emissão de nota oficial sobre o tema no contexto das Eleições 2020; (ii) previsão de medida para o apontamento nas políticas da plataforma envolvida; (iii) engajamento (quantidade de compartilhamentos, visualizações e comentários); e (iv) gravidade do apontamento.” (Brasil, 2021, p. 71).

Ou seja, o alcance de uma mensagem por meio das plataformas digitais era um dos critérios para o encaminhamento ou não de um apontamento. Confirmado o preenchimento dos requisitos, era realizado o encaminhamento às plataformas digitais parceiras que, então,

realizavam uma análise de acordo com seus termos e políticas de uso, podendo adotar as seguintes medidas, a depender destes termos:

“redução no alcance da publicação; marcação do conteúdo como violador das políticas contra a desinformação; inclusão de etiqueta relativa à checagem de fatos; emissão de alertas no momento de compartilhamento; desmonetização; remoção de conteúdo; suspensão ou banimento da conta do usuário; entre outras consequências.” (Brasil, 2021, p. 71).

Tendo em vista que a remoção do conteúdo é considerada medida mais gravosa e, portanto, excepcional, somente foi adotada a solicitação de remoção do conteúdo de acordo com o seguinte rol de ocorrências:

“(i) risco à segurança sanitária do processo de votação; (ii) incentivo ao absenteísmo ou outra desinformação destinada a impedir o voto ou causar confusão quanto à realização da eleição; (iii) incitação a atos violentos, especialmente contra servidores, membros e colaboradores da Justiça Eleitoral ou mesários; (iv) discurso de ódio; (v) indícios de comportamento inautêntico; (vi) indícios de financiamento irregular; (vii) falseamento da origem, por meio da utilização indevida de símbolos oficiais da Justiça Eleitoral; e (viii) outros casos excepcionais envolvendo crimes e/ou circunstâncias graves capazes de impactar negativamente o Processo Eleitoral.” (Brasil, 2021, p. 71).

O relatório de ações e resultados do TSE para o Programa de 2020 aponta que o alto número de arquivamentos decorre do privilégio pela liberdade de expressão, de maneira que em casos nos quais havia dúvidas quanto ao enquadramento nos critérios para identificação da ocorrência de desinformação, como em caso de críticas contundentes ou não passíveis de checagem, optou-se pelo arquivamento (Brasil, 2021, p. 75).

Após ser encaminhado às plataformas digitais, os encaminhamentos eram tratados de acordo com os termos de uso de serviço, regras de moderação de conteúdo e comportamento e políticas de comunidade das próprias mídias. Essas regras não necessariamente eram voltadas especificamente ao processo eleitoral e, quando eram, poderiam ser de difícil acesso, não estarem disponíveis em português ou estavam pouco claras. Entretanto, a atuação colaborativa das plataformas foi um fator fundamental para o funcionamento do Programa de 2020, adotando medidas para combater a disseminação de desinformação sobre o processo eleitoral (Brasil, 2021, p. 78).

Os Memorandos de Entendimento assinados pelas plataformas digitais e o Tribunal Superior Eleitoral para o pleito de 2020, por se tratar de iniciativas conjuntas, não previam hipóteses de responsabilização das plataformas pelo descumprimento aos termos da parceria. O Memorando de Entendimento-TSE nº 41/2020, assinado com o *Facebook*, para a rede social

homônima e o *Instagram*, previa, na cláusula terceira, a instituição da ferramenta Megafone, a ser implementada na *timeline* dos usuários da plataforma *Facebook*, com informações confiáveis sobre as eleições e o desenvolvimento de *stickers* na plataforma *Instagram*, com a mesma finalidade (Brasil, 2020c, p. 3). O Memorando de Entendimento-TSE nº 40/2020, assinado com o *Google*, sobre o mecanismo de busca e a plataforma *YouTube*, previa a facilitação da comunicação entre o TSE e as plataformas, e se comprometia a colaborar com a execução das medidas previstas, notadamente os canais de denúncias e de recebimento de ordens judiciais já existentes nas plataformas (Brasil, 2020b, p. 3).

Entre os casos apontados ao Programa nas eleições de 2020, o relatório destaca dez casos emblemáticos que exemplificam a tomada de decisão no combate à desinformação, de acordo com os critérios definidos. Assim, em casos mais graves e de identificação inequívoca da ocorrência de ilícito, foi adotada a medida mais gravosa, de exclusão do conteúdo, após encaminhamento do Comitê Banco de Dados às plataformas digitais. Ou seja, não há exigência para a exclusão do conteúdo, para o etiquetamento ou outra medida, mas um encaminhamento com indicação da ação sugerida e da gravidade do fato, para que a tomada de decisão possa ser elaborada pela plataforma (Brasil, 2021, p. 79-80)

Exemplos de exclusão de conteúdos são o caso do Programa Voto em Casa, que afirmava se tratar de uma iniciativa do Poder Público, mediante a pandemia de Covid-19, para que os eleitores pudessem votar *online*, através de inscrição em site. Foi identificado que se tratava da prática de desinformação e *phishing*, ou seja, uso de ferramenta enganosa com o objetivo de coletar dados e informações do usuário. O site para suposto acesso ao Programa Voto em Casa foi excluído, mediante encaminhamento ao provedor de hospedagem, além de ter sido aberta investigação pela Polícia Federal e publicada nota oficial de esclarecimento pelo TSE. Caso semelhante foi a publicação de um vídeo no *YouTube* que incentivava o absenteísmo, ao supostamente explicar como o eleitor poderia utilizar um aplicativo que burla a localização de GPS para comprovar, por meio do aplicativo e-Título, que o eleitor não estaria no domicílio eleitoral, de maneira a poder justificar a ausência às eleições. Após o encaminhamento ao *YouTube*, o vídeo foi excluído por descumprir os termos da plataforma. outro caso que justificou a exclusão do perfil que publicou desinformação foi o de página no Instagram que utilizava o logotipo das Eleições 2020 e do TSE para se passar por uma página oficial, tendo sido excluída pela plataforma (Brasil, 2021, 80-81; 89).

Outros casos, porém, não justificaram a exclusão do conteúdo ou, até mesmo, apesar do encaminhamento por parte do Comitê Banco de Dados, após a avaliação, não tiveram medidas tomadas pelas plataformas digitais, tendo em vista que estas, por sua vez, entenderam não se

tratar de violação aos termos de uso. Por exemplo, em um caso de um vídeo que alegava a possibilidade de fraude nas urnas eletrônicas com uma suposta “réplica” caseira do aparelho, o Comitê Banco de Dados identificou que se tratava de desinformação, uma vez que a “urna” do vídeo não se assemelhava à real, bem como constatou o critério de repercussão do conteúdo que havia sido publicado por meio do *YouTube* e replicado e disseminado por meio de outras plataformas tendo alcançado, apenas no *Facebook*, 15 mil compartilhamentos. Houve encaminhamento tanto à Assessoria de Comunicação do TSE, que elaborou nota de esclarecimento oficial, quanto à Coalizão para Checagem- Eleições 2020, o que gerou publicação de conteúdo informativo por parte de agências de checagem de fatos, quanto para as plataformas *YouTube*, *Instagram* e *Facebook*. Enquanto o *Facebook* optou por etiquetar o conteúdo enquanto “sem contexto”, indicando o *link* com informações oficiais do TSE e de agências de checagem de fato, o *YouTube* optou por não tomar nenhuma medida, por não entender que o caso violava as políticas de uso da plataforma (Brasil, 2021, p. 92-93).

Caso semelhante ocorreu em vídeo que afirmava haver divergências entre o boletim da urna eletrônica impresso e o que constava do aplicativo do TSE, veiculado por *Facebook*, *Instagram* e *Twitter*, em que, na verdade, estavam sendo comparados boletins de urna de zonas eleitorais distintas. Apesar da verificação quanto à inveracidade dos fatos narrados no vídeo, não houve uniformidade no tratamento oferecido ao caso, de forma que “alguns dos conteúdos foram removidos das plataformas, alguns sofreram inclusão de aviso de conteúdo falso e outros foram mantidos” (Brasil, 2021, p. 90-91).

Tendo em vista essa divergência entre os critérios e as análises dos encaminhamentos feitos pelo TSE para as plataformas digitais, concluiu-se que é necessário aprimorar tanto o canal de encaminhamento de apontamentos, tanto para que seja dado *feedback* das plataformas em relação às ações tomadas após a notificação, quanto para que sejam mais bem elaboradas e aclaradas as políticas das plataformas em relação às eleições brasileiras (Brasil, 2021, p. 133).

O TSE se propõe, ainda, tendo em vista que por meio do Programa de 2020 desenvolveram-se estratégias de combate à disseminação em massa de desinformação, a colaborar na elaboração de novas normas que possam regulamentar o tema no Brasil. Assim, propõe-se a, por um lado, elaborar relatórios que permitam aprimorar a atuação do próprio TSE nas eleições seguintes, e por outro participar e colaborar no processo legislativo, notadamente na discussão do PL nº 2.630 – Projeto de Lei Brasileira de Liberdade, Responsabilidade e Transparência na Internet, também conhecido como “PL das *fake news*” (Brasil, 2020a, p. 28-29).

3.6.3. O Programa Permanente de Enfrentamento à desinformação da Justiça Eleitoral: Plano Estratégico Eleições 2022

Na esteira do planejamento, execução e avaliação do Programa desenvolvido pelo TSE para as eleições de 2020, o Tribunal optou por tornar o Programa, que originalmente havia sido criado para durar apenas as eleições municipais de 2020, permanente, sendo instituído pela Portaria-TSE nº 510, de 4 de agosto de 2021 o Programa Permanente de Enfrentamento à desinformação da Justiça Eleitoral.

A opção por tornar o programa permanente deriva da percepção de que os efeitos da desinformação nas mídias digitais sobre as eleições não se circunscrevem estritamente ao período eleitoral, mas que a produção organizada de conteúdos desinformativos, discurso de ódio, ameaças cibernéticas, incitação ao extremismo e à violência política, que gera danos à integridade eleitoral, se enraíza e estrutura ainda que temporalmente distante do pleito, inclusive por meio de novas plataformas digitais, como serviços de mensageria com pouco ou nenhum controle que passaram a ser usados por grande parte da população (Brasil, 2022a, p. 13-14), como é o caso do Telegram que, apesar do alto número de usuários no Brasil, deixou de colaborar voluntariamente com o Tribunal Superior Eleitoral no contexto eleitoral (Brasil, 2022a, p. 57).

O Marco Normativo do Programa Permanente inclui importantes inovações legais posteriores ao Programa com Foco nas Eleições de 2020, quais sejam: o Decreto nº 10.222/2020, que aprova a Estratégia Nacional de Segurança Cibernética; a Lei nº 14.197/2021, chamada Lei de Crimes contra o Estado Democrático de Direito; a Resolução-TSE nº 23.650/2021 que institui a Política Geral de Privacidade e Proteção de Dados Pessoais no Âmbito da Justiça Eleitoral; e a Resolução-TSE nº 23.650/2019, atualizada pela Resolução-TSE nº 23.671/2021, sobre propaganda eleitoral, utilização e geração do horário gratuito e condutas ilícitas na campanha eleitoral (Brasil, 2022a, p. 15).

Já quanto às referências teóricas, o Programa Permanente expandiu significativamente os documentos utilizados enquanto fundamentação para a elaboração do planejamento e estratégias no combate à desinformação, em relação ao Programa de 2020. Enquanto aquele indicava sete documentos enquanto referencial teórico, o Programa Permanente conta com 45 documentos, subdivididos em categorias: declarações conjuntas da Organização das Nações Unidas, Organização para a Segurança e Cooperação na Europa e Organização dos Estados Americanos (3 documentos); Sistema Interamericano de Direitos Humanos (1 documento);

documentos no âmbito da União Europeia (8 documentos); outros relatórios e documentos (25 documentos); documentos nacionais (8 documentos) (Brasil, 2022a, p. 16-21).

Deste amplo referencial teórico o Programa Permanente destaca quinze tópicos que devem orientar a atuação no combate à disseminação de desinformação e discurso de ódio, destacando-se, especialmente, ações coordenadas, sistemáticas, transparentes e com regras claras e proporcionais, com a participação dos diversos atores envolvidos na comunicação midiática digital, incluindo-se as plataformas digitais, os políticos e os influenciadores digitais. Destaca-se, ainda, a necessidade de transparência quanto à moderação de conteúdo realizada pelas plataformas, e a necessidade de ações que reduzam os incentivos econômicos para a propagação de desinformação e discurso de ódio (Brasil, 2022a, p. 22).

A transparência e clareza das políticas de moderação adotadas pelas plataformas é fator fundamental para a proteção à liberdade de expressão e para garantir isonomia no tratamento dos diferentes casos. A partir da interlocução com as plataformas, portanto, o TSE se propôs a elaborar soluções para tornar mais claras e transparentes tais políticas,

“detalhando, inclusive, o protocolo de atuação em casos de identificação de comportamentos inautênticos coordenados e publicação de informações enganosas sobre o processo eleitoral, bem como hipóteses correlacionadas de violência política, discursos de ódio e demais questões que impactem negativamente as eleições.” (Brasil, 2022a, p.54)

ao mesmo tempo em que sugere a elaboração e publicação de relatórios por parte das plataformas sobre as medidas tomadas, tanto para permitir a fiscalização, quanto para estar disponíveis para jornalistas e pesquisadores.

Além disso, o Programa prevê a instituição de canais de comunicação nas próprias plataformas, disponíveis tanto para os usuários quanto para o TSE, para tornar mais célere o processo de monitoramento, bem como prevê o estabelecimento de um procedimento especial de atuação excepcional para os casos de disseminação de desinformação grave que possa comprometer a higidez do pleito, “como alegações infundadas de fraude no resultado das eleições, ameaças graves e anúncio falso sobre o cancelamento das eleições” (Brasil, 2022a, p. 54).

O conceito de desinformação adotado pelo Programa deriva da análise elaborada pelo Conselho da Europa sobre “*information disorder*”, de autoria de Wardle e Derakhshan (2017), analisado anteriormente nesta pesquisa. Assim, o Programa reconhece as peculiaridades que diferenciam as diversas espécies de desinformação, utilizando um conceito amplo que abarca

os conceitos de desordem informacional e manipulação informacional, tratando por desinformação, portanto:

“qualquer informação ou conteúdo – independentemente do formato, meio de apresentação ou canal de veiculação, seja em texto, áudio, vídeo, notícia ou publicação em rede social – identificado como falso, equivocado, enganoso, impreciso, manipulado, fabricado, fraudulento, ilícito ou odioso. Desse modo, a caracterização de um conteúdo como desinformativo independe da intencionalidade do agente (abarcando tanto a noção de *disinformation* como a de *misinformation*). Também se consideram abarcadas pelo conceito de desinformação as informações fora de contexto, manipuladas, editadas maliciosamente, com falseamento de fonte ou apresentadas de forma sensacionalista, ou, ainda, instrumentalizadas para fins ilegítimos (compreendendo a noção de *malinformation*, ilustrada pelo caso de divulgação maliciosa de incidentes cibernéticos contra sistemas *online* de organismos eleitorais).” (Brasil, 2022a, p. 23-24).

A estrutura do Programa Permanente é muito semelhante à do Programa de 2020, de maneira que destacaremos apenas as distinções entre eles, dando-se enfoque, de acordo com o recorte da presente pesquisa, para as medidas que envolvem a cooperação entre o Poder Público e as plataformas digitais, notadamente no combate à disseminação de desinformação, por meio de

“acordos de cooperação com provedores de aplicação de internet que favoreçam a mobilização dessas plataformas digitais e de seus recursos tecnológicos em prol da disseminação de informação verdadeira sobre o processo eleitoral e do combate a comportamentos inautênticos” (Brasil, 2022a, p. 54).

Para o Programa Permanente foi instituída uma Assessoria Especial de Enfrentamento à Desinformação, responsável por executar as ações, e, dentro do Grupo Gestor, o Comitê Estratégico de Combate à Desinformação e o Grupo de Análise e Monitoramento. Um dos objetivos de tornar o Programa permanente é tornar constante o aprendizado e comunicação da Justiça Eleitoral com os parceiros, capacitação contínua, fortalecimento da atuação em rede e a execução de estratégias de médio e longo prazo (Brasil, 2022a, p. 5-9).

Diferentemente dos seis eixos de atuação do Programa de 2020, o Programa Permanente concentrou as ações em apenas três eixos, para coordenar melhor e tornar mais claras as ações. Os três eixos são informar, capacitar e responder (Brasil, 2022a, p. 27).

O eixo de resposta à disseminação de desinformação foi pautado na velocidade das ações, verificação dos fatos e na colaboração das plataformas nas ações, bem como com enfoque na transparência, e conta com três níveis de atuação: o primeiro é voltado ao monitoramento das redes sociais para identificar, especialmente, a disseminação de desinformação e comportamentos inautênticos, como disparo em massa de mensagens; o segundo é a adoção de medidas céleres para o combate aos conteúdos identificados no primeiro nível, notadamente voltado ao combate às ações coordenadas de disseminação de desinformação; com atuação coordenada entre as plataformas, MPE, Polícia Federal, instituições de checagem, entre outros; e, por fim, a elaboração de estratégias de caráter sistêmico e estrutural para o combate à desinformação, dotado de duas esferas: 1) aperfeiçoamento do ordenamento jurídico e 2) da atuação das plataformas digitais, aprimorando as regras e políticas das plataformas, conferindo maior consistência, efetividade e rapidez na atuação e estimulando a transparência (Brasil, 2022a, p. 28-29). Nesse sentido, foram estabelecidos oito projetos dentro deste eixo, quais sejam:

“(i) coalizão permanente para checagem; (ii) engajamento das plataformas digitais e seus recursos tecnológicos no enfrentamento de redes estruturadas de desinformação e comportamento inautêntico; (iii) canal de denúncia de disparo em massa de conteúdos em parceria com o WhatsApp; (iv) criação de rede de monitoramento de práticas de desinformação contra o processo eleitoral; (v) contenção de desinformação no Telegram; (vi) parceria e interlocução com a Polícia Federal e o Ministério Público Eleitoral; (vii) Comitê Estratégico de Ciberinteligência e (viii) revisão e elaboração de normas que combatam a prática da desinformação no âmbito da Justiça Eleitoral, como forma de prevenção do fenômeno em tela.” (Brasil, 2022a, p. 6).

Tendo em vista a necessidade de tomar medidas rápidas contra de desinformação disseminada pelas mídias digitais, de maneira a evitar o dano causa pela demora na reação institucional, as ações do Programa foram guiadas pelas diretrizes do chamado Plano 10-20-60, que estabeleceu o prazo de dez minutos para que fosse feito o registro de uma ocorrência por meio de sistema próprio, elaborando o chamado apontamento; vinte minutos para análise e classificação da ocorrência, identificando se é caso de desinformação, *phishing*, crime eleitoral, entre outros; e, por fim, sessenta minutos para encaminhamento, seja às plataformas parceiras do programa, à secretaria de comunicação, ou demais encaminhamentos previstos no programa (Brasil, 2022a, p. 30-31).

As ações planejadas para as eleições de 2022 denotam o reconhecimento do TSE quanto a especificidades do funcionamento da comunicação digital. As medidas que envolvem a cooperação das próprias plataformas digitais envolvem soluções para problemas como a criação de filtros-bolha e câmaras-de-eco, buscando minimizar os efeitos dos “operações de influência”

e “vácuos de informação”, por meio, por exemplo, do estabelecimento de uma rede de disseminação de informações verdadeiras que envolve um papel destacado das mídias digitais em priorizar a apresentação destes conteúdos em seus serviços, em detrimento de outros materiais que seriam veiculados sob a lógica do capitalismo de vigilância (Brasil, 2022a, p. 39-40).

Os acordos para atuação conjunta com os provedores de aplicações de *internet* foram aprimorados, em relação àqueles estabelecidos para as eleições de 2020, prevendo a continuidade de iniciativas daquela oportunidade, porém visando o aprimoramento especialmente no que diz respeito à uniformidade e transparência na tomada de decisão das plataformas parceiras.

Se o acordo feito com o *Google* para as eleições de 2020 contava com pouca delimitação das ações que seriam desenvolvidas, com maior enfoque para a formação de servidores do TSE sobre o funcionamento das plataformas geridas pela empresa, o Memorando de Entendimento-TSE nº 1/2022 traz com maior especificidade tanto as medidas de treinamento para servidores do TSE, especificando oito sessões de treinamento sobre serviços da empresa, tais quais o *Google News Initiative*, a *DialogFlow*, *Google Play*, *Google Search*, *YouTube*, canais de denúncia e *Google Ads*, quanto para medidas afirmativas de disseminação de informação confiável, como seleção editorial de aplicativos disponíveis na *Google Play Store* e a publicação de um *Doodle* relativo às eleições, além de prever também a publicação de um conteúdo informativo sobre o próprio funcionamento das plataformas (Brasil, 2022b, p. 2-4).

Já quanto à contenção de desinformação, enquanto o acordo de 2020 apenas reforçava a existência de canais de denúncia já existentes nas plataformas geridas pela empresa, o Memorando de Entendimento-TSE nº 1/2022 prevê a publicação de *trend hub* do serviço *Google Search*, durante o período eleitoral, indicando, através dos dados coletados pela empresa, as principais pesquisas realizadas no mecanismo de busca, além de reforçar os já existentes canais de denúncia, inclusive por meio do *Trusted Flagger*, programa que foi criado pela empresa para fornecer ferramentas para denúncias e *feedback* de decisões para instituições confiáveis, do qual o TSE faz parte desde dezembro de 2020 (Brasil, 2022b, p. 4-5). E, quanto à transparência, a empresa se comprometeu a produzir relatório sobre anúncios políticos no Brasil, bem como a atualizar seus termos de serviço do *Google Ads* e *Display & Video 360* para incluir a identificação obrigatória de anunciantes desta natureza (Brasil, 2022b, p. 5).

No mesmo sentido, o Memorando de Entendimento-TSE nº 03/2022, com o *Facebook*, *Serviços Online do Brasil Ltda* prevê, assim como em 2020, a continuidade da ferramenta megafone na rede social *Facebook* e de *stickers* no *Instagram*. Adicionalmente, previu a criação

de rótulo eleitoral no *Facebook* e *Instagram* que, ao identificar que um conteúdo postado diz respeito às eleições de 2022, o rotulava, de maneira a direcionar os usuários para páginas oficiais sobre o processo eleitoral, inclusive da própria Justiça Eleitoral. Foi ainda criado *chatbot* para perguntas e respostas sobre as eleições no *Instagram*. As ações de capacitação dos servidores da Justiça Eleitoral incluíram seminários sobre o funcionamento das plataformas, produção de cartilhas educativas, *workshops* sobre discurso de ódio e extremismo e incentivo à participação feminina na política (Brasil, 2022c, p. 2-4).

Quanto às ações voltadas ao combate à desinformação, o Memorando de Entendimento-TSE nº 03/2022 possibilita maior capacidade de identificação de conteúdo que viola os termos e serviços das plataformas e que possibilitam uma reação mais célere e, portanto, mais eficaz: o uso do API (interface de programação de aplicativo) da biblioteca de anúncios das plataformas pelo TSE, por meio de uma conta de desenvolvedor no Facebook, com a finalidade de elencar os anúncios sobre política ou eleições; e a instituição de um canal extrajudicial não-vinculativo de comunicação de denúncias de conteúdos que sejam desinformação e que, potencialmente, descumpram os termos da plataforma (Brasil, 2022c, p. 4-5).

3.6.4. Moderação de conteúdo realizada pelo Facebook

O *Facebook* realiza, conforme exposto anteriormente, moderação de conteúdo nas redes sociais geridas pela empresa, seja por iniciativa própria, por meio dos Padrões de Comunidade, em relação ao *Facebook*, ou pela Diretrizes de Comunidade, no *Instagram*, ou em função de requerimento de terceiros, sempre com base nas Políticas Corporativas de Direitos Humanos da empresa e em estrutura multiparticipativa, ou seja, em cooperação com organizações da sociedade civil, governos, acadêmicos, entidades internacionais e as normas especiais da Organização das Nações Unidas (Facebook, 2023a).

Nas políticas voltadas ao combate à desinformação nas plataformas do *Facebook*, a empresa afirma que, em razão da dificuldade de estabelecer um conceito objetivo ou um rol de conteúdos que possam ser considerados desinformação, em função da ampla gama de situações limítrofes do que poderia ou não ser considerado falso, optou por uma abordagem voltada principalmente para conteúdos com desinformação em recortes específicos, quais sejam: quando há risco de violência derivado da disseminação daquele conteúdo; quando o conteúdo pode afetar processos políticos; no caso de mídias altamente enganosas. Quanto a temas relacionados à disseminação de desinformação e que podem ser mais bem categorizados, no

caso de contas falsas, fraude e comportamento inautêntico coordenado, a empresa é dotada de políticas específicas mais bem delimitadas. Para cumprir com o objetivo, a plataforma enfoca em combater a disseminação viral de conteúdo com desinformação, bem como realiza acordos com agências verificadoras de fatos em relação aos conteúdos que tem maior alcance (Facebook, 2023b).

Quanto às categorias de conteúdo com desinformação sobre os quais a plataforma atua, incluem-se: 1) possibilidade de agressão física ou violência, em que são excluídos desinformação ou boatos não verificáveis- ambos avaliados por organizações parceiras que, no primeiro caso, verifica que o conteúdo é falso e, no segundo, uma afirmação cuja fonte é muito difícil ou impossível ser verificada, que carece de fontes oficiais, quando ela não é específica o suficiente para ser considerada falsa ou quando é muito absurda para que possa ser considerada real- e que possam resultar em ameaça de violência e, em última instância, risco de morte; 2) saúde, que se subdivide em três categorias: desinformação que possa contribuir para a rejeição de vacinas; desinformação sobre saúde pública em meio a uma emergência de saúde pública, criada em razão da pandemia de Covid-19; e promover ou defender supostas curas que possam ser prejudiciais à saúde. Nesse caso, o conteúdo é avaliado de acordo com as autoridades públicas de saúde; 3) desinformação que possa afetar processos políticos, especialmente a ocorrência de eleições e censos, em especial aquelas que possam prejudicar a participação em processos eleitorais; 4) mídia manipulada, que envolve conteúdos editados de maneira a parecer que o indivíduo que aparece no vídeo falou algo que não corresponde à realidade ou nos casos de produtos de inteligência artificial, por exemplo o chamado *deepfake*, que possa gerar desinformação. Nesse caso, o Facebook opta por atuar preventivamente em razão de especialistas identificarem que tais conteúdos podem viralizar rapidamente e que dificilmente podem ser corrigidos pela disponibilização de conteúdos informativos (Facebook, 2023b).

A empresa realiza análise do conteúdo por meio de ferramentas tecnológicas que identificam automaticamente o conteúdo violador, por meio de aprendizado de máquina e inteligência artificial, principalmente em relação a conteúdos que são fácil e objetivamente identificáveis, como no caso de nudez, por exemplo, sendo que mais de 90% dos conteúdos violadores das políticas das plataformas geridas pela empresa são identificados antes de qualquer denúncia, por meio de análise automatizada. Já em casos que dependem de contextualização ou que contém mais nuances, é realizada a verificação humana, seja por parte de milhares de analistas contratados para verificar o cumprimento aos termos das plataformas geridas pela empresa, seja em casos mais complexos e recorrentes com atuação das equipes de Operações Globais e Política de Conteúdo do próprio Facebook, da verificação de fatos de

agências de verificação ou, ainda, do Comitê de Supervisão, que gera recomendações sobre os conteúdos (Facebook, 2023b).

A análise automatizada é elaborada através de modelos de aprendizado de máquina e previsões, de forma que são criadas ferramentas capazes de identificar o conteúdo de uma foto ou um texto, por exemplo, e indicar o descumprimento das políticas das plataformas e, após tal análise, um outro sistema de inteligência artificial identifica se deve ser tomada alguma medida sobre o conteúdo, como a remoção; rebaixamento, ou seja, a diminuição do alcance da publicação; ou ainda se o conteúdo deve ser encaminhado para análise humana. A tecnologia é atualizada ao ser confrontada com a posterior análise humana, em cima dos conteúdos postados todos os dias na plataforma, que aperfeiçoa, pela repetição, através de milhares de decisões revisadas por humanos, a capacidade de identificação. Ademais, a inteligência artificial é altamente eficaz para enfrentar conteúdos recorrentes, ou seja, é ideal para combater campanhas de desinformação, memes e conteúdos virais que possam ser veiculados rapidamente. Os conteúdos recorrentes, entretanto, podem conter pequenas alterações que dificultam a avaliação automatizada, de forma que a empresa tem atuado, nos últimos anos, no aperfeiçoamento das tecnologias de identificação de conteúdo (Facebook, 2023b).

A empresa utiliza ferramentas tecnológicas que atuam na identificação, classificação e tomada de medidas em face a conteúdos violadores de suas políticas nas plataformas, entre eles: o *Linformers*, que analisa conteúdos no *Facebook* e *Instagram*, tanto de imagem quanto de textos, que identifica conteúdos violadores, especialmente discurso de ódio, violência e *bullying*, cujo código foi disponibilizado para experts em inteligência artificial externos, para compreensão e sugestões de aprimoramento; o *Reinforced Integrity Optimizer*, ou Otimizador de Integridade Reforçada, que realiza aprendizado de máquina para melhor identificar discurso de ódio; o *SimSearchNet*, que é uma ferramenta que identifica nuances e diferenças sutis entre imagens para tomada de medidas contra a desinformação; e as ferramentas XLR e XLR-M, que através do aprendizado de vários idiomas torna mais facilitada a aplicação das políticas da empresa nos mais de 160 idiomas utilizados nas plataformas, cujo código também foi disponibilizado para terceiros (Facebook, 2023b).

Essas inteligências artificiais possibilitam a análise da vasta quantidade de conteúdo publicado diariamente nas redes sociais *Facebook* e *Instagram*, diminuindo o escopo de materiais que serão encaminhados para análise humana. Caso o conteúdo dependa de análise humana, entretanto, este poderá ser encaminhado pela inteligência artificial através de uma análise quanto ao próprio conteúdo, ou ainda em relação a métricas de alcance, como a rápida viralização, ou se usuários estiverem indicando, seja pelos sistemas de notificação das

plataformas, seja por meio de mecanismos das próprias plataformas, como comentários ou reações, descrença sobre aquele conteúdo. Durante o período de análise, o conteúdo pode ter seu alcance limitado pela plataforma (Facebook, 2023b).

O conteúdo é, então, analisado por verificadores independentes, ou seja, não vinculados à empresa, certificados pelo *International Fact-Checking Network* (IFCN), podendo ser classificado como: falso, que são conteúdos que não se baseiam em fatos, como citações falsas, teorias da conspiração, conteúdo inventado de sites que pretendem se passar por verdadeiros e elaborados por agências de notícia, conteúdo que é autêntico, mas que busca comprovar evento não relacionado; alterado, que são conteúdos que foram editados ou sintetizados de formas que possam enganar as pessoas, mas que não incluem trechos de mídia ou mídia fora do contexto, e conteúdos criados por inteligência artificial ou aprendizado de máquina que possam fazer parecer que determinado indivíduo disse algo que não foi dito; parcialmente falso, ou seja, que contém imprecisões sobre fatos, incluindo materiais que contém imprecisões sobre dados, datas e horas, que possam ser considerados uma interpretação dos fatos, mas que estão equivocados, ou uma mistura de informações verdadeiras e falsas em que as falsas não predominam ou, ainda, conteúdo apresentado como opinião, mas que contém distorções sobre fatos; sem contexto, ou seja, o uso de um conteúdo verídico para insinuar uma alegação falsa, sem que seja dita diretamente; sátira, que usa a ironia, exagero ou absurdo para tecer comentários sobre determinado fato, especialmente em relação a conteúdos políticos, religiosos ou sociais; verdadeira, ou seja, o conteúdo que não contém imprecisões ou enganação (Facebook, 2023b).

Os verificadores analisam, prioritariamente, informação falsa viral; boatos sem base em fatos; conteúdo apresentado como opinião, mas que contém informações falsas; alegações falsas que são oportunas, populares e relevantes, e, por outro lado, para proteger o debate público e a liberdade de expressão, não analisam conteúdo sem afirmação verificável, conteúdo verdadeiro no momento da publicação; conteúdo de opinião que não divulgue informações falsas; opiniões e falsas de políticos; afirmações irrelevantes ou que tenham pequenas imprecisões (Facebook, 2023b).

Após a análise dos conteúdos, a empresa pode aplicar, de acordo com a classificação dada pelos verificadores e pelas tecnologias, medidas para: 1) remover; 2) reduzir; e 3) informar. Caso o conteúdo viole frontalmente as políticas de uso das plataformas, o *Facebook* o excluirá, informando o usuário que postou, e identificando, se possível, qual a política específica descumprida. É possível recorrer da decisão da plataforma, de forma que o conteúdo passará por nova avaliação, tanto no caso de remoção do conteúdo, como em outras medidas como a rotulação (Facebook, 2023b). Desde 2020, ainda, é possível que um recurso quanto à

violação das políticas das plataformas possa ser apresentado junto ao Comitê de Supervisão, que lida com casos de ampla repercussão e recorrentes, para análise pormenorizada (Facebook, 2023c).

A empresa trabalha, ainda, com um sistema de contabilização de advertências, em função da quantidade e gravidade das violações postadas por um perfil, página ou grupo, com validade de um ano, que pode gerar, seja pela recorrência ou gravidade, exclusão ou suspensão do usuário da plataforma. As restrições de conta, página ou grupos são aplicadas somente na plataforma *Facebook*, e não no *Instagram*, e podem implicar em limitações como restrições em recursos específicos da plataforma, como publicar em grupos, criar conteúdo e comentar em *posts*, páginas e grupos podem ter seu alcance reduzido. A conta, página ou grupo pode, em ambas as redes sociais, ser desativada, em razão da reincidência ou gravidade da violação, como no caso de contas falsas ou publicação de exploração sexual infantil (Facebook, 2023b).

A plataforma conta com política específica para figuras públicas, definidas enquanto representantes políticos a nível nacional ou estadual, candidatos a cargos políticos, pessoas com mais de um milhão de seguidores e pessoas que recebem ampla cobertura de notícias, que incitem atos de violência ou agitações civis, que podem resultar em medidas mais imediatas e enérgicas que as políticas padrão da empresa, de acordo com o histórico da figura público e da gravidade da violação; o potencial de influência da figura pública; e a gravidade dos atos de violência resultantes daquele conteúdo. Podem ser aplicadas ainda medidas mais gravosas após o término de medidas anteriores em relação à reincidência nesses casos (Facebook, 2023b).

O *Facebook* publica relatórios de transparência sobre os conteúdos que são objeto de restrição em seus serviços *Instagram* e *Facebook*, tanto em relação ao descumprimento de termos de uso da própria plataforma, por meio do *Community Standards Enforcement Report*, ou Relatório de Aplicação de Padrões da Comunidade, quanto em razão de solicitação de terceiros em função de cumprimento da legislação local (Facebook, 2023b).

Os dados sobre a aplicação dos padrões de comunidade são compilados e publicados em relatório quadrimestral, subdivididos em 4 grupos e 11 categorias: 1) violência e comportamento criminoso: a) organizações perigosas, como grupos terroristas ou grupos organizados de ódio; b) produtos e serviços restritos ou proibidos, como drogas ou armas de fogo; c) violência ou incitação à violência; 2) segurança: d) suicídio ou automutilação; e) nudez, abuso físico ou exploração sexual de crianças; f) bullying ou assédio; 3) conteúdo questionável: g) discurso de ódio; h) conteúdo que contém violência; i) nudez adulta e atividade sexual; 4) integridade e autenticidade: j) contas falsas; k) *spam* (Facebook, 2023b).

A empresa destaca que as políticas e classificação de conteúdo foram muito alteradas ao longo do tempo, demonstrando que a abordagem tanto em relação às medidas tomadas, quanto aos relatórios de transparência, tem sido afetada pelos avanços nas discussões e no reconhecimento dos impactos decorrentes da veiculação deste tipo de conteúdo nas mídias digitais. Conforme exposto anteriormente, e em consonância com as políticas da empresa que não buscam classificar exatamente o que constitui desinformação, o relatório não trata este tipo de conteúdo como uma categoria, o que dificulta a análise dos dados sobre as medidas tomadas especificamente contra a desinformação e sua evolução ao longo do tempo, estando os casos distribuídos entre as demais categorias. Ademais, diferentemente dos dados sobre a restrição de conteúdos em razão do cumprimento de legislação local, os dados referentes às políticas de comunidade tanto do *Facebook* quanto do *Instagram* são globais, não havendo dados específicos quanto a aplicação no Brasil (Facebook, 2023b).

Nesse sentido, a regulamentação das plataformas digitais é fundamental para a transparência das medidas tomadas pela empresa, mas também para que seja possibilitado a agentes externos, inclusive pesquisadores, analisar as políticas aplicadas pela empresa que resultam em moderação de conteúdo e, portanto, em governança privada dos discursos. Há, recentemente, a elaboração de diversas normas que obrigam as plataformas à elaboração deste tipo de relatório, que são elaborados a partir de normas e padrões mais específicos, voltados ao enfrentamento de questões em que a plataforma não atua adequadamente. Assim, além dos relatórios sobre a aplicação das políticas de comunidade e do cumprimento de legislação local, a empresa publica relatórios de acordo com normas nacionais ou regionais específicas, quais sejam: *Digital Services Act*, da União Europeia, em relação ao número mensal de usuários ativos, em cumprimento ao art. 24 da referida norma; *KoPI-G*, da Áustria, *EU Child Sexual Abuse Material*, no combate ao abuso sexual infantil na Europa; *India Monthly Report under the Information Technology (Intermediary Guidelines and Digital Ethics Code) Rules*, que estabelece um código de funcionamento e ética para as plataformas; *NetzDG*, lei alemã referida anteriormente, que responsabiliza as plataformas; *Turkey Transparency Report for User Applications under Law no. 5651*, quanto à remoção de conteúdos que afetam a personalidade e privacidade; *EU Terrorist Content Online*, sobre atividade terrorista na Europa (Facebook, 2023b).

Quanto aos dados sobre moderação de conteúdo de acordo com as políticas das plataformas, identifica-se que as categorias de violência e incitação à violência, contas falsas e *spam* guardam relação direta com a desinformação, de acordo com os critérios definidos pela própria plataforma, enquanto há uma categoria específica para o discurso de ódio. Os dados

globais totalizados publicados pela empresa nestas categorias podem lançar luz na maneira como as políticas são efetivamente aplicadas e, principalmente, sobre o volume de conteúdo avaliado e sobre o qual a empresa toma medidas restritivas (entre elas a remoção, avisos quanto ao conteúdo ou exclusão de contas). Os dados sobre violência e incitação à violência passaram a ser publicados separadamente desde o terceiro trimestre de 2021, portanto no período de julho a setembro, enquanto os dados sobre discurso de ódio, contas falsas e *spam* são publicados desde o último trimestre de 2017, de outubro a dezembro (Facebook, 2023b).

A forma como a empresa contabiliza as ações quanto aos conteúdos violadores de suas políticas modificou-se ao longo do tempo, tendo sido a métrica utilizada atualmente estabelecida em 2018. De acordo com essa métrica, cada foto, texto ou *post* no *Facebook* corresponde a um conteúdo, enquanto no *Instagram* um *post* com mais de uma foto e texto conta como apenas um conteúdo. Ou seja, um *post* com quatro fotos no *Facebook* corresponde a cinco conteúdos, enquanto o mesmo material no *Instagram* corresponde a apenas um conteúdo. São contabilizados apenas conteúdos que efetivamente foram veiculados nas redes sociais, de forma que conteúdos bloqueados antes de serem publicados não são incluídos nos dados. Os conteúdos passaram a ser categorizados em 2017, para aprimorar a transparência sobre os conteúdos objeto de restrições e, quando relacionados a mais de uma categoria, são contabilizados apenas naquela mais relevante em relação ao próprio conteúdo (Facebook, 2023b).

Os dados destas três categorias demonstram que, de fato, o *Facebook* exerce governança privada do discurso, de acordo com os termos de uso e políticas de suas plataformas, atuando para remover ou utilizar outros meios de restrição sobre conteúdo postado pelos usuários. O número de conteúdos afetados é muito alto, atingindo o pico de 2,2 bilhões de contas falsas identificadas no *Facebook* em apenas um trimestre, de janeiro a março de 2019; 2,9 bilhões de conteúdo com *spam* entre outubro e novembro de 2019; 21,7 milhões de conteúdos com violência ou incitação à violência no *Facebook* entre janeiro e março de 2022; e 31,5 milhões com discurso de ódio no *Facebook* entre abril e junho de 2021. Tendo em vista o alto volume de conteúdo que é postado diariamente nas redes sociais geridas pela empresa, apesar da grande quantidade de conteúdo violador identificado, este representa apenas uma pequena parcela do que é veiculado nas plataformas. Quanto às contas falsas, a empresa estima que cerca de 4 a 5% das contas mensalmente ativas no *Facebook* no segundo quadrimestre de 2023 eram contas falsas, enquanto os conteúdos com violência ou incitação à violência e discurso de ódio representam, no pico de todo o período disponibilizado, de 0,04% a 0,05% entre junho e setembro de 2021 e 0,1% a 0,11% entre junho e setembro de 2020, respectivamente. Quanto ao

spam, a empresa afirma que não consegue mensurar a quantidade de conteúdo veiculado neste momento (Facebook, 2023b).

Outra informação relevante derivada dos dados é a de que determinados conteúdos são mais fáceis de serem objetivamente identificados, de maneira que a atuação de tecnologias de detecção é mais facilmente aplicável, tornando viável a atuação proativa da empresa. No caso das contas falsas, em que se trata de comportamento inautêntico, a empresa prioriza o combate a contas que pretendem gerar danos, criadas para disseminar conteúdo em campanhas de *spam* e que são, geralmente, financiadas. As contas são identificadas, em sua maioria, de maneira proativa, por meio de uso de tecnologia, geralmente em apenas alguns minutos após a criação. Ao longo de todo o período em que os dados são publicados, as contas falsas no *Facebook* sobre as quais a empresa tomou medidas proativamente representam cerca de 99% de todas as contas identificadas, sendo o outro 1% reportado por usuários, dado semelhante ao *spam* (Facebook, 2023b).

Quanto aos conteúdos com violência ou incitação à violência e discurso de ódio, a notificação dos usuários representa uma parcela mais significativa da atuação da plataforma. Quanto ao conteúdo com violência ou incitação à violência, entre julho de 2021 e setembro de 2022 a empresa identificava proativamente mais de 94% dos conteúdos violadores, atingindo o pico de 98,2% no trimestre de abril a junho de 2022. Desde então, entretanto, a participação dos usuários foi ampliada, representando 14,9% dos conteúdos violadores identificados no período de abril a junho de 2023. É importante destacar, ainda, que o aumento do percentual de participação dos usuários na identificação de conteúdo coincide com a diminuição do número de ações tomadas em face a conteúdos com violência ou incitação à violência, que no mesmo período reduziu de 19,3 milhões para 10,6 milhões, uma redução de cerca de 45% (Facebook, 2023b).

Em relação aos conteúdos com discurso de ódio, o percentual identificado proativamente pela empresa variou mais ao longo da série histórica, representando menos de um quarto dos conteúdos entre outubro e dezembro de 2017 (23,6%, em face a 76,4% que foram notificados por usuários), quando se inicia o relatório, com aumento paulatino até o período de abril a junho de 2020, quando atingiu um platô entre 94% e 97,6% até abril e junho de 2022, voltando a registrar números inferiores a 90% após este período. Trata-se, portanto, de categoria em que a denúncia dos usuários representa parcela maior da atuação da plataforma (Facebook, 2023b).

Já os dados de restrição de conteúdos em razão de cumprimento de legislação local, ou seja, mediante denúncia e solicitação do governo, reguladores, tribunais, entidades não

governamentais ou dos próprios usuários, são compilados em semestres e categorizados por país e natureza da publicação que foi objeto de restrição, ou seja, se é um *post*, uma página ou grupo, um comentário, um perfil, um evento ou um álbum. Portanto, além de a empresa publicar relatório sobre o descumprimento dos termos de serviço, ainda é produzido relatório específico sobre a remoção de conteúdo que se dá em cumprimento a legislação local (Facebook, 2023b).

Quando a empresa recebe uma denúncia em relação a algum conteúdo, por qualquer via disponível, estabelece um procedimento de moderação que identificará se o material deve ser removido. Em razão do alto volume de conteúdo veiculado e da dificuldade em averiguar caso-a-caso, bem como das nuances quanto à avaliação sobre a necessidade de exclusão de um conteúdo, o Facebook identificou a necessidade de instituir um órgão que fosse capaz de, através da estrutura multiparticipativa e da análise quanto aos casos recorrentes, elaborar diretrizes e decisões que pudessem orientar a moderação de conteúdo. Ao longo de 2019 foi desenvolvido o projeto do Comitê de Supervisão, instituído definitivamente em 2020, que elabora, por meio de um processo de apelação quanto às decisões sobre moderação de conteúdo, decisões sobre casos recorrentes e recomendações para as políticas da plataforma. O Comitê segue as recomendações do *Guiding Principles on Business and Human Rights*, ou Princípios Orientadores sobre Empresas e Direitos Humanos, da ONU, demonstrando a aproximação da empresa com organizações internacionais nas medidas de governança (Facebook, 2023b).

O procedimento para avaliação sobre o conteúdo notificado, seja por meio de tribunal, pedido do governo, reguladores, entidades não governamentais ou dos usuários é semelhante. Inicialmente o *Facebook* realiza uma análise se o conteúdo reportado contraria os padrões de comunidade, no caso da plataforma *Facebook*, ou as diretrizes de comunidade, se na plataforma *Instagram*. Caso o conteúdo esteja em desacordo com as normas das plataformas, este é globalmente excluído. Entretanto, caso este não contrarie os padrões da empresa, é feita uma análise legal por equipe jurídica que investiga se o pedido de exclusão é procedimentalmente válido e se o conteúdo descumprir a norma local. Caso a equipe jurídica identifique que o procedimento de requisição é irregular ou que não há descumprimento de norma local, a notificação é rejeitada (Facebook, 2023b).

Trata-se, portanto, de etapa que pode rejeitar a denúncia e arquivar o caso reportado, entretanto, não gera, por si só, a tomada de medida de restrição pela empresa. Para tanto, uma vez que a denúncia ultrapasse a análise jurídica, será encaminhada para análise de *due diligence* em direitos humanos, nos termos do compromisso assumido pelo *Facebook* de acordo com as diretrizes do *Global Network Initiative*, instituição multissetorial que reúne especialistas no tema, e a *Corporate Human Rights Policy* do *Facebook*. O conteúdo passa por uma análise de

acordo com diretrizes internacionais de direitos humanos, avaliando-se cinco critérios: legalidade, legitimidade, necessidade, proporcionalidade e riscos externos. Ou seja, ainda que o procedimento de notificação tenha sido válido e o conteúdo viole norma nacional, este somente será excluído após avaliação mediante princípios internacionais de direitos humanos e, ainda, caso seja identificado que o conteúdo é irregular, este é indisponibilizado apenas na localidade em que é considerado ilegal (Facebook, 2023b).

A *Corporate Human Rights Policy*, ou Política Corporativa de Direitos Humanos, em tradução literal, em uma espécie de preâmbulo do documento, informa os princípios que norteiam a moderação de conteúdo exercida pela empresa: dar voz às pessoas; servir a todos; promover oportunidades econômicas; construir conexões e uma comunidade; manter as pessoas à salvo e garantir a privacidade. O documento, segundo a empresa, é pautado no *Guiding Principles on Business and Human Rights*, ou Princípios Orientadores sobre Empresas e Direitos Humanos, da ONU, bem como em diversos documentos internacionais de direitos humanos, como a Declaração Universal dos Direitos do Homem, da Convenção Internacional em Direitos Civis e Políticos; a Convenção Internacional em Direitos Econômicos, Sociais e Culturais; a Declaração de Princípios Fundamentais e Direitos do Trabalho, da Organização Internacional do Trabalho e, em casos específicos, na Convenção Internacional para a Eliminação de Todas as Formas de Discriminação Racial; na Convenção Internacional para a Eliminação de Todas as Formas de Discriminação contra a Mulher; na Convenção de Direitos da Criança; na Convenção de Direitos das Pessoas com Deficiência; na Carta de Direitos Fundamentais da União Europeia e na Convenção Americana de Direitos Humanos (Meta, 2023).

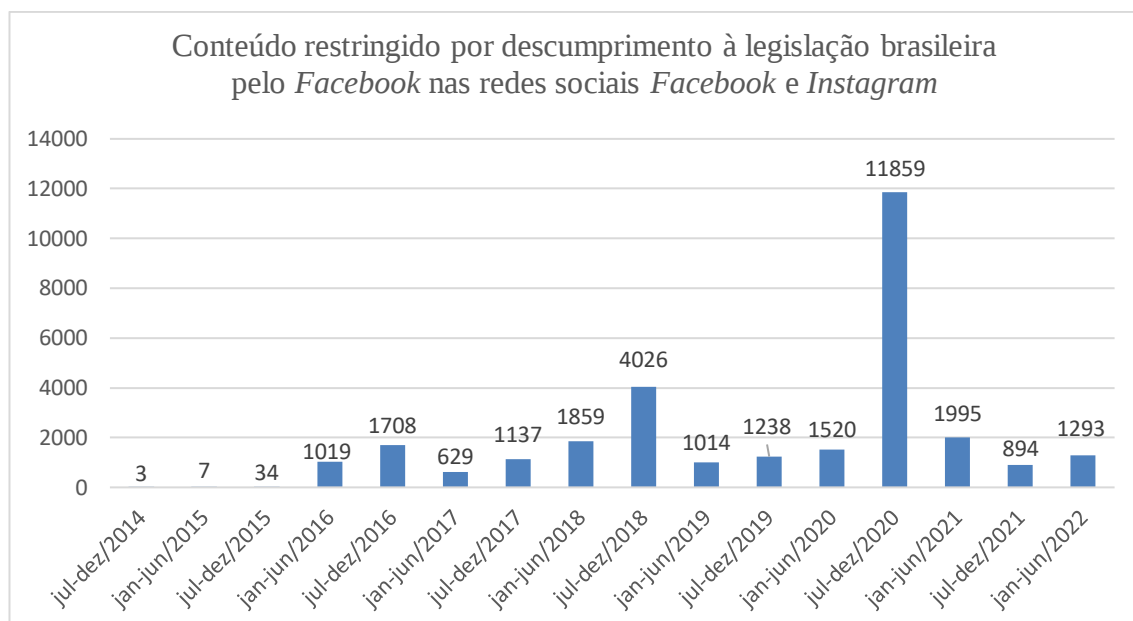
Além disso, o documento afirma que a empresa buscar integrar direitos humanos e tecnologia baseadas em inteligência artificial desde sua construção, teste e desenvolvimento, e que reconhece os Princípios sobre Inteligência Artificial da OCDE, que destaca que as inteligências artificiais devem respeitar o Estado de Direito, direitos humanos e valores democráticos, incluindo a liberdade, dignidade e autonomia, proteção da privacidade e dos dados, a não-discriminação e equidade, a diversidade, justiça, justiça social e os direitos do trabalho reconhecidos internacionalmente (Meta, 2023).

Tanto na fase jurídica, em que se avaliam as normas a nível nacional, quanto na fase de análise de consonância com padrões internacionais de direitos humanos, a empresa estabelece um procedimento regulatório interno que afeta terceiros, notadamente os usuários, e realiza juízo legal sobre os conteúdos veiculados na plataforma. Quanto aos critérios de direitos humanos analisados pela empresa, avalia-se, quanto a: 1) legalidade: se a norma nacional está

em consonância com os padrões de direitos humanos, respeitando do Estado de Direito, conferindo segurança jurídica e protegendo os usuários de arbitrariedades, inclusive quanto a normas procedimentais; 2) legitimidade: se a notificação respeita os direitos coletivos, a proteção da segurança nacional, a ordem pública, a saúde pública ou a moral; 3) necessidade: este critério está relacionado com a garantia da liberdade de expressão, em que o abuso deve ser considerado uma exceção, de forma que apenas um restrito rol de conteúdos pode justificar a tomada de medidas, protegendo-se o debate público e político em temas de interesse público. Essa avaliação leva em consideração, ainda, critérios quantitativos de alcance, engajamento, viralização e idade do conteúdo; 4) proporcionalidade: qualquer medida que seja tomada deve levar em consideração o efeito esperado e ser implementada de maneira mais restrita possível; 5) riscos externos: esse critério leva em consideração situações como risco de bloqueio das plataformas, procedimentos criminais ou prisões de funcionários da empresa ou usuários, segurança e eventuais riscos de dano (Facebook, 2023b).

Após a análise e em qualquer dos casos em que seja tomada alguma ação por parte da empresa, tanto o denunciante quanto o usuário que tenha publicado o conteúdo é notificado da decisão tomada e, ainda, a menos que haja ordem legal explícita proibindo a empresa de fazê-lo, o conteúdo é substituído, nas plataformas, por uma nota que informa a exclusão em razão de cumprimento de legislação local (Facebook, 2023b).

A série histórica de restrição de conteúdos em razão de cumprimento da legislação local sobre o Brasil se inicia no segundo semestre de 2014, período que coincide com o recorte temporal da presente pesquisa, que se inicia com a entrada em vigor do Marco Civil da Internet, e termina, até o momento, no primeiro semestre de 2022. Os dados demonstram que, inicialmente, havia poucos casos de restrição de conteúdo pelo descumprimento à lei brasileira, sendo reportados apenas 3 casos no segundo semestre de 2014. Todavia, o número de casos foi se ampliando ao longo do tempo, com destaque para os segundos semestres dos anos de 2018 e 2020, período eleitoral no Brasil e início da atuação proativa do TSE no combate à desinformação, por meio do Conselho Consultivo sobre Internet e Eleições em 2018 e do Programa de Enfrentamento em 2020, em que foram restringidos em torno de 4 mil e 11,8 mil conteúdos respectivamente (Facebook, 2023b).



Fonte: Facebook, 2023b.

Os casos de restrição de conteúdo em função da legislação local no Brasil entre o segundo semestre de 2014 até o primeiro semestre de 2016 no *Facebook* e *Instagram*, são especialmente derivados de decisões judiciais, cíveis ou criminais, que determinaram a exclusão de conteúdo em razão de difamação, segundo relatório da empresa. A partir do segundo semestre de 2016, que também é período eleitoral, apesar de ainda não existir estratégia específica da justiça eleitoral brasileira, além da indicação do cumprimento de decisões judiciais de natureza civil e criminal, há a indicação de cumprimento de decisões da justiça eleitoral, o que se mantém por todo o período. O que se verifica sobre os dados de restrição de conteúdo durante todo o período é que, desde 2016, afora os períodos eleitorais descritos acima, há certa regularidade na quantidade de casos, orbitando em torno de mil casos por semestre, com média total para o período, excluídos os dois semestres eleitorais de 2018 e 2020, de 1.024 casos, sendo o pico no primeiro semestre de 2021 com quase 2 mil casos e a mínima no primeiro semestre de 2017, com apenas 629 casos. Nesse sentido, pode-se afirmar, em consonância com os relatórios de ação do TSE para as eleições de 2018 e 2020, que os Programas de Enfrentamento à Desinformação aumentaram em aproximadamente 4 e 12 vezes, respectivamente, a quantidade de conteúdos que foram restringidos no *Facebook* e no *Instagram* ao longo da série histórica (Facebook, 2023b).

O *Facebook* destaca ainda, que entre o segundo semestre de 2020 e o primeiro semestre de 2022, cumpriu ordem de restrição global, ou seja, para além das fronteiras nacionais, de 50 páginas e perfis em função de decisões do Supremo Tribunal Federal, anotando, ainda, que

recorreu judicialmente de tais decisões. Além disso, 23 itens foram objeto de restrição no segundo semestre de 2020 a pedido do Ministério da Agricultura e 145 no primeiro semestre de 2022 a pedido da Agência Nacional de Vigilância Sanitária-ANVISA (Facebook, 2023b).

Apesar de os dados de moderação de conteúdo pelo descumprimento dos termos e políticas das plataformas não serem publicados em relação aos conteúdos de determinado país, é possível averiguar, em comparação com os dados globais de restrição de conteúdo em função do cumprimento de legislação local, que a maior parte dos conteúdos que são objeto de ação por parte das plataformas é objeto ou de ação proativa da empresa, que identifica e age sobre aquele conteúdo unilateralmente, ou de notificação dos usuários, sobre a qual a própria empresa decide unilateralmente. Os casos de pedidos de restrição de conteúdo em função de solicitação de terceiros para cumprimento de legislação local, portanto, representam uma parcela ínfima de todo o conteúdo moderado nas plataformas geridas pelo *Facebook* (Facebook, 2023b).

A título de exemplo, o pico global de conteúdo restrito em cumprimento à legislação local no *Facebook* de toda a série histórica é de pouco menos de 70 mil conteúdos entre janeiro e junho de 2022, enquanto para o mesmo período, em cumprimento aos termos e políticas da plataforma, a empresa atuou sobre mais de 2,5 bilhões de conteúdos identificados como *spam*, 3 bilhões de contas falsas, 41 milhões de conteúdos com violência ou incitação à violência e 28,6 milhões que continham discurso de ódio. Ou seja, apenas em relação a essas quatro categorias, desconsiderando as demais, a quantidade de conteúdo restrito em razão de cumprimento à legislação local representa apenas a 0,001% do conteúdo moderado pela plataforma (Facebook, 2023b).

É necessário ter parcimônia ao tirar conclusões sobre esses dados, uma vez que se trata de dados relativamente pouco transparentes, pouco detalhados e produzidos de maneira unilateral pela provedora de aplicações de *internet*, no caso duas redes sociais com ampla popularidade no Brasil. Entretanto é possível, em consonância com a revisão bibliográfica realizada nesta dissertação, chegar a algumas conclusões: 1) que as plataformas exercem, de fato, governança privada dos discursos, ao avaliar e restringir conteúdos publicados por terceiros; e 2) que o vasto volume de conteúdo postado diariamente nas plataformas digitais é incompatível com os mecanismos tradicionais de moderação de conteúdo, tornando ineficaz e ineficiente políticas exclusivamente pautadas em decisões judiciais.

3.7. Análise combinada dos tópicos

Da análise da documentação legislativa e jurídica deste capítulo, é possível averiguar que a comunicação digital, por ter como característica ser multiparticipativa, ou seja, com a atuação de diversos atores, especialmente o Estado, Organizações Internacionais, plataformas digitais, organizações da sociedade civil e usuários, se molda a partir de interações e interrelações que, não necessariamente, vão se cristalizar na forma institucional.

De fato, com o desvelamento do caso *Cambridge Analytica*, com a influência, difícil de ser mensurada, sobre as eleições estadunidenses e o *Brexit*, houve uma mudança nos rumos da moderação do conteúdo veiculado pelas mídias digitais, especialmente por parte de atuação das próprias plataformas que oferecem os serviços, chamadas pelo Marco Civil da Internet de provedoras de aplicação *online*, no que pode ser denominado governança privada dos discursos, e da atuação institucional, surgindo mundo afora normas, da esfera nacional e regional, como a *Net-zDG* e o *Digital Services Act*, no âmbito da Alemanha e da União Europeia, respectivamente, que buscam, especialmente por meio da responsabilização das plataformas digitais sobre o conteúdo postado por terceiros, bem como pela instituição de mecanismos de transparência, minimizar os efeitos negativos da ampla disseminação de conteúdos perniciosos, como as *fake news* e o discurso de ódio.

No caso analisado nesta pesquisa, é possível identificar que a conjuntura crítica que deu forma à regulamentação da comunicação digital no Brasil, por meio do Marco Civil da Internet, se deu a partir da descoberta do Caso Snowden, ou seja, da perspectiva de que a participação dos usuários nas mídias digitais pudesse ser objeto de investigação e vigilância por parte do poder público, alinhada à concepção tradicional de violação à privacidade e liberdade de expressão. Nesse sentido, a isenção de responsabilidade das plataformas digitais sobre o conteúdo postado por terceiros, incluída no art. 19 do Marco Civil da Internet, tinha como enfoque o combate a esta forma de vigilância. Até então, tendo em vista a complexidade da comunicação digital, constituída por emaranhados de algoritmos de difícil compreensão e protegidos por segredos empresariais, em uma espécie de caixa-preta, era difícil de identificar uma outra espécie de vigilância, realizada pelas próprias plataformas, que se desvelou posteriormente pelo caso *Cambridge Analytica* e sobre o qual academia, sociedade, instituições públicas e organizações internacionais vem se debruçando ao longo dos últimos anos.

Há um movimento, identificado a partir da revisão bibliográfica e da aplicação do método de mapeamento de arranjo-jurídico institucional, no sentido de ampliar as medidas de moderação de conteúdo nas mídias digitais, notadamente voltada ao comportamento inautêntico e veiculação de conteúdos perniciosos como incentivo à violência e discurso de ódio. No Brasil, esse movimento pode vir a ser institucionalizado por meio da aprovação de nova lei que trate

da moderação de conteúdos, como foi aventado pela chamada Lei das *Fake News*, que chegou a ser votada no Senado Federal, mas não na Câmara dos Deputados, ou por meio da decisão sobre a inconstitucionalidade do artigo 19 do Marco Civil da Internet, em casos julgados pelo Supremo Tribunal Federal.

Até o momento da conclusão desta pesquisa, entretanto, o quadro institucional ainda é regido pela isenção de responsabilidade do Marco Civil da Internet. Não obstante, o Programa de Enfrentamento à Desinformação, instituído pelo Tribunal Superior Eleitoral, e os dados sobre moderação de conteúdo nas plataformas *Facebook* e *Instagram*, permitem inferir que este movimento de ampliação da governança dos discursos nas plataformas digitais já vem ocorrendo no Brasil ao longo dos últimos anos, em consonância com uma tendência internacional. O atual quadro institucional centralizaria toda a moderação de conteúdos nas decisões judiciais caso-a-caso, excetuados os casos de direitos autorais e pornografia de vingança, o que se demonstra totalmente inviável frente à ampla veiculação de conteúdos pelas mídias digitais, de forma que medidas alternativas que, apesar de não serem exigidas, são permitidas pela legislação brasileira, vem sendo adotadas.

Se por um lado a isenção de responsabilidade das plataformas digitais sobre o conteúdo postado por terceiros, juntamente com a ausência de perspectiva dos efeitos negativos ocasionados pela atuação de algoritmos de recomendação na formação de filtros-bolha e câmaras de eco nas mídias digitais, resultou em uma inércia no combate à disseminação de *fake news* e discurso de ódio nas mídias digitais, a responsabilização por si só pode ter como efeito secundário um aumento indevido do poder das plataformas, enquanto agentes privados, gerando uma espécie de censura colateral.

A dificuldade de se estabelecer conceitos fechados, ou mesmo um rol taxativo que possibilite a identificação de *fake news* e discurso de ódio, de maneira que não há sequer consenso acadêmico quanto a estes, demonstra a necessidade de que um arranjo jurídico-institucional que busque enfrentar a disseminação de tais conteúdos conte com a participação dos setores envolvidos na comunicação digital, como aquele que foi instituído pelo Tribunal Superior Eleitoral e pelas próprias plataformas.

Em suma, a responsabilização das plataformas digitais sobre o conteúdo postado por terceiros, que vem sendo adotada internacionalmente, tem o condão de incentivar que as plataformas exerçam governança privada dos discursos, frente à inércia incentivada pela isenção de responsabilidade. Todavia, sem transparência e a possibilidade de auditoria externa às plataformas, as Democracias se fiarão excessivamente à atuação de empresas privadas que, não necessariamente, estão vinculadas ao interesse público, mas sim a condições de mercado.

A atuação multiparticipativa na moderação de conteúdos demonstra-se necessária e sua configuração não deve ser engessada, tendo em vista a necessidade de acompanhar as mudanças que ocorrem rapidamente no setor, devendo ser estabelecidas balizas legais que regulamentarão a relação entre os diversos atores envolvidos no processo de comunicação digital que, hoje, estão demasiadamente centralizadas nas próprias plataformas, de maneira a empoderar especialmente os usuários frente às novas ameaças à liberdade de expressão derivadas do capitalismo de vigilância.

CONCLUSÃO

Conforme exposto na introdução, a pergunta de pesquisa que guiou esta dissertação pode ser sintetizada da seguinte maneira: a regulamentação da comunicação digital, com a responsabilização das plataformas digitais sobre o conteúdo postado por terceiros, pode ser um meio eficaz e democraticamente legítimo para reduzir a disseminação sistemática de conteúdos perniciosos? A partir do desenvolvimento teórico e da análise dos documentos e dados levantados nessa dissertação, é possível elaborar algumas conclusões.

Quanto à regulamentação da comunicação digital, tem-se que, no Brasil, a isenção de responsabilidade das plataformas digitais sobre o conteúdo postado por terceiros, instituída pelo Marco Civil da Internet, foi inspirada por normativas de outros países, notadamente dos Estados Unidos, e na perspectiva tradicional de proteção ao princípio constitucional da liberdade de expressão. Nesse sentido, diversos aspectos da arquitetura das plataformas digitais passaram ao largo da previsão legal no ordenamento jurídico, especialmente o amplo volume de conteúdo que é veiculado diariamente nas mídias digitais e o modelo de negócio das plataformas, baseado no maior engajamento e participação dos usuários, independentemente do conteúdo que está sendo veiculado. Não se trata de afirmar que as plataformas coadunam com *fake news* ou com discurso de ódio e sim, que, por se tratar de empresas privadas, cujo objetivo (legítimo) primário é o lucro, estratégias de enfrentamento sistemático à disseminação de conteúdo prejudicial à institucionalidade democrática são relegadas, quando muito, ao segundo plano.

O questionamento subjacente à própria necessidade de alguma espécie de regulamentação da comunicação digital é, portanto, se existem conteúdos que, por sua natureza, não devem/podem ser veiculados pela comunicação social e, desta maneira, são passíveis de regulação. Nesse ponto, é interessante destacar que o próprio ordenamento jurídico brasileiro institui exceções à isenção de responsabilidade das plataformas sobre o conteúdo postado por terceiros, sendo que a inércia para retirar ou para adotar estratégias alternativas de minimização do alcance de tais conteúdos, podem ser responsabilizadas mediante mera notificação ou, até mesmo, sem notificação – é o caso da violação aos direitos autorais (art. 19, §2º, do Marco Civil da Internet) e da pornografia de vingança (art. 21 do Marco Civil da Internet).

Ao mesmo tempo, as próprias mídias tradicionais – e até mesmo reuniões em praça pública – encontram limitações dentro do texto constitucional, tais como a classificação indicativa etária para programas de televisão (art. 220, §3º, I, Constituição Federal); a proibição de veiculação

de propaganda de produtos nocivos à saúde ou ao meio ambiente (art. 220, §3º, II, Constituição Federal); a informação dos efeitos negativos do tabaco, bebidas alcoólicas, agrotóxicos, medicamentos e terapias (art. 220, §4º, Constituição Federal); ~~bem como regras éticas~~ e a preferência por conteúdos com finalidade educativa, artística, cultural e informativa no rádio e televisão (art. 221, Constituição Federal). Assim, o princípio da liberdade de expressão, como se sabe, não é absoluto, comportando sopesamento em função de outras garantias ou princípios fundamentais e devendo ser interpretado em unidade com a totalidade do sistema jurídico. Por conseguinte, é possível – e não é inteiramente novo – afirmar que existem conteúdos que comportam regulamentação. Contudo, as *fake news* e o discurso de ódio se enquadrariam dentro dessa categoria?

Neste ponto há mais desacordos do que concordâncias entre sociedade civil, usuários, as plataformas e governos. Um dos poucos consensos gira em torno do diagnóstico quanto aos dos efeitos nocivos da disseminação em larga escala desses conteúdos através de um capitalismo de vigilância sobre os dados dos usuários das plataformas digitais. Nesse sentido, O desvelamento de casos de empresas de análise psicométrica que buscam influenciar vontades dos eleitores-consumidores explicitou o funcionamento dos algoritmos das plataformas, com sua geração de câmaras-de-eco e filtros bolha, resultando tanto em repercussões em eleições quanto em atos de violência. A real dimensão dos produtos destas técnicas de marketing digital e da psicopolítica dado-pulsional é difícil de ser mensurada, mas é factível identificar algum impacto negativo da disseminação de *fake news* e discurso de ódio nas democracias contemporâneas e ponderar que este não é insignificante, o que, amparado nos dados analisados nesta pesquisa, justifica, por si só, alguma espécie de regulamentação da comunicação digital.

Daí decorrem outros questionamentos: quais os conceitos de *fake news* e discurso de ódio? É possível delimitá-los?

Ao longo do primeiro capítulo, procuramos desenvolver um conceito de *fake news* e discurso de ódio que levasse em consideração a estrutura e arquitetura do mercado de tecnologia e das mídias digitais, enquanto elementos centrais para a disseminação de tais conteúdos na sociedade digital, buscando apreender as etapas e estratégias da produção industrial de *fake news* e discurso de ódio com a finalidade de influenciar vontades.

É fundamental destacar que esse conceito não pode ser excessivamente engessado, ou demasiadamente restritivo, estabelecendo um rol taxativo de condutas a serem moderadas, uma vez que suas caracterizações e impactos variam conforme a realidade social e cultural de onde

o conteúdo é veiculado. Desta constatação decorre a necessidade de que o sistema de moderação de conteúdo nas mídias digitais seja multiparticipativo e transparente, envolvendo os diversos atores, em um processo de constante aprimoramento. Entretanto, para os fins desta pesquisa, optamos por elaborar balizas mais genéricas que possibilitassem aferir a legitimidade democrática das possibilidades de regulamentação.

Quanto às *fake news*, optamos por entendê-la enquanto ferramenta comunicacional da pós-verdade, veiculada por mídias hiper-partidárias, através de uma análise psicométrica, com o objetivo de gerar engano ou desinformação sobre determinado acontecimento, ao se passar por notícia ou novidade (*news*) e construir uma versão inverídica sobre este, pautando-se, portanto, na estratégia de disputar sentidos sobre a ignorância dos fatos reais. Optamos, assim, por um conceito vinculado à produção e veiculação sistemática destes materiais por meio das mídias digitais, através da atuação profissional de técnicos em *marketing* digital que exercem a chamada psicopolítica dado-pulsional. Através desta caracterização, uma das conclusões parciais da pesquisa é a de que estratégias de regulamentação devem ser especialmente voltadas às técnicas de manipulação do debate público industrialmente realizada, antes do compartilhamento de conteúdo entre os próprios usuários – que constitui uma etapa secundária do processo de disseminação e viralização de *fake news*.

Já quanto ao discurso de ódio, dados os escopos desta dissertação, o recurso a referências mais extensas foi limitado, de maneira que [não] comportou discussões mais aprofundadas a cerca das nuances do tema. Entretanto, discernimos que se trata da comunicação pública – portanto realizada em espaços não-privados, como as mídias digitais – que, conscientemente, degrada simbolicamente grupos histórica ou sistematicamente discriminados, podendo ser direcionada a ‘categorias’ distintas como origem, cor da pele, gênero, religião, orientação sexual, etc. Posto isso, deduzimos que a particularidade da veiculação do discurso de ódio nos meios digitais está na possibilidade – graças à arquitetura das plataformas – de formação de redes de usuários com interesses comuns. Assim, conteúdos que seriam mantidos alheios às mídias tradicionais, isolando indivíduos intolerantes, passam a ser veiculados pelas mídias digitais, a constituir elemento de conexão entre indivíduos pessoas com interesses semelhantes. Por se tratar de uma discriminação sistemática e estrutural, que não comporta conciliação e que está latente no tecido social, o discurso de ódio é especialmente prejudicial à democracia e, por este motivo, a regulamentação que combata à disseminação destes materiais é eivada de legitimidade democrática.

Logo, mesmo que não seja possível estabelecer um rol de conteúdos que se enquadram dentro dos conceitos de *fake news* ou de discurso de ódio, cabe ao direito estabelecer as balizas tanto para a identificação de tais comportamentos, quanto para a instituição de mecanismos de moderação de conteúdos, como sistemas de notificação e exclusão de conteúdo. Uma estratégia seria a instituição legal de procedimentos para que os conteúdos possam ser analisados, estabelecidos de maneira transparente tanto para os usuários, quanto para a sociedade e governos, ou seja, que possam ser objeto de auditoria externa, cujo cumprimento isente as plataformas de responsabilização. De outra forma, a falta de parâmetros para a remoção de mensagens somada a um rol de conteúdos – ainda que exemplificativo – poderia ser limitada e tornar-se, rapidamente, obsoleta, ou a exclusão preventiva e desarrazoada de conteúdos controversos, o que implicaria em censura colateral e, conseqüentemente, afrontaria ao princípio constitucional da liberdade de expressão.

Retomando-se a questão central desta pesquisa: “a regulamentação da comunicação digital, com a responsabilização das plataformas digitais sobre o conteúdo postado por terceiros, pode ser um meio eficaz e democraticamente legítimo para reduzir a disseminação sistemática de conteúdos perniciosos?”, podemos concluir, portanto, que: 1) a regulamentação da comunicação digital se mostra imprescindível para a manutenção da sociedade democrática; 2) a responsabilização das plataformas digitais sobre o conteúdo postado por terceiros é possível, desde que bem estabelecidas as balizas sobre os conteúdos que devem ser moderados, bem como instituídos mecanismos de moderação, com participação de todos os atores envolvidos no processo de comunicação social por meio das mídias digitais e transparente; e 3) a responsabilização das plataformas digitais é meio democraticamente legítimo para reduzir a disseminação sistemática de conteúdos perniciosos como *fake news* e discurso de ódio nas mídias digitais, tendo em vista que a capacidade de vigilância das plataformas digitais e de especialistas em *marketing* digital que exercem o chamado *microtargeting* implicam em uma nova forma de vigilância privada que afronta a liberdade de expressão, atuando sobre a formação de vontade e a autonomia. A eficácia de uma regulamentação está vinculada ao estabelecimento de estratégias sistemáticas que busquem atuar sobre a cadeia de produção e disseminação industrial de conteúdos que contém discurso de ódio ou *fake news* e que busquem influenciar o debate público.

No que concerne à hipótese de que, no Brasil, a isenção de responsabilidade das plataformas digitais sobre o conteúdo postado por terceiros instituída por força do artigo 19 do Marco Civil da Internet – cujo objetivo declarado foi a proteção da liberdade de expressão –, teve enquanto

efeito não esperado ou não pretendido a ampla disseminação de conteúdos perniciosos concluímos que a inclusão desta previsão normativa derivou do não conhecimento, à época, do sistema de vigilância privada instituído por meio da comunicação digital, o que somente se tornou notório a partir da divulgação dos escândalos da Cambridge Analytica. Consequentemente, tal dispositivo encontra-se ultrapassado (ainda que seja bastante recente) e vai na contramão de regulamentos aprovados mundo afora, como o *Digital Services Act* na União Europeia e o *Net-zDG* na Alemanha, e da própria atuação das plataformas.

Além disso, as principais medidas institucionais tomadas para enfrentar a disseminação de *fake news* e discurso de ódio partiram do Tribunal Superior Eleitoral, especialmente durante os períodos eleitorais de 2018, 2020 e 2022, que atuou em conjunto com as plataformas para criar sistemas de notificação e moderação de conteúdos nas mídias digitais. Entretanto um sistema de moderação de conteúdo pautado em decisões judiciais é incondizente tanto com a velocidade quanto com o volume de conteúdos veiculados pelas mídias digitais diariamente. Ainda, mecanismos que foram criados tendo em vista a comunicação de massa tradicional, como a previsão constitucional da vedação do anonimato, ou o dever de “identificabilidade” para ulterior responsabilização, tornam-se obsoletos mediante as novas tecnologias de comunicação.

As alterações recentes na legislação internacional e na postura das plataformas, entretanto, tem resultado em novas políticas no combate à disseminação de *fake news* e discurso de ódio, em um processo que passa por ações institucionais propriamente ditas, como os Programas do Tribunal Superior Eleitoral estudados nessa dissertação, ou na discussão do chamado Projeto de Lei das *Fake News*, e, principalmente, na ação das plataformas ao exercer governança privada dos discursos, modificando o panorama instituído pelo Marco Civil da Internet.

Este é o quadro do arranjo jurídico-institucional no momento de encerramento desta dissertação, entretanto a dinamicidade do tema e as discussões que estão ocorrendo em diversas frentes, no Brasil especialmente pelo Congresso Nacional, na discussão do PL das *Fake News* e no Supremo Tribunal Federal, por meio do julgamento do Recurso Extraordinário nº 1037396/SP, Tema de Repercussão Geral 987, e do Recurso Extraordinário nº 1057258/RJ, Tema de Repercussão Geral 533, sobre a constitucionalidade do art. 19 do Marco Civil da Internet, são movimentos que podem implicar em um ponto de inflexão à regulamentação da comunicação digital, que pode ser objeto de observação posterior.

REFERÊNCIAS BIBLIOGRÁFICAS

Alves, Marco Antônio Souza; Maciel, Emanuella Ribeiro Halfeld. O Fenômeno das fake news: definição, combate e contexto. *Revista Internet & Sociedade*. Vol 1, Nº 1, fev 2020.

Amato, L. F.; Saba, D. T.; Loschiavo Leme de Barros, M. A. Sociologia Jurídica das Fake News Eleitorais: Uma Observação Sistêmica das Respostas Judiciais e Legislativas em Torno das Eleições Brasileiras de 2018. *Direito Público*, [S. l.], v. 18, n. 99, 2021. DOI: 10.11117/rdp.v18i99.5377. Disponível em: <https://www.portaldeperiodicos.idp.edu.br/direitopublico/article/view/5377>. Acesso em: 22/08/2022.

Amato, Lucas Fucci. Fake news: regulação ou metarregulação? *Revista de Informação Legislativa: RIL*, Brasília, DF, v. 58, n. 230, p. 29-53, abr./jun. 2021. Disponível em: https://www12.senado.leg.br/ril/edicoes/58/230/ril_v58_n230_p29. Acesso em: 23/08/2022

Balkin, Jack. Free speech in the algorithmic society: big data, private governance, and new school speech regulation. *UC Davis Law Review*, v. 51, 2018. Acesso em: 07/03/2023. Disponível em: <https://openyls.law.yale.edu/bitstream/handle/20.500.13051/4699/51_U.C.D._law_Review_1149_2018_.pdf?sequence=2&isAllowed=y>.

Baptista, Rodrigo. Redes sociais influenciam voto de 45% da população, indica pesquisa do DataSenado. *Senado Notícias*: Brasília, 12 dez 2019. Disponível em: <<https://www12.senado.leg.br/noticias/materias/2019/12/12/redes-sociais-influenciam-voto-de-45-da-populacao-indica-pesquisa-do-datasenado>>. Acesso em: 21 jan 2022.

Bucci, Eugênio. *A Superindústria do Imaginário: como o capital transformou o olhar em trabalho e se apropriou de tudo que é visível*. Belo Horizonte: Autêntica, 2021.

Bucci, Maria Paula Dallari; Gaspardo, Murilo. Mapeamento de arranjos jurídico-institucionais: um roteiro metodológico para estudos das relações entre direito e política. *REI - REVISTA ESTUDOS INSTITUCIONAIS*, [S. l.], v. 10, n. 1, p. 1–36, 2024. Disponível em: <https://www.estudosinstitucionais.com/REI/article/view/777>. Acesso em: 31 jan. 2024.

Castells, Manuel. *A Era da Informação: Economia Sociedade e Cultura*. Volume I: *A Sociedade em rede*. 6ª ed. 14ª reimpressão. São Paulo: Editora Paz e Terra Ltda, 2011.

- Castells, Manuel. *Ruptura: A crise da democracia liberal*. 1ª ed. Rio de Janeiro: Zahar, 2018.
- Deibert, Ronald. J. Três duras verdades sobre as redes sociais. *Journal of Democracy em Português*, São Paulo, Volume 8, n. 1, p. 27-50, maio de 2019.
- Dourado, Tatiana Maria Silva Galvão. *Fake News na eleição presidencial de 2018 no Brasil*. Tese (doutorado) - Universidade Federal da Bahia, Faculdade de Comunicação, 2020.
- Empoli, Giuliano da. *Os Engenheiros do Caos*. São Paulo: Vestígio, 2019.
- Gabriel, Iason. Toward a theory of justice for artificial intelligence. In: *Daedalus*, Vol. 151, n. 2, maio/2022. Cambridge: MIT Press, 2022.
- Gasser, Urs; Schultz, Wolfgang. *Governance of Online Intermediaries: observations from a Series of National Case Studies*. Cambridge: Berkman Center Research Publication, 2015. Disponível em: <https://dash.harvard.edu/handle/1/16140636>. Acesso em: 09/11/2021.
- Gillespie, T. A relevância dos algoritmos. V. 6, n. 1, p. 95–121, 2018.
- Han, Byung-Chul. *Infocracia: digitalização e crise da democracia*. Petrópolis: Editora Vozes, 2022.
- Hartmann, Ivar. A new framework for online content moderation. *Computer Law & Security Review*, v. 36, 2020.
- Holland, Adam; Bavitz, Chris; Hermes, Jeff; Sellars, Andy; Budish, Ryan; Lambert, Michael; Decoster, Nick. *NoC Online Intermediaries Case Studies Series: Intermediary Liability in the United States*. In: Gasser, Urs; Schultz, Wolfgang. *Governance of Online Intermediaries: observations from a Series of National Case Studies*. Cambridge: Berkman Center Research Publication, 2015. Disponível em: <https://dash.harvard.edu/handle/1/16140636>. Acesso em: 09/11/2021.
- Horbach, Carlos Bastide. Political Speech, Freedom of Expression and Fake News: The Role of the Brazilian Electoral Justice in the 2018 Presidential Elections. In: Morais, Carlos Blanco de; Mendes, Gilmar Ferreira; Vesting, Thomas (org.). *The Rule of Law in Cyberspace*. Cham: Springer International Publishing, 2022. p. 187-201.
- Lemos, Ronaldo; Steibel, Fabro; Affonso de Souza, Carlos; Nolasco, Juliana. A Bill of Rights for the Brazilian Internet (“Marco Civil”)- A Multistakeholder Policymaking Case. p. 200-212.

In: Gasser, Urs; Budish, Ryan; Myers West, Sarah (eds.). *Multistakeholder as Governance Groups: Observations from Case Studies*. Boston: Berkman Center Research Publication, 2015.

Macedo Júnior, Ronaldo Porto. "Fake news" e as novas ameaças à liberdade de expressão. A liberdade de expressão e as novas mídias. Tradução. São Paulo: Perspectiva, 2020. p. 231-247. (Kindle).

Maranhão, Juliano; Campos, Ricardo. Fake news e autorregulação regulada das redes sociais no Brasil: fundamentos constitucionais. In: Abboud, Georges; Júnior, Nelson; Campos, Ricardo. *Fake News e Regulação*. São Paulo: Editora Revista dos Tribunais, 2019. p. 321-335. (Kindle)

Marichal, José. *Facebook democracy: the architecture of disclosure and the threat to public life*. Nova York: Routledge, 2016.

Morozov, Evgeny. *Big Tech: a ascensão dos dados e a morte da política*. São Paulo: Ubu Editora, 2018.

Oliveira, Filipe. Facebook chega a 127 milhões de usuários mensais no Brasil. Folha de São Paulo: São Paulo, 18 jul. 2018. Disponível em: <<https://www1.folha.uol.com.br/tec/2018/07/facebook-chega-a-127-milhoes-de-usuarios-mensais-no-brasil.shtml>>. Acesso em: 19 jan 2022.

Pasquale, Frank. Toward a Fourth Law of Robotics: Preserving Attribution, Responsibility, and Explainability in an Algorithmic Society. In: *Ohio State Law Journal*, Vol. 78, n. 21, p. 1243-1255, julho/2017a.

Pasquale, Frank. *The Automated Public Sphere*. University of Maryland Legal Studies Research Paper n° 2017-31, 2017b. Disponível em: <chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://digitalcommons.law.umaryland.edu/cgi/viewcontent.cgi?article=2609&context=fac_pubs>. Acesso em: 13/06/2023.

Queiroz, Christina. Terra sem lei: falta de consenso sobre o conceito de desinformação dificulta esforço regulatório para controlar disseminação de conteúdo falso em plataformas digitais. *Revista Fapesp*: São Paulo, 2022. Disponível em: <<https://revistapesquisa.fapesp.br/terra-sem-lei/>>. Acesso em: 22/08/2022.

Ribeiro, Márcio Moretto; Ortellado, Pablo. O que são e como lidar com as notícias falsas: dos sites de notícias falsas às mídias hiperpartidárias. São Paulo: Sur-Revista Internacional de Direitos Humanos, 2018. Vol. 15, nº 27, p. 71-83. Disponível em: < sur-27-portugues-marcio-moretto-ribeiro-pablo-ortellado.pdf (conectas.org)>. Acesso em: 05/01/2023.

Salinas, Natasha Schmitt Caccia. Avaliação legislativa no Brasil: apontamentos para uma nova agenda de pesquisa sobre o modo de produção das leis. In: UNICEUB. Revista Brasileira de Políticas Públicas. Vol 3, n 2, dez/2013. Brasília: Uniceub, 2013.

Souza, Carlos Affonso; Lemos, Ronaldo. NoC Online Intermediaries Case Studies Series: Brazilian Courts and the Internet- rulings before and after the Marco Civil on Intermediary Liability. In: Gasser, Urs; Schultz, Wolfgang. Governance of Online Intermediaries: observations from a Series of National Case Studies. Cambridge: Berkman Center Research Publication, 2015. Disponível em: <https://dash.harvard.edu/handle/1/16140636>. Acesso em: 09/11/2021.

Sponholz, Liriam. O papel dos discursos de ódio (online) na ascensão da extrema direita: um aporte teórico. Confluências, Niterói, v. 22, n. 3, p. 220-243, dez. 2020/mar. 2021.

Vesting, Thomas. A Mudança da Esfera Pública pela Inteligência Artificial. In: Abboud, Georges; Júnior, Nelson; Campos, Ricardo. Fake News e Regulação. São Paulo: Editora Revista dos Tribunais, 2019. p. 193-210. (Kindle)

Wardle, Claire; Derakhshan, Hossein. Information disorder: Toward an interdisciplinary framework for research and policy making. Strasbourg: Council of Europe, 2017. Disponível em: <<http://tverezo.info/wp-content/uploads/2017/11/PREMS-162317-GBR-2018-Report-desinformation-A4-BAT.pdf>>. Acesso em: 10/01/2023.

Wardle, Claire. Fake news. It' s complicated. First Draft, 2017. Disponível em: <<https://firstdraftnews.org/fake-news-complicated/>>. Acesso em: 28 dez. 2019.

FONTES DOCUMENTAIS

Brasil. Lei nº 12.965/2014, Marco Civil da Internet. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2014/lei/12965.htm. Acesso em: 27/07/2022. Brasil. Câmara dos Deputados. Projeto de Lei nº 2.126, de 2011, do Poder

Executivo, que “Estabelece princípios, garantias, direitos e deveres para o uso da Internet no Brasil.”. Brasília: Câmara dos Deputados, 2011. Disponível em: <https://www.camara.leg.br/proposicoesWeb/prop_mostrarintegra?codteor=912989&filename=Tramitacao-PL%202126/2011>. Acesso em: 31/07/2023.

Brasil. Câmara dos Deputados. Substitutivo Oferecido em Plenário em Substituição à Comissão Especial destinada a proferir Parecer ao Projeto de Lei nº 2.126, de 2011, do Poder Executivo, que “estabelece princípios, garantias, direitos e deveres para o uso da Internet no Brasil”. Brasília: Câmara dos Deputados, 2014a. Disponível em: <<https://www.camara.leg.br/proposicoesWeb/fichadetramitacao?idProposicao=606108>>. Acesso em: 27/07/2023.

Brasil. Câmara dos Deputados. Subemenda Substitutiva Global às Emendas de Plenário ao Projeto de Lei nº 2.126, de 2011, que “estabelece princípios, garantias, direitos e deveres para o uso da Internet no Brasil”. Brasília: Câmara dos Deputados, 2014b. Disponível em: <https://www.camara.leg.br/proposicoesWeb/prop_mostrarintegra?codteor=1225789&filename=Tramitacao-PL%202126/2011>. Acesso em: 27/07/2023.

Brasil. Senado Federal. Emenda nº 41 ao Projeto de Lei da Câmara nº 21/2014 de Autoria do Senador Aloysio Nunes Ferreira. Brasília: Senado Federal, 2014c. Disponível em: <https://legis.senado.leg.br/sdleg-getter/documento?dm=3847392&ts=1630411509775&disposition=inline&_gl=1*11lm1x4*_ga*MjcxNjE5NjI5LjE2ODg1NjQzMzc.*_ga_CW3ZH25XMK*MTY5MTQ5NjcxMi42LjEuMTY5MTQ5Nzc5NC4wLjAuMA..>Acesso em: 08/08/2023.

Brasil. Senado Federal. Parecer da Comissão de Constituição, Justiça e Cidadania, sobre o Projeto de Lei da Câmara nº 21, de 2014 (nº 2.126, de 2011, na origem), do Poder Executivo, que estabelece princípios, garantias, direitos e deveres para o uso da internet no Brasil. Brasília: Senado Federal, 2014d. Disponível em: <https://legis.senado.leg.br/sdleg-getter/documento?dm=3847478&ts=1630411510440&disposition=inline&_gl=1*1vbv4th*_ga*MjcxNjE5NjI5LjE2ODg1NjQzMzc.*_ga_CW3ZH25XMK*MTY5MDkxMTAwOS40LjEuMTY5MDkxNDg5OS4wLjAuMA..>. Acesso em: 08/08/2023.

Brasil. Senado Federal. Parecer da Comissão de Ciência, Tecnologia, Inovação, Comunicação e Informática do Senado Federal, designada, em caráter simultâneo, com as Comissões de Constituição, Justiça e Cidadania e de Meio Ambiente, Defesa do Consumidor e Fiscalização e

Controle, sobre o Projeto de Lei da Câmara nº 21, de 2014 (nº 2.126, de 2011, na origem), do Poder Executivo, que estabelece princípios, garantias, direitos e deveres para o uso da internet no Brasil. Brasília: Senado Federal, 2014e. Disponível em: < https://legis.senado.leg.br/sdleg-getter/documento?dm=3847469&ts=1630411510339&disposition=inline&_gl=1*1vbw4th*_ga*MjcxNjE5NjI5LjE2ODg1NjQzMzc.*_ga_CW3ZH25XMK*MTY5MDkxMTAwOS40LjEuMTY5MDkxNDg5OS4wLjAuMA..>. Acesso em: 08/08/2023.

Brasil. Tribunal Superior Eleitoral. Portaria nº 949, de 7 de dezembro de 2017. Institui Conselho Consultivo de Internet e Eleições. Brasília: Tribunal Superior Eleitoral, 2017. Disponível em: < <https://www.tse.jus.br/legislacao/compilada/prt/2017/portaria-no-949-de-7-de-dezembro-de-2017>>. Acesso em: 15/08/2023.

Brasil. Tribunal Superior Eleitoral. Portaria nº 382, de 24 de maio de 2019. Institui grupo de trabalho incumbido de elaborar propostas de novas linhas de ação do Tribunal Superior Eleitoral sobre desinformação e eleições. Brasília: Tribunal Superior Eleitoral, 2019a. Disponível em: < Institui grupo de trabalho incumbido de elaborar propostas de novas linhas de ação do Tribunal Superior Eleitoral sobre desinformação e eleições.>. Acesso em: 15/08/2023.

Brasil. Tribunal Superior Eleitoral. Portaria nº 663, de 30 de agosto de 2019. Institui o Programa de Enfrentamento à Desinformação com Foco nas Eleições 2020 e disciplina a sua execução. Brasília: Tribunal Superior Eleitoral, 2019b. Disponível em: < <https://www.tse.jus.br/legislacao/compilada/prt/2019/portaria-no-663-de-30-de-agosto-de-2019>>. Acesso em: 15/08/2023.

Brasil. Tribunal Superior Eleitoral. Programa de Enfrentamento à Desinformação com Foco nas Eleições 2020: plano estratégico. Brasília: Tribunal Superior Eleitoral, 2020a. Disponível em: < https://www.justicaeleitoral.jus.br/desinformacao/arquivos/Programa_de_enfrentamento_web.pdf>. Acesso em: 15/08/2023.

Brasil. Tribunal Superior Eleitoral. Memorando de Entendimento-TSE nº 40/2020. Memorando de Entendimento com Google Brasil Internet Ltda ao Programa de Enfrentamento à Desinformação com Foco nas Eleições 2020. Brasília: Tribunal Superior Eleitoral, 2020b. Disponível em: < https://www.justicaeleitoral.jus.br/parcerias-digitais-eleicoes/assets/arquivos/memorando_google.pdf>. Acesso em: 29/08/2023.

Brasil. Tribunal Superior Eleitoral. Memorando de Entendimento-TSE nº 41/2020. Memorando de Entendimento com Facebook Brasil ao Programa de Enfrentamento à Desinformação com Foco nas Eleições 2020. Brasília: Tribunal Superior Eleitoral, 2020c. Disponível em: <https://www.justicaeleitoral.jus.br/parcerias-digitais-eleicoes/assets/arquivos/memorando_facebook.pdf>. Acesso em: 29/08/2023.

Brasil. Tribunal Superior Eleitoral. Programa de Enfrentamento à Desinformação com Foco nas Eleições 2020: relatório de ações e resultados. Brasília, Tribunal Superior Eleitoral, 2021. Disponível em: <https://www.justicaeleitoral.jus.br/desinformacao/arquivos/Programa_de_enfrentamento_resultados.pdf>. Acesso em: 15/08/2023.

Brasil. Tribunal Superior Eleitoral. Programa Permanente de Enfrentamento à Desinformação no âmbito da Justiça Eleitoral: plano estratégico: eleições 2022. Brasília: Tribunal Superior Eleitoral, 2022a. Disponível em: <<https://www.justicaeleitoral.jus.br/desinformacao/arquivos/programa-permanente-de-enfrentamento-a-desinformacao-novo.pdf>>. Acesso em: 30/08/2023.

Brasil. Tribunal Superior Eleitoral. Memorando de Entendimento-TSE nº 1/2022. Memorando de Entendimento com Google Brasil Internet Ltda ao Programa de Enfrentamento à Desinformação com Foco nas Eleições 2022. Brasília: Tribunal Superior Eleitoral, 2022b. Disponível em: < Brasil. Tribunal Superior Eleitoral. Memorando de Entendimento-TSE nº 40/2020. Memorando de Entendimento com Google Brasil Internet Ltda ao Programa de Enfrentamento à Desinformação com Foco nas Eleições 2020. Brasília: Tribunal Superior Eleitoral, 2022b. Disponível em: < https://www.justicaeleitoral.jus.br/parcerias-digitais-eleicoes/assets/arquivos/memorando_google.pdf>. Acesso em: 29/08/2023.>. Acesso em: 29/08/2023.

Brasil. Tribunal Superior Eleitoral. Memorando de Entendimento-TSE nº 03/2022. Memorando de Entendimento com Facebook Serviços Online do Brasil Ltda ao Programa de Enfrentamento à Desinformação com Foco nas Eleições 2022. Brasília: Tribunal Superior Eleitoral, 2022c. Disponível em: <https://www.tse.jus.br/++theme++justica_eleitoral/pdfjs/web/viewer.html?file=https://www.tse.jus.br/comunicacao/noticias/arquivos/assinatura-de-acordos-plataformas-digitais/memorando-tse-e-facebook/@@download/file/MoU%20TSE_FB%20%281%29.pdf>. Acesso em: 29/08/2023.

Brasil. Tribunal Superior Eleitoral. Relatório de Gestão 2022. Brasília: Tribunal Superior Eleitoral, 2023. Disponível em: <
https://www.tse.jus.br/++theme++justica_eleitoral/pdfjs/web/viewer.html?file=https://www.tse.jus.br/transparencia-e-prestacao-de-contas/contas-publicas/prestacao-anual-de-contas/arquivos/relatorio-de-gestao-2022-final/@@download/file/Relatorio_Gestao_2022_>.
 Acesso em: 21/09/2023.

Data Reportal. Digital Global Overview Report 2023. Acesso em: 16/02/2023. Disponível em:
 <<https://datareportal.com/reports/digital-2023-global-overview-report>>.

Facebook. Termos e Políticas, 2023a. Disponível em: <
https://www.facebook.com/policies_center/>. Acesso em: 06/10/2023.

Facebook. Transparency Center, 2023b. Disponível em: < <https://transparency.fb.com/>>.
 Acesso em: 06/10/2023.

Facebook. Comitê de Supervisão, 2023c. Disponível em: < <https://www.oversightboard.com/>>.
 Acesso em: 06/10/2023.

OECD. The Economic and Social Role of Internet Intermediaries. Paris: OECD, 2010.
 Disponível em: <https://www.oecd.org/sti/ieconomy/44949023.pdf>. Acesso em: 09/11/2021.

Meta. Corporate Human Rights Policy, 2023. Disponível em: < <https://about.fb.com/wp-content/uploads/2021/03/Facebooks-Corporate-Human-Rights-Policy.pdf>>. Acesso em:
 06/10/2023.