

CENTRO UNIVERSITÁRIO DE CASCAVEL - UNIVEL
CURSO DE PÓS-GRADUAÇÃO STRICTO SENSU EM DIREITO
ÁREA DE CONCENTRAÇÃO EM DIREITO, INOVAÇÕES E REGULAÇÕES

ADRIANE NOGUEIRA FAUTH DE FREITAS

**A AUTORREGULAÇÃO REGULADA COMO MODELO REGULATÓRIO DA
LIBERDADE DE EXPRESSÃO NAS REDES SOCIAIS**

CASCAVEL – PR
2023

ADRIANE NOGUEIRA FAUTH DE FREITAS

**A AUTORREGULAÇÃO REGULADA COMO MODELO REGULATÓRIO DA
LIBERDADE DE EXPRESSÃO NAS REDES SOCIAIS**

Dissertação apresentada ao Programa de Pós-Graduação Stricto Sensu em Direito – Nível de Mestrado, área de concentração em Direito, inovações e regulações, Linha de Pesquisa em Direito e Inovações Tecnológicas, do Centro Universitário de Cascavel – UNIVEL.

Orientador: Prof. Dr. Júlio César Garcia

CASCADEL – PR
2023

FICHA CATALOGRÁFICA

Dados Internacionais de Catalogação na Publicação (CIP)

Catalogação na fonte elaborada pela Bibliotecária

Andrea Figueredo Ferreira CRB 9/1703

S586p Freitas, Adriane Nogueira Fauth de

A autorregulação regulada como modelo regulatório da liberdade de expressão nas redes sociais -- Cascavel, 2023.

124 p.

Orientador: Júlio Cesar Garcia

Dissertação (Mestrado) – Centro Universitário Univel, Programa de Pós-Graduação em Direito, 2023.

FOLHA DE APROVAÇÃO

ADRIANE NOGUEIRA FAUTH DE FREITAS

A AUTORREGULAÇÃO REGULADA COMO MODELO REGULATÓRIO DA LIBERDADE DE EXPRESSÃO NAS REDES SOCIAIS

Esta dissertação foi julgada adequada para a obtenção do título de Mestre em Direito e aprovada em sua forma final pelo Programa de Pós-Graduação Stricto Sensu em Direito, área de concentração em Direito, Inovações e Regulações, do Centro Universitário de Cascavel – UNIVEL.

BANCA EXAMINADORA

Orientador: Prof. Dr. Júlio César Garcia
Centro Universitário de Cascavel (UNIVEL)

Convidado: Prof. Dr. Fabro Boaz Steibel
Centro Universitário de Cascavel (UNIVEL)

Convidada: Prof. Dra. Aldacy Rachid Coutinho
Centro Universitário de Cascavel (UNIVEL)

Cascavel, 22 de maio de 2023

AGRADECIMENTOS

Gostaria de agradecer ao Prof. Dr. Júlio César Garcia, pela orientação, paciência, encorajamento e, especialmente, humanidade e gentileza que foram essenciais para minha motivação e crescimento acadêmico.

Minha gratidão à Univel, aos Coordenadores e professores do mestrado que contribuíram direta ou indiretamente para o sucesso deste trabalho.

Gostaria também de estender meu agradecimento à minha família, em especial meu esposo Emanuel, que esteve ao meu lado desde o início desta jornada. Seu amor, apoio incondicional e compreensão foram pilares fundamentais para minha realização acadêmica. Agradeço por sempre me incentivar a perseguir meus sonhos e por estar presentes em todos os momentos, mesmo nos mais desafiadores. À minha filha Helena, pela compreensão pelas ausências da mamãe. Ao bebê Emanuel, prestes a nascer, que em momentos difíceis me motivou a continuar.

RESUMO

A criação das plataformas de redes sociais revolucionou as condições sociais com as quais a liberdade de expressão é exercida; velozes, sem limite de tempo e espaço, essas novas tecnologias são a principal ferramenta para a construção, debate e compartilhamento de opiniões e ideias, sendo utilizadas como verdadeiros espaços públicos para o exercício de direitos e liberdades fundamentais. De forma paradoxal, esse novo ambiente, apesar de ampliar o “livre mercado de ideias” e amplificar também a audiência, acabou por desencadear inúmeros desafios que podem colocar em risco a democracia e a própria liberdade de expressão. Dessa forma, a disponibilização do espaço de manifestação para os usuários dos serviços das plataformas e não mais para aqueles que detêm a infraestrutura comunicacional implicou em uma modificação social capaz de exigir uma mudança regulatória. Diante deste cenário, analisa-se o seguinte problema de pesquisa: qual modelo regulatório, estatal ou de autorregulação pelas próprias plataformas, é mais adequado para o exercício da liberdade de expressão nos termos garantidos pela Constituição Federal? A hipótese sustentada é de que a autorregulação regulada é o modelo regulatório mais adequado para a liberdade de expressão nas redes sociais, em especial no novo espaço público por elas criado. A pesquisa é realizada com base no método dedutivo, com emprego da revisão bibliográfica.

PALAVRAS-CHAVE: fake news; moderação de conteúdos; direito digital; regulação; redes sociais.

ABSTRACT

The creation of social media platforms has revolutionized the social conditions under which free speech is exercised; fast, without time and space limits, these new technologies are the main tool for the construction, debate and sharing of opinions and ideas, being used as true public spaces for the exercise of fundamental rights and freedoms. Paradoxically, this new environment, despite expanding the “free market of ideas” and expanding the audience, ended up triggering numerous challenges that could endanger democracy and free speech itself. In this way, making the space for manifestation available to users of platform services and no longer to those who own the communication infrastructure, implied a social change capable of demanding a regulatory change. Given this scenario, this research seeks to answer the following research problem: which regulatory model, state or self-regulation by the platforms themselves, is more appropriate for exercising freedom of expression in the terms guaranteed by the Federal Constitution? The supported hypothesis is that regulated self-regulation is the most appropriate regulatory model for freedom of expression in social networks, especially in the new public space they create. The research is based on the deductive method, using a literature review.

KEYWORDS: fake news; content moderation; digital law; regulation; social network.

SUMÁRIO

INTRODUÇÃO	8
1 A LIBERDADE DE EXPRESSÃO	10
1.1 ÂMBITO DE PROTEÇÃO DA LIBERDADE DE EXPRESSÃO	10
1.2 TITULARES E DESTINATÁRIOS DA LIBERDADE DE EXPRESSÃO E FORMAS DE EXPRESSÃO	21
1.3 LIMITES DA LIBERDADE DE EXPRESSÃO	28
2 LIBERDADE DE EXPRESSÃO NAS REDES SOCIAIS	48
2.1 TRANSIÇÃO DA LIBERDADE DE EXPRESSÃO DO MEIO FÍSICO PARA O VIRTUAL: A ESFERA PÚBLICA DIGITAL	48
2.2 OS DESAFIOS DA LIBERDADE DE EXPRESSÃO NAS REDES SOCIAIS	56
2.3 DA MUDANÇA SOCIAL PARA A MUDANÇA REGULATÓRIA	66
3 A REGULAÇÃO DA LIBERDADE DE EXPRESSÃO NAS REDES SOCIAIS	71
3.1 A REGULAÇÃO DAS REDES SOCIAIS: MODELOS, OBJETO E LIMITES NO CENÁRIO NACIONAL	72
3.2 A AUTORREGULAÇÃO: MODERAÇÃO DE CONTEÚDO	78
3.3 A REGULAÇÃO ESTATAL E A AUTORREGULAÇÃO REGULADA DAS REDES SOCIAIS	92
CONCLUSÃO	110
REFERÊNCIAS BIBLIOGRÁFICAS	113

INTRODUÇÃO

As novas tecnologias, especificamente as redes sociais e o advento da *web 2.0*, mudaram o contexto social em que a liberdade de expressão é exercida. Isso porque essas plataformas criaram um novo espaço de colaboração e conexão entre as pessoas e são hoje a principal ferramenta para a construção, debate e compartilhamento de opiniões e ideias, sendo utilizadas como verdadeiros espaços públicos para o exercício de direitos e liberdades fundamentais e exercendo um papel relevante para a democracia: a de mediadoras do debate público.

Como consequência dessa mudança, de forma relativamente positiva, houve uma amplificação do espaço comunicacional. Em contrapartida, a configuração desse novo espaço híbrido, na medida em que ele é privado por natureza, regido por regras de mercado, mas que permite o exercício de liberdades públicas, desencadeou inúmeros desafios, os quais têm colocado em risco a própria liberdade de expressão.

Essa mudança social não está apenas no surgimento de uma nova tecnologia, mas sim na revolução provocada na infraestrutura da liberdade de expressão, uma vez que a disponibilização do espaço de manifestação pertence não mais àqueles que detêm a infraestrutura, mas aos usuários dos serviços das plataformas.

Ademais, são os atores privados que definem, unilateralmente, as regras e condições para o exercício da liberdade de expressão (políticas e termos de uso), numa espécie de poder normativo, e são eles que primeira e predominantemente solucionam conflitos entre direitos fundamentais, por meio da moderação de conteúdo, agindo por vezes de forma incompatível com os valores constitucionais.

Esse novo cenário de mudanças na complexidade social, a instrumentalização das plataformas para propagação de discursos de ódio e desinformação, as consequências negativas suportadas pela sociedade, bem como as acusações de censura privada pelas plataformas, têm desencadeado um debate atual e urgente no Brasil e no mundo a respeito da eficiência dos modelos de regulação atual das redes, uma vez que os novos desafios não comportam mais soluções antigas.

Dessa forma, o objetivo do presente trabalho é analisar a necessidade de se estabelecer um novo modelo regulatório de proteção da liberdade de expressão no ordenamento jurídico brasileiro, com limites definidos aos usuários, às plataformas e aos próprios governantes, por intermédio da chamada autorregulação regulada.

No primeiro capítulo apresentou-se um arcabouço teórico e conceitual necessário à compreensão do problema, a partir da definição do alcance, titularidade e limites da liberdade de expressão, com aportes do direito comparado e delimitando-se conceitos relevantes, como discursos de ódio e *fake news*.

No segundo capítulo analisou-se a transição do espaço de expressão do ambiente físico para o virtual, conceituando as redes sociais e pontuando alguns desafios que surgiram ao exercício da liberdade de expressão no meio digital, demonstrando, ao final, como essa mudança social aponta para a necessidade de atualizações e ajustes do modelo regulatório.

No terceiro e último capítulo adentrou-se no debate regulatório, delimitando, primeiramente, o espectro regulatório da liberdade de expressão, analisando seu objeto, que são as relações jurídicas existentes nas redes, e traçando limites; na sequência, analisou-se a autorregulação ou regulação privada feita pelas plataformas, tendo como norte a moderação de conteúdo. Finalmente, apresentou-se o modelo de autorregulação regulada no cenário regulatório nacional, fazendo uma abordagem dos principais aspectos dos marcos regulatórios nacionais: a lei 12.965/2014 (Marco Civil da Internet) e o projeto de lei 2.630/2020 (Lei de Transparência e Responsabilidade na Internet).

O trabalho tem como abordagem o método dedutivo, baseando-se em pesquisas bibliográficas.

1 A LIBERDADE DE EXPRESSÃO

A liberdade de expressão, direito humano fundamental, é a essência da democracia livre e pluralista e um dos principais pilares das sociedades civilizadas e tolerantes. Ela corresponde a uma das mais antigas exigências da humanidade, razão pela qual se encontra presente desde as primeiras fases do constitucionalismo moderno.

Contudo, as mudanças sociais provocadas pela revolução digital, em especial pela *web 2.0*, alteraram o ambiente em que predominantemente se exercia a liberdade de expressão, que se desloca dos espaços públicos para os espaços privados das plataformas digitais, principalmente para as redes sociais.

Dada a amplitude do tema e para que se possa compreender esse fenômeno, os desafios que se aventam e como ele impacta diretamente na regulação da liberdade de expressão, este primeiro capítulo conceitual delimita o seu âmbito de proteção, estabelecendo um acordo semântico sobre a definição de liberdade de expressão e seus sentidos e dimensões, seus titulares e destinatários, os modos e meios de expressão, e traçando os limites ao seu exercício.

1.1 ÂMBITO DE PROTEÇÃO DA LIBERDADE DE EXPRESSÃO

A definição semântica que se pretende dar à liberdade de expressão neste trabalho é a de gênero, um direito-mãe ou *Mütterrecht* (MACHADO; BRITO, 2020) em seu sentido mais amplo, que se decodifica em uma série de outras liberdades ou direitos, tais como: a liberdade de manifestação do pensamento, de opinião, exteriorizada pelo direito de reunião e associação; a liberdade de consciência e de crença (religiosa); a liberdade de comunicação e informação, incluindo a liberdade de imprensa; e a liberdade de expressão da atividade intelectual, artística e científica.

Essa definição encontra escopo no art. 19 da Declaração Universal de Direitos Humanos, que apresenta a liberdade de expressão como uma reação consciente às práticas abusivas dos regimes autoritários vencidos na segunda Guerra Mundial:

Artigo 19. Todo ser humano tem direito à liberdade de opinião e expressão; esse direito inclui a liberdade de, sem interferência, ter opiniões e de procurar,

receber e transmitir informações e ideias por quaisquer meios e independentemente de fronteiras. (ONU, 1948, s/p).

Nesse mesmo sentido, a Convenção Americana de Direitos Humanos (Pacto de San José da Costa Rica) estabelece em seu artigo 13:

13.1. Toda pessoa tem direito à liberdade de pensamento e de expressão. Esse direito compreende a liberdade de buscar, receber e difundir informações e ideias de toda natureza, sem consideração de fronteiras, verbalmente ou por escrito, ou em forma impressa ou artística, ou por qualquer outro processo de sua escolha. (CIDH, 1969, s/p).

Dessa forma, passa-se a analisar cada uma das liberdades mencionadas para que se possa compreender o real conteúdo e extensão da liberdade de expressão que se pretende analisar. Primeiramente, importa delimitar o sentido de “liberdade” aqui adotado, pois o desenvolvimento do tema levou a doutrina a produzir uma rica classificação que amplia a sua compreensão, passando pelos sentidos positivo e negativo, interno e externo, subjetivo e objetivo.

Para Isaiah Berlin (1969), a liberdade apresenta dois grandes sentidos: o negativo e o positivo. O sentido negativo, também chamado de liberdade política, é a ausência de constrictão, a área em que se pode agir sem sofrer limitações de terceiros: a liberdade significa o direito de fazer o que se quer sem ninguém o impedir, numa esfera de autonomia individual. É a liberdade “de”. Para o autor, esse sentido poderia ser incorporado à seguinte pergunta: “qual é a área em que o sujeito – um indivíduo ou um grupo de indivíduos – está livre, ou se deveria permitir que fosse, da interferência dos outros?” (p. 11).

O sentido positivo, por sua vez, está relacionado ao desejo de autogovernar e à liberdade de alcançar o que se pretende. É a liberdade “para”, que poderia ser caracterizada como uma resposta à seguinte pergunta: “O que ou quem é a fonte de controle ou de interferência que pode determinar que alguém faça, ou seja, uma coisa e não outra?” (BERLIN, 1969, p. 11).

Para José de Melo Alexandrino (2010 *apud* TORNADA, 2018), a liberdade de expressão também possui dois sentidos, um negativo e o outro positivo. O sentido negativo se coaduna com o sentido apresentado por Isaiah Berlin e significa o direito de não ser coibido de expor e divulgar seus pensamentos, ideias, opiniões e demais formas de expressão, assumindo a condição precípua de direito de defesa, de não

censura. Já o sentido positivo, por sua vez, se refere ao direito de os indivíduos terem acesso aos meios de expressão, o que não implica em um acesso livre aos meios de comunicação social ou em um conceito multifuncional das liberdades de comunicação¹ (SARLET, 2022).

A liberdade de expressão possui ainda duas dimensões que, longe de serem antagônicas, em verdade se complementam: a interna ou substancial e a externa ou instrumental. A dimensão interna ou substancial se refere ao direito de formar a própria opinião, ao pluralismo político, à plenitude do desenvolvimento da personalidade individual a partir de sua interação com a realidade e de suas percepções dela, o que acentua a liberdade de expressão como corolário da dignidade humana. Já a dimensão externa ou instrumental se refere ao direito de exteriorizar esse pensamento por meio do exercício das liberdades mencionadas anteriormente (SILVA, 2022).

Deve-se compreender a liberdade de expressão em suas dimensões subjetiva e objetiva. Numa dimensão subjetiva, ela opera como direito subjetivo, tanto individual como coletivo, exigindo-se perante o Estado um dever de abstenção (garantia da liberdade e da ausência de constrição indevida) e de prestação estatal por meio da proteção e promoção da liberdade de expressão, seja pela regulação ou pelo Poder Judiciário, que tem o dever de concreta e abstratamente controlar a constitucionalidade dos atos que violem ou coíbam a liberdade de expressão. Já na sua dimensão objetiva, a liberdade de expressão assume uma perspectiva de direito transindividual e representa um valor central na comunidade política (SARLET, 2022).

Para esta pesquisa, a liberdade de expressão nas redes sociais deve ser compreendida em todos os sentidos e dimensões apresentados, uma vez que serão vislumbrados desafios para a expressão envolvendo possíveis atos de censura estatal e privada nas redes. Isso afeta o sentido negativo e as dimensões externas e subjetivas, bem como as limitações de acesso à informação e ao pluralismo, pelas *fake news* e formação das *echo chambers* ou bolhas de eco; e também afeta o sentido positivo e a dimensão interna, além do fomento aos discursos de ódio que têm colocado em risco a própria democracia, atacando a dimensão objetiva da liberdade de expressão.

¹ O sentido positivo e o acesso aos meios de comunicação serão analisados no tópico da titularidade da liberdade de expressão.

Uma vez analisados os sentidos e dimensões da liberdade de expressão, passa-se a delimitar a extensão de seu conteúdo, a partir dos direitos que dela se desdobram, apresentando suas peculiaridades e relações com a expressão, a saber: a liberdade de opinião e manifestação do pensamento; a liberdade de reunião e associação; a liberdade de consciência e de crença; a liberdade intelectual (científica e de pesquisa); a liberdade artística e a liberdade de informação e comunicação.

A liberdade de opinião deve ser considerada como a base, o alicerce da liberdade de expressão, e se refere ao direito de exprimir, realizar juízos de valor e se manifestar sobre fatos, opiniões, valores e ideias de terceiros por meio da anuência, da crítica, do protesto, da discordância e da livre circulação de ideias (SARLET, 2022). Por sua vez, a liberdade de manifestação do pensamento deve ser vista como um dos aspectos externos e instrumentais da liberdade de opinião, uma vez que representa o direito à exteriorização da opinião ou de não a exteriorizar, de mantê-la em segredo, de ficar em silêncio² (MENDES; BRANCO, 2022).

A liberdade de reunião e associação também estão estritamente ligadas ao exercício da dimensão instrumental da liberdade de expressão e, por conseguinte, do sistema democrático, porque permitem uma efetiva participação popular no processo político e na tomada de decisões no âmbito público. O direito ou liberdade de reunião³ é considerado como um direito-meio para o exercício de um direito-fim, que é a liberdade de expressão⁴. Por assim dizer, é um direito individual, mas de manifestação coletiva (MENDES; BRANCO, 2022).

A liberdade de associação⁵ apresenta-se também como direito autônomo, mas diretamente ligado à liberdade de expressão, uma vez que autoriza os indivíduos a se

² A CF considera o direito ao silêncio como direito individual fundamental em seu art. 5º, LXIII, devendo ser compreendido tanto no aspecto penal quanto em outras esferas: “LXIII - o preso será informado de seus direitos, entre os quais o de permanecer calado, sendo-lhe assegurada a assistência da família e de advogado” (SILVA, 2022, p. 243).

³ “Art. 5º XVI - todos podem reunir-se pacificamente, sem armas, em locais abertos ao público, independentemente de autorização, desde que não frustrem outra reunião anteriormente convocada para o mesmo local, sendo apenas exigido prévio aviso à autoridade competente” (CF/88).

⁴ Nesse sentido, o Supremo Tribunal Federal (ADPF 187/DF) define: “O Supremo Tribunal Federal, em ambos os casos, deixou claramente consignado que o direito de reunião, enquanto direito-meio, atua em sua condição de instrumento viabilizador do exercício da liberdade de expressão, qualificando-se, por isso mesmo, sob tal perspectiva, como elemento apto a propiciar a ativa participação da sociedade civil, mediante exposição de ideias, opiniões, propostas, críticas e reivindicações, no processo de tomada de decisões em curso nas instâncias de Governo.”

⁵ A liberdade de associação, na Constituição Federal, abrange diversas outras faculdades: “Art. 5º XVII - é plena a liberdade de associação para fins lícitos, vedada a de caráter paramilitar;

reunirem em caráter permanente com traços de estabilidade, ou seja, não de forma esporádica, para atender interesses em comum, desde que tenham finalidades lícitas (MENDES; BRANCO, 2022).

Observa-se que o surgimento das redes sociais impactou diretamente no exercício dessas duas liberdades, pois elas facilitam o encontro entre pessoas com interesses em comum, que é uma característica desses direitos. As redes sociais têm se apresentado como uma ferramenta central para o aumento do número e dimensão das manifestações populares: desde 2013, elas têm sido usadas como instrumento de organização dos movimentos, ampliando a sua divulgação e mobilização⁶.

A liberdade de consciência e de crença⁷, formas de expressão do pensamento, apesar de semelhantes, não se confundem. A liberdade de consciência é mais ampla e se refere à faculdade do indivíduo de adotar valores éticos, existenciais, morais e filosóficos, formando juízos e ideias sobre si e sobre o seu meio externo, incluindo a formação e a exteriorização da sua consciência, bem como a faculdade de não seguir nenhuma delas, acolhendo o ceticismo (SILVA, 2022). Nesse sentido negativo, que assegura a inviolabilidade desse direito, impede-se uma atuação demasiadamente paternalista do Estado, que não deve determinar o que é certo ou errado, deixando para cada indivíduo a liberdade de pensar e agir conforme suas próprias convicções (HESSE, 1995 *apud* SARLET, 2022).

A liberdade de crença se refere ao aspecto religioso, espiritual, ao direito de ter ou não ter uma crença ou religião. Nela está incluído o direito de escolha de qualquer seita religiosa, bem como o direito de mudar de religião. Contempla também a liberdade de culto religioso, que é a exteriorização da fé professada com a prática de

XVIII - a criação de associações e, na forma da lei, a de cooperativas independem de autorização, sendo vedada a interferência estatal em seu funcionamento;

XIX - as associações só poderão ser compulsoriamente dissolvidas ou ter suas atividades suspensas por decisão judicial, exigindo-se, no primeiro caso, o trânsito em julgado;

XX - ninguém poderá ser compelido a associar-se ou a permanecer associado;

XXI - as entidades associativas, quando expressamente autorizadas, têm legitimidade para representar seus filiados judicial ou extrajudicialmente" (CF/88).

⁶ De acordo com o estudo feito pelo grupo Máquina PR/BRandviewer, as manifestações populares de 2013 impactaram 94 milhões de pessoas pelas redes sociais. A rede mais usada para distribuição de conteúdo foi o Twitter (49,3% das citações), seguida pelo Facebook (47,1%) e Google+ (1,9%). (Disponível em: <https://exame.com/tecnologia/ativismo-digital-atingiu-94-milhoes-de-brasileiros/>. Acesso em: 12 mai. 2023).

⁷ "Art. 5º VI - é inviolável a liberdade de consciência e de crença, sendo assegurado o livre exercício dos cultos religiosos e garantida, na forma da lei, a proteção aos locais de culto e a suas liturgias" (CF/88).

ritos, cerimônias e liturgias, incluindo a liberdade de organização religiosa e o proselitismo. É, portanto, o direito de adotar determinada crença, criar sua própria religião ou de recusar qualquer uma delas, abraçando o ateísmo ou o agnosticismo (SILVA, 2022).

Ressalta-se que a liberdade religiosa é precursora da liberdade de expressão, uma vez que surge a partir do conceito de tolerância, que nasce no terreno da controvérsia religiosa. Conforme Norberto Bobbio (2011, p. 149), “o reconhecimento da liberdade religiosa deu origem aos Estados não confessionais; o reconhecimento da liberdade política, aos Estados democráticos.”

Durante anos a Igreja tentou conter a propagação de ideias que fossem contrárias aos dogmas católicos, aquilo que chamava de heresia⁸, o que desencadeou conflitos que se estenderam por toda a Europa. Ao final, viu-se que a luta era inútil, pois nenhum dos lados conseguiria efetivamente a vitória. A partir dessa percepção negativa é que surge a noção de tolerância religiosa⁹, que posteriormente se deslocaria para o contexto político, embrião da liberdade de expressão (RUSSELL, 2017). Como se verá adiante, a intolerância religiosa propagada por meio de discursos de ódio nas mídias digitais, em especial contra minorias religiosas, é um dos desafios da regulação da liberdade de expressão.

Como decursivo da liberdade de expressão tem-se as manifestações intelectuais, artísticas e científicas, que são garantidas pela Constituição Federal, juntamente com a comunicação, que será analisada na sequência, independentemente de licença ou censura¹⁰.

A liberdade intelectual é um gênero que abrange a produção científica e filosófica, a pesquisa – a liberdade de se escrever um livro, tese, dissertação ou artigo.

⁸ Basta lembrar da condenação de Giordano Bruno à fogueira em 1593 por suas teorias similares ao modelo de Copérnico, bem como do julgamento de Galileu Galilei perante a Inquisição por sua defesa do heliocentrismo.

⁹ O conceito de tolerância nesse momento é meramente negativo, o que Bobbio denomina de “primeira dimensão de tolerância” e compreende a ideia de aceitar e tolerar aqueles que não se submetiam aos preceitos católicos, mas não reconhecendo que haja no outro uma verdade em si, apenas aceitando aqueles que desconheciam a verdade religiosa (BOBBIO, 2012).

¹⁰ “Art. 5º IX - é livre a expressão da atividade intelectual, artística, científica e de comunicação, independentemente de censura ou licença; [...]”

Art. 220. A manifestação do pensamento, a criação, a expressão e a informação, sob qualquer forma, processo ou veículo não sofrerão qualquer restrição, observado o disposto nesta Constituição. § 2º É vedada toda e qualquer censura de natureza política, ideológica e artística.” (CF/88).

As expressões artísticas¹¹ se referem às artes plásticas, poesia, música, literatura, teatro e cinema e estão inseridas dentro do direito à cultura, uma vez que são consideradas como patrimônio cultural brasileiro, nos termos do art. 216, III da CF¹² (SILVA, 2022).

O surgimento da internet e das redes sociais ampliou o acesso às produções artísticas e intelectuais. Contudo, tem gerado discussões regulatórias quanto à proteção da propriedade intelectual, aqui incluído o direito autoral, uso de imagem e produção intelectual. O uso de inteligência artificial, como o ChatGPT, se mostrou capaz de reproduzir textos, compor músicas e poemas, e tem suscitado debates sobre a existência ou não de direitos autorais decorrentes das suas criações¹³ (ALLEMAND; MAGRO, 2023).

Por fim, a liberdade de expressão está intrinsicamente ligada aos conceitos de informação e comunicação e possui especial relevância nessa pesquisa porque a expansão informacional provocada pelo surgimento das novas tecnologias, em especial as redes sociais, alterou sensivelmente as condições sociais em que a liberdade de expressão passou a ser exercida (DANTAS; SANTOS, 2020).

A informação, sob um panorama jurídico, pode ser analisada a partir de três perspectivas: o direito de informar (liberdade de informação), o direito de se informar (direito de busca e acesso à informação) e o direito de ser informado (direito de receber informações). Todas elas encontram respaldo constitucional (SARLET, 2022).

A liberdade de informação não se confunde com o direito à informação, visto que a primeira se refere ao direito de informar¹⁴, dar e transmitir informações por qualquer meio. Nesse sentido, identifica-se com a própria liberdade de manifestação do pensamento¹⁵, apresentando-se, numa dimensão subjetiva e negativa, como um

¹¹ A respeito da liberdade artística destaca-se importante decisão do Tribunal Constitucional Alemão, que entende que essa liberdade compreende tanto a obra em si quanto os impactos e efeitos decorrentes dela (SCHWABE; MARTINS, 2006).

¹² “Art. 216. Constituem patrimônio cultural brasileiro os bens de natureza material e imaterial, tomados individualmente ou em conjunto, portadores de referência à identidade, à ação, à memória dos diferentes grupos formadores da sociedade brasileira, nos quais se incluem: III - as criações científicas, artísticas e tecnológicas” (CF/88).

¹³ G1. Disponível em: <https://g1.globo.com/pop-arte/noticia/2023/04/11/chatgpt-e-direito-autoral-entenda-a-treta-juridica-que-ronda-a-relacao-entre-inteligencia-artificial-e-arte.ghtml>. Acesso em: 10 maio 2023.

¹⁴ José Afonso da Silva destaca a liberdade de ensinar, referente ao papel do professor na transmissão do conhecimento, como forma de informação e manifestação do pensamento (SILVA, 2022).

¹⁵ “O direito ou liberdade de informar coincide com a liberdade de manifestação do pensamento pela palavra, por escrito ou por qualquer outro meio de difusão; o direito de ser informado indica o interesse

direito de defesa, de modo que o seu titular não seja impedido de expor e difundir suas ideias e opiniões. Em contrapartida, o direito à informação se desdobra no direito de se informar e o direito de ser informado¹⁶ (SARLET, 2022).

O direito de se informar refere-se ao direito de ter acesso às informações¹⁷, a liberdade individual de poder buscar e escolher as fontes, bem como de não ser impedido de recolhê-las. Esse acesso deve incluir toda e qualquer informação abrangida pela liberdade de expressão.

Já o direito de ser informado se manifesta na possibilidade de qualquer indivíduo receber informações dos Poderes Públicos¹⁸ que sejam de interesse particular, coletivo ou geral, para o exercício livre, responsável e consciente das liberdades públicas, viabilizando o controle social e exercício da cidadania ativa – *status activus processualis*¹⁹ – a partir da transparência e publicidade dos atos públicos (SARLET, 2022).

Para Ingo Sarlet e Molinaro (2014) há, portanto, um “direito da informação” que abrange tanto a liberdade de informação quanto o direito a ela (ser informado e se informar – direito de acesso às informações) (p. 14).

Nesse sentido, observa-se uma noção liberal da liberdade de informação, de abstenção estatal e liberdade individual, mas também uma noção social, na medida em que há um dever prestacional do Estado em garantir um sistema informacional adequado e que permita aos indivíduos terem acesso a informações atuais, verdadeiras e completas, no intuito de satisfazer um direito público à informação²⁰.

sempre crescente da coletividade para que tanto os indivíduos como a comunidade estejam informados para o exercício consciente das liberdades públicas” (GRECCO, 1974, p. 38 *apud* SILVA, 2022, p. 247).

¹⁶ José Afonso da Silva apresenta uma classificação distinta da informação. Para ele, a liberdade de informação se desdobra no direito de informar (manifestação do pensamento) e de ser informado (receber informações), sendo que o direito à informação é tratado como direito coletivo, se referindo a um direito da coletividade e de acesso à informação (SILVA, 2022).

¹⁷ Essa perspectiva foi adotada pela Constituição Federal de 1988, no Art. 5º XIV – “é assegurado a todos o acesso à informação e resguardado o sigilo da fonte, quando necessário ao exercício profissional” (CF/88).

¹⁸ “Art. 5º XXXIII - todos têm direito a receber dos órgãos públicos informações de seu interesse particular, ou de interesse coletivo ou geral, que serão prestadas no prazo da lei, sob pena de responsabilidade, ressalvadas aquelas cujo sigilo seja imprescindível à segurança da sociedade e do Estado” (CF/88).

¹⁹ Expressão cunhada por Peter Haberle, que apresenta uma quinta função e status aos direitos fundamentais: o *status activus processualis*, por meio do qual o cidadão atua como agente ativo e participante nos processos públicos.

²⁰ Sobre o dever Estatal de garantir o acesso a informações adequadas, destaca-se a decisão proferida pelo Supremo Tribunal Federal na ADPF/690, publicada em 15 de abril de 2021, que entendeu que a interrupção da divulgação de informações epidemiológicas referentes à evolução da pandemia da COVID-19 violava o direito fundamental à informação. (ADPF/690).

Esse sentido legitima especialmente a atuação regulatória do Estado sobre as plataformas digitais.

Destaca-se ainda a necessidade da informação (prestada e acessada) ser plural, na medida em que a formação da opinião e a participação dos cidadãos de forma consciente e responsável no ambiente social depende do acesso a informações diversas, inclusive contraditórias, capazes de oportunizar ao indivíduo a sua própria autonomia. Assim, a informação se apresenta como direito, mas também como garantia da democracia e da própria liberdade de expressão.

Isso, por outro lado, não impede que a legislação estabeleça limites tanto ao acesso a informações (direito de se informar) quanto à disponibilidade delas pelo poder público (direito de ser informado) quando em conflito com outros direitos fundamentais, como será analisado mais adiante.

Dada a especificidade da pesquisa, importa ainda diferenciar o recente direito à comunicação do direito à informação, tratados por vezes como sinônimos. Tais termos se tangenciam, mas não se confundem, pois enquanto a informação se refere mais propriamente ao conteúdo da mensagem, a comunicação enfatiza o processo de comunicar – a troca entre dois polos, o fluxo de informações, um transmissor e um receptor que interagem por meio de símbolos, signos, sinais, vozes e escrita, de modo que a comunicação necessariamente se faz numa via de mão dupla, atingindo a cadeia de produção da comunicação social (VANNUCHI, 2018). Para Sarlet e Molinaro (2014), a liberdade de comunicação deve ser compreendida como uma versão atualizada da antiga liberdade de imprensa.

A necessidade de se reconhecer um direito à comunicação partiu da percepção de que as mudanças tecnológicas demonstraram que o direito à informação e liberdade de expressão não contemplam de forma satisfatória as garantias necessárias ao exercício da comunicação.

Um importante documento que inaugura no cenário internacional o direito à comunicação como direito humano fundamental é o relatório *Um mundo, muitas Vozes – Comunicação e Informação na nossa época*, aprovado em outubro de 1980 pela Organização das Nações Unidas para a Educação, a Ciência e a Cultura

(UNESCO), que ficou conhecido como relatório MacBride²¹. Esse relatório denunciava a concentração dos grandes meios de comunicação e as desigualdades sociais e tecnológicas entre os países desenvolvidos e subdesenvolvidos, como desencadeadoras de um desequilíbrio nos fluxos de informação²² (UNESCO, 1983). Apesar de terem passado mais de quatro décadas da sua elaboração, o relatório se mostra atual e pertinente para a problemática levantada na pesquisa.

Uma das relevantes contribuições do relatório é o reconhecimento do direito de comunicar, diretamente relacionado com a democratização da informação na medida em que garante o direito do indivíduo de produzir a informação e de que ela seja ouvida, e não somente o direito de acessar e consumir informações já elaboradas.

Essa garantia de produção da informação e interação é a marca registrada das redes sociais, que revolucionaram o processo comunicacional justamente por não produzirem diretamente o conteúdo como nas mídias tradicionais, mas criarem espaços para que os indivíduos possam, além de produzir conteúdo, também divulgá-lo. Nesse sentido, Vera Veiga França esclarece:

Vários autores têm contribuído para alargar o conceito de comunicação; adoto aqui, para fins de nossa discussão, uma formulação e uma perspectiva que temos chamado de “relacional”, e se expressa na seguinte formulação: a comunicação é um processo de globalidade, em que sujeitos interlocutores, inseridos em uma dada situação, e através da linguagem, produzem e estabelecem sentidos, conformando uma relação e posicionando-se dentro dela. Nesta concepção, a comunicação não é reduzida a uma dinâmica de transmissão, mas é entendida como interação [...] (FRANÇA, 2016, p. 158 *apud* VANNUCHI, 2018, p. 170).

Por extensão, o direito de comunicar implicaria, segundo a UNESCO, na garantia de acesso das comunidades periféricas não apenas à informação produzida (nos países desenvolvidos), mas sobretudo no acesso aos meios de comunicação (sentido positivo da liberdade de expressão), para que pudessem incidir no espaço

²¹ O nome do relatório se deu em homenagem ao presidente da comissão de notáveis e defensor dos direitos humanos Sean MacBride. Apenas como curiosidade, o escritor colombiano Gabriel García Márquez integrou a comissão de notáveis que elaborou o referido relatório.

²² Esse relatório surge em um contexto histórico marcado pela Guerra Fria, de modo que o tema ganhava relevância: protestos dos países referenciados no documento como subdesenvolvidos contra os fluxos dominantes de notícias provenientes das nações industrializadas; ausência de livre circulação da informação; liberdade de imprensa; valor da informação; infraestruturas dos serviços de informação; função, direitos e deveres dos jornalistas e, principalmente, a potência dos meios de comunicação social para colaborar com a resolução de problemas mundiais, sobretudo os impeditivos ao fortalecimento dos ideais democráticos.

público, interagindo e se expressando. O relatório MacBride ainda denunciava os problemas decorrentes dos avanços tecnológicos como capazes de se tornarem obstáculos à democratização da comunicação:

A modernização das tecnologias de comunicação torna o controle popular mais vital, mas também mais difícil. O desenvolvimento de sistemas de informação e bancos de dados em larga escala leva ao acúmulo de enormes quantidades de dados de importância essencial nas esferas social científica, econômica e política. O acesso a essas fontes de informação pode obviamente ser restrito por aqueles que estão no controle ou pode ser ampliado para um amplo público de usuários potenciais. A pressão do público deve ser exercida pela participação democrática nas decisões sobre difusão de conteúdos ou distribuição de informações e contra a concentração de controle sobre a comunicação. Por outro lado, o público deve estar sempre atento à acumulação e distribuição injustificadas oficiais ou privadas de dados pessoais sobre indivíduos. (...). Em suma, os avanços tecnológicos podem se tornar obstáculos ou ameaças à democratização da comunicação; ao mesmo tempo, esses novos serviços também podem levar a estruturas descentralizadas e mais democráticas. (UNESCO, 1983, p. 167).

No cenário nacional, a Comunicação Social foi regulada em capítulo específico na Constituição Federal de 1988 (art. 220 – 224)²³, no título da Ordem Social e, dado o contexto histórico de sua elaboração, buscou garantir ao máximo a liberdade de imprensa, a liberdade de expressão e o direito da informação. Esse capítulo complementa todas as demais liberdades já analisadas.

De acordo com Uadi Lammêgo Bulos (2022), a liberdade de comunicação social possui um sentido amplo e um estrito. Em sentido amplo, abrange toda e qualquer forma de exteriorização do pensamento por meio de qualquer meio de informação de massa, atingindo número indeterminado de pessoas; em sentido estrito, que se insere no próprio sentido amplo, se refere ao ato de expor opinião nos veículos de comunicação social, tais como jornais, revistas, rádio e televisão, bem como os veículos mais recentes, como sites da internet, blogs e outros aplicativos (MARTINS, 2022).

²³ A Comunicação Social foi objeto de previsão na Emenda Constitucional 1/1969, contudo estava inserida no contexto único da liberdade de manifestação do pensamento: “Art. 153 § 8º É livre a manifestação de pensamento, de convicção política ou filosófica, bem como a prestação de informação independentemente de censura, salvo quanto a diversões e espetáculos públicos, respondendo cada um, nos termos da lei, pelos abusos que cometer. É assegurado o direito de resposta. A publicação de livros, jornais e periódicos não depende de licença da autoridade. Não serão, porém, toleradas a propaganda de guerra, de subversão da ordem ou de preconceitos de religião, de raça ou de classe, e as publicações e exteriorizações contrárias à moral e aos bons costumes.” (Emenda Constitucional 1/1969).

Em síntese, pode-se considerar que o direito à comunicação se apresenta de forma mais ampla do que o direito à informação, enquanto esse se refere ao direito de receber, buscar e ter acesso a informações precisas e confiáveis, inclusive dos poderes públicos. O direito à comunicação abrange todos os anteriores e também o direito de participação na comunicação social e formação democrática da informação.

Dessa forma, a liberdade de expressão aqui tratada assume uma função de cláusula geral para as diversas liberdades que envolvem alguma forma de comunicação. Ela é considerada como o direito de o indivíduo poder buscar, receber, acessar, expressar e difundir suas ideias, opiniões e informações por intermédio dos meios de comunicação social, incluindo as redes sociais, bem como de expressar suas atividades, sejam elas intelectuais, artísticas ou científicas e das mais variadas formas, inclusive a religiosa.

Como se pontuou e se aprofundará adiante, as redes sociais afetaram todos os sentidos (positivo e negativo) e dimensões (interna, externa, subjetiva e objetiva) da liberdade de expressão em todos os seus aspectos, desde a forma como se busca, recebe e transmite a informação que formará a opinião até a sua expressão, seja ela de carácter religioso, filosófico, artístico, científico ou de comunicação.

Assim, muito embora se reconheçam peculiaridades entre as diversas liberdades, em especial os diferentes âmbitos de proteção e limites, optou-se, nos demais tópicos, por analisá-las a partir de uma concepção geral, afastando-se de uma abordagem compartimentada; não como um conglomerado de direitos, e sim como partes que se conectam, integradas de forma sistemática no conceito de liberdade de expressão.

Uma vez compreendido o alcance quanto ao conteúdo da liberdade de expressão, faz-se necessário analisar quem são os titulares desse direito e os seus destinatários, bem como as formas de expressão, definindo assim o arcabouço teórico e conceitual necessário para a problematização do tema.

1.2 TITULARES E DESTINATÁRIOS DA LIBERDADE DE EXPRESSÃO E FORMAS DE EXPRESSÃO

O presente tópico visa avaliar uma possível variação de amplitude do direito à liberdade de expressão, a depender do seu titular e apesar de sua universalidade,

bem como definir os destinatários desse direito: se são apenas o poder público ou se o dever de garantia, proteção e promoção da liberdade de expressão se estende aos particulares ou entes privados, aqui especialmente as plataformas de redes sociais, como destinatárias diretas da liberdade de expressão, encerrando com as formas de expressão.

Os titulares (sujeitos ativos) desse direito devem ser compreendidos como todas as pessoas físicas, independentemente de características pessoais (como nacionalidade, gênero e idade) ou jurídicas, públicas ou privadas. Trata-se de um direito universal à autodeterminação individual e coletiva, na medida em que todos, individual e coletivamente, têm o direito à livre circulação de ideias e informações (MACHADO; BRITO, 2020). Como destinatários (sujeitos passivos) desse direito, tem-se diretamente o poder público (eficácia vertical) e direta e indiretamente os particulares (eficácia horizontal), tendo eficácia irradiante para todo o ordenamento jurídico (SARLET, 2022).

A eficácia horizontal, entre particulares, pode ser exemplificada pelo direito de resposta ou pleitos indenizatórios, bem como e especialmente pelo dever de proteção e garantia da liberdade de expressão, que também podem ser exigidos em face das plataformas de redes sociais²⁴.

Assim, a proteção à liberdade de expressão pode ser invocada quando a ameaça parte do Estado, numa relação vertical, e quando a ameaça parte de outros indivíduos ou empresas privadas, como é o caso das relações envolvendo as redes sociais, que constituem estruturas corporativas que exercem poder sobre os espaços nos quais as pessoas se manifestam, determinando o que pode ou não ser dito (MARTINS HARTMANN; SARLET, 2019).

Dessa forma, há que se reconhecer efeitos diretos *prima facie* no âmbito das relações jurídicas travadas nas redes sociais, em uma relação de complementariedade entre a vinculação do Estado (eficácia vertical) e a vinculação dos atores privados (eficácia horizontal). Esse reconhecimento é condição para que se discuta a regulação das redes no país.

²⁴ Sobre a eficácia direta e horizontal dos direitos fundamentais nas mídias sociais, recomenda-se o artigo de Ingo Wolfgang Sarlet, *Direitos Fundamentais e Direito privado: a proteção da liberdade de expressão nas mídias sociais* (2019).

Esse direito, apesar de universal, pode apresentar amplitude diversa a depender do seu titular (MARTINS, 2022), como por exemplo a figura dos parlamentares, os quais usufruem de uma liberdade de expressão mais ampla que os cidadãos em geral, dada a sua imunidade material (*freedom of speech*).

A CF/88 estabelece, em seu art. 53, que deputados e senadores são invioláveis, civil e penalmente, por quaisquer de suas palavras, opiniões e votos, ou seja, eles não cometem crimes ou ilícitos civis ao expressarem suas opiniões²⁵, desde que as manifestações guardem conexão com o desempenho da função legislativa e sejam proferidas em razão desta. Tal liberdade não é absoluta e não deve servir como escudo protetivo para a prática de atividades ilícitas²⁶.

Ademais, essa imunidade parlamentar é reconhecida inclusive em manifestações proferidas fora do recinto parlamentar, como em entrevistas, pronunciamentos, palestras e inclusive nas redes sociais. Nesse sentido, o projeto de lei 2630/2020²⁷, que será melhor analisado posteriormente, estende a imunidade material prevista no texto constitucional para as plataformas de redes sociais, de forma alinhada à jurisprudência dos tribunais superiores²⁸.

Essa relativização quanto à amplitude da liberdade de expressão encontra correspondência nas políticas de conteúdo das redes sociais. Em 2019, o Twitter alterou sua política de conteúdo, determinando que manifestações contrárias às regras da plataforma não fossem imediatamente excluídas quando proferidas por candidatos ou membros do governo (presidente, deputados, senadores) com mais de 100.000 seguidores. Nesse caso, a publicação exige um duplo clique para ser acessada (RAMOS, 2021).

²⁵ O reconhecimento da imunidade não impede que os parlamentares sejam representados por quebra de decoro parlamentar.

²⁶ Destaca-se a decisão do Supremo Tribunal Federal, na AP 1044/DF, envolvendo o ex-deputado Daniel Silveira, que o condenou pelo crime do art. 286 do Código Penal (incitação ao crime), mesmo tendo sido cometido na tribuna da Câmara dos Deputados, por entender que discursos de ódio não estariam protegidos pela liberdade de expressão e pela imunidade parlamentar.

²⁷ “Art. 33 § 6º A imunidade parlamentar material, na forma do art. 53 da Constituição Federal, estende-se aos conteúdos publicados por agentes políticos em plataformas mantidas pelos provedores de redes sociais e mensageria privada.” (PL 2630/2020).

²⁸ Segundo a Ação Originária 2002: “[...] Art. 53 da Constituição Federal. Imunidade parlamentar. Ofensas em entrevistas a meios de comunicação de massa e em postagens na rede social “WhatsApp”. O ‘manto protetor’ da imunidade alcança quaisquer meios que venham a ser empregados para propagar palavras e opiniões dos parlamentares. Precedentes. Possível aplicação da imunidade a manifestações em meios de comunicação social e em redes sociais.” (AO 2022 DF).

A liberdade de expressão no âmbito da expressão artística também possui maior amplitude. Isso porque a arte em si visa mexer com as emoções, possuindo um caráter inovador, provocador, subversivo e até mesmo agressivo que pode chocar determinadas pessoas, de modo que a mera discordância do público não pode ser capaz de impedir a sua exibição²⁹. Assim, estão amparadas pela liberdade artística aquelas expressões que contrariem radicalmente as concepções predominantes e ortodoxas em dado momento histórico-cultural.

Em posição de destaque no exercício da liberdade de expressão encontram-se os professores, por meio da chamada liberdade de cátedra, prevista na Constituição Federal, art. 206 e 207, que garantem a “liberdade de aprender, ensinar, pesquisar, divulgar o pensamento, a arte e o saber³⁰”, direito esse que deve ser garantido inclusive para aulas presenciais, online e manifestações em redes sociais.

Sobre o tema, o Supremo Tribunal Federal julgou inconstitucional a Lei 7.800/2016 do Estado de Alagoas, conhecida como “Lei da escola sem partido”, que estabelecia um dever de neutralidade política e ideológica aos professores. De acordo com o Supremo, isso seria incompatível com os valores e princípios constitucionais, especialmente com a proteção ao pluralismo de ideias e concepções pedagógicas, além de implicar na “não tolerância de diferentes visões de mundo, ideologias e perspectivas políticas em sala³¹”, em verdadeira afronta à liberdade de expressão.

Em contrapartida, para juízes e promotores, a liberdade de expressão apresenta amplitude menor, pois a CF/88³² veda atuação político-partidária às referidas autoridades. Além da evidente proibição de filiação a partidos políticos, tal vedação tem se estendido a expressões de apoio ou críticas a candidatos ou lideranças políticas.

A resolução 305/2019 do Conselho Nacional de Justiça definiu regras para o uso das redes sociais por magistrados, visando compatibilizar o exercício da liberdade de expressão com os valores de independência e imparcialidade inerentes a função

²⁹ Relembra-se a decisão do STF na Rcl 38782, envolvendo o grupo Porta dos Fundos, que autorizou a exibição do *Especial de Natal* por entender que a obra não incitava a violência contra grupos religiosos, mas apenas realizava crítica por meio de sátira a elementos cristãos.

³⁰ CF/88.

³¹ ADI 5.580 Alagoas.

³² “Art. 95, parágrafo único, III. Os juízes gozam das seguintes garantias: Parágrafo único. Aos juízes é vedado: III - dedicar-se à atividade político-partidária. Previsão idêntica no art. 128, II, e para os membros do Ministério Público.” (CF/88).

jurisdicional. O art. 4º da resolução define condutas vedadas aos magistrados nas redes sociais:

emitir opinião que demonstre atuação em atividade político-partidária ou manifestar-se em apoio ou crítica públicos a candidato, lideranças políticas ou partidos políticos”, essa vedação, contudo, “não abrange manifestações, públicas ou privadas, sobre projetos e programas de governo, processos legislativos ou outras questões de interesse público, de interesse do Poder Judiciário ou da carreira da magistratura, desde que respeitada a dignidade do Poder Judiciário (BRASIL, 2019b).

Dessa forma, há uma maior relativização – aqui no sentido de limitação – da liberdade de expressão dos magistrados, equilibrando-se com a dignidade e independência do Poder Judiciário.

Por fim, os jornalistas³³ e os veículos de comunicação social assumem um “direito qualificado de liberdade de expressão” (MACHADO; BRITO, 2020), posto que assumem papel de destaque e protagonismo na produção e difusão das informações, amplificando o debate público.

Especialmente em relação ao sujeito passivo da liberdade de expressão, no contexto privado (eficácia horizontal), e à figura do jornalista discute-se a possibilidade de admitir uma liberdade de expressão em face do veículo de comunicação social, na hipótese, por exemplo, de discordar da linha editorial adotada pela empresa, numa espécie de liberdade de imprensa *interna corporis*.

Para Gilmar Mendes (2022), embora a liberdade de imprensa seja regida pela pluralidade, não há razões para se impor a liberdade de expressão do jornalista nessa relação privada, por entender que a eficácia horizontal não se aplica automaticamente e deve ser compatibilizada com outros interesses envolvidos. Ademais, eventual manifestação do jornalista pode gerar consequências civis tanto para ele quanto para a empresa³⁴. Dessa forma, caso a orientação editorial possa afetar a honra, consciência ou outros valores morais do jornalista, ele poderá romper com o contrato

³³ O jornalista, para exercer a liberdade de expressão profissionalmente, não precisa ter diploma de curso superior, como decidiu o STF no Recurso Extraordinário 511.961.

³⁴ Segundo a Súmula 221 do Superior Tribunal de Justiça: “São civilmente responsáveis pelo ressarcimento de dano, decorrente de publicação pela imprensa, tanto o autor do escrito quanto o proprietário do veículo de divulgação.” (Súmula 221 – STF).

sem qualquer ônus, o que se convencionou chamar de “cláusula de consciência” (*clause de conscience*)³⁵ (FARIAS, 2001).

Observa-se que a liberdade de expressão, inobstante ser um direito de titularidade universal, pode ser mais ou menos relativizada a depender do seu titular, daquele que está a realizar a expressão e do contexto em que é realizada, podendo variar de uma liberdade mais ampla, como é para congressistas, artistas e jornalistas, até uma liberdade mais restrita, como para juízes e membros do Ministério Público.

O termo “expressão” também deve ser entendido de forma bastante ampla, contemplando todas as formas de manifestação, desde que não violentas, como a palavra verbal ou não verbal³⁶, a palavra escrita, um filme, uma pintura, uma peça de teatro, fotografias, danças, desenhos, símbolos, uma tatuagem, uma passeata³⁷, postagens em blogs, participações em podcasts ou qualquer outro ato que possa expressar um pensamento, uma determinada opinião, incluindo o direito de não a expressar (SARLET, 2022).

Considera-se ainda como liberdade de expressão os chamados “atos simbólicos”, como por exemplo a queima de uma bandeira ou o uso de um símbolo³⁸.

³⁵ Edilson Farias, em *Liberdade de expressão e comunicação* (2001, p. 131), destaca: “O direito estipulado em favor do comunicador social, atribuindo-lhe a faculdade de não cumprir imposições profissionais contrárias à sua convicção interior, é o que se convencionou denominar de *clause de conscience*. A cláusula de consciência tem sua origem na lei francesa, de 29 de março de 1935, que conferia ao jornalista a faculdade de rescindir seu contrato e obter uma indenização correspondente ao valor pago, nas hipóteses de demissão arbitrária ou sem justa causa, caso o órgão ao qual prestava seus serviços profissionais modificasse a sua orientação editorial de tal monta que pudesse afetar a honra, a consciência ou outros valores morais do jornalista.”

³⁶ Destaca-se a decisão do Supremo Tribunal Federal no HC 83996/RJ Trata-se de Habeas Corpus impetrado pelo réu, que teria realizado gestos obscenos para a plateia após a apresentação, em sinal de protesto às vaías do público. Entendeu o Supremo que não havia crime, uma vez que manifestações não verbais, como gestos, podem ser consideradas como liberdade de expressão. A decisão levou em consideração, ainda, o público-alvo do espetáculo para a não caracterização do crime. (HC 83996/RJ). Outro exemplo de expressão não verbal que encontra proteção na jurisprudência brasileira são as tatuagens. O Supremo Tribunal Federal, no RE 898.450, entendeu que tatuagens ou outras formas de marcas permanentes realizadas intencionalmente no corpo do indivíduo por sua livre escolha consistem em autêntica forma de liberdade de expressão. O caso envolvia a possibilidade de se restringir o acesso de indivíduos que possuem tatuagens a determinados cargos públicos. Dessa forma, o STF entendeu que o candidato possuir uma tatuagem pelo corpo não pode ser, por si, um impedimento para o exercício de cargo público, desde que a tatuagem não expresse ideologias terroristas, extremistas e contrárias às instituições democráticas, ou seja, qualquer marca que propague a intolerância e o discurso de ódio. (RE 898.450).

³⁷ Determinado comportamento, apenas por si, pode caracterizar uma expressão. Destaca-se, por exemplo, a decisão do Supremo Tribunal Federal a respeito da Marcha da Maconha: entendeu-se que a manifestação pela descriminalização do uso da droga não pode ser considerada como apologia ao crime (STF ADPF 187).

³⁸ A Suprema Corte dos Estados Unidos, no caso *Tinker v. De Moines Independent Community School District* (1969), decidiu que o uso de uma faixa preta no braço pelos estudantes dentro de uma escola

Mesmo que não haja palavras, tais atos devem ser considerados como formas de expressão (WARBURTON, 2020).

Destaca-se, em relação às expressões simbólicas, que o seu grau de aceitação pode variar de acordo com a cultura de um determinado Estado. A Suprema Corte Americana considerou como inconstitucional uma lei estadual que criminalizava a conduta de queimar a bandeira, por entender o ato como um exercício da liberdade de expressão (MENDES; BRANCO, 2022). Já na Alemanha, em 2020, tornou-se ilegal a queima de bandeiras estrangeiras por entender-se que tal manifestação não estaria acobertada pela liberdade de expressão, antes seria expressão de ódio e agressão³⁹.

No Brasil, a prática de destruir ou ultrajar a bandeira e outros símbolos nacionais já foi considerada crime. O Decreto 898 de 1969 previa pena de detenção de 2 a 4 anos para esse tipo de ação, mas foi revogado pela Lei de Segurança Nacional de 1978, que definiu a pena de reclusão de 1 a 4 anos. Em 1983, o crime foi extinto por uma nova versão da Lei de Segurança Nacional. Atualmente, o comportamento pode ser considerado como contravenção penal⁴⁰, prevista no art. 35 da Lei 5.700/1971, que trata dos símbolos nacionais.

A expressão ainda pode se dar entre interlocutores presentes ou ausentes. A entre presentes pode ser de pessoa para pessoa, como em um diálogo ou conversa, ou de uma pessoa para outras pessoas, como em uma palestra, discurso ou *live*. Já a entre ausentes pode ser feita para pessoas determinadas, como em uma correspondência, telefone e mensagens, ou indeterminadas, como em um livro, rádio, televisão, redes sociais abertas etc. (SILVA, 2022).

A expressão pode ser realizada tanto no ambiente físico quanto virtual. Muitas vezes uma manifestação realizada de pessoa para pessoa no meio virtual pode se transformar em uma manifestação para outras pessoas, dado o alcance permitido nesse ambiente. É fácil imaginar, por exemplo, um *podcast* em que os interlocutores conversam entre si (pessoa para pessoa), mas com transmissão pelo YouTube (pessoa para pessoas), alcançando indivíduos indeterminados.

em protesto à Guerra do Vietnã era um ato protegido pela liberdade de expressão. (Disponível em: <https://www.oyez.org/cases/1968/21>. Acesso em: 10 mai. 2023.)

³⁹ A lei foi proposta após manifestantes queimarem bandeiras de Israel em protestos ocorridos em Berlim. (DW. Disponível em: <https://www.dw.com/pt-br/alemanha-torna-ilegal-queima-de-bandeiras-estrangeiras/a-53454896>. Acesso em: 10 maio 2023).

⁴⁰ Há o projeto de Lei 2.303/2022, que propõe a criminalização dos atos que desrespeitem os símbolos nacionais. O projeto encontra-se aguardando decisão do Plenário do Senado.

A liberdade de expressão envolve a escolha do meio e forma para a exteriorização do pensamento e das informações. Esses meios são os próprios veículos de comunicação, que podem ou não ser sociais, desde uma conversa pessoal até o uso de veículos de comunicação social, como livros, jornais, periódicos, televisão, rádio, os espaços públicos e, atualmente, as redes sociais. Isso porque o meio e a forma de exteriorização não apenas transmitem um tipo de informação, mas também afetam diretamente “a percepção, o pensamento e a forma de experiência humana” (ABBOUD; CAMPOS, 2020, p. 151).

Dessa forma, conclui-se que a liberdade de expressão, apesar de ser um direito universal, pode alcançar maior ou menor amplitude a depender do seu titular, além de ter como destinatários tanto o Estado quanto particulares, aqui incluídas as plataformas privadas das redes sociais, que possuem tanto o dever de não restringir a liberdade de expressão (aspecto negativo), quanto de proteger e promover essa liberdade (aspecto positivo). A exata compreensão da titularidade e dos destinatários se apresenta como um elemento estruturante da regulação, pois é a partir deles que se estabelecem as relações jurídicas nas redes sociais que serão objeto de regulação.

Finalizando a base teórica que se pretende dar à pesquisa, é necessário compreender os limites da liberdade de expressão, os quais podem variar a depender do ordenamento jurídico analisado.

1.3 LIMITES DA LIBERDADE DE EXPRESSÃO

O presente tópico visa apresentar os principais limites à liberdade de expressão, com aportes do direito comparado, por meio do cotejo entre dois modelos de proteção, o americano e o alemão, e apresentando as suas diferenças e influências para a definição dos limites da expressão no cenário brasileiro.

A liberdade de expressão não se confunde com licença ou uma autorização para toda e qualquer forma de expressão e não se trata de um direito absoluto de se expressar como, onde e quando quiser, sobre quem ou tudo o que quiser. Ela não pode ser vista como uma zona livre em que se possa fazer tudo e na qual nada importa; muito pelo contrário, a expressão importa.

É cediço que os direitos fundamentais têm como principal característica a sua limitabilidade. Não há direito fundamental absoluto porque razões de relevante

interesse público ou exigências derivadas do princípio da convivência das liberdades legitimam, ainda que excepcionalmente, a adoção – por parte dos órgãos estatais e, como se verá, também por entes privados – de medidas restritivas das prerrogativas individuais ou coletivas, desde que respeitados os termos estabelecidos pela própria Constituição, segundo a ADPF 130/DF.

E é nesse ponto que reside a dificuldade mais sensível aos filósofos e juristas, que é determinar as exceções à liberdade de expressão, situações em que certo valor concorrente tenha prevalência sobre a liberdade sem que se enseje em um ato de censura indesejável. Isso porque qualquer ato de censura tolerado pode abrir caminho para um terreno escorregadio que poderá desembocar em um governo arbitrário, colocando em risco a própria democracia. Ronald Dworkin explica que

A liberdade de expressão é uma condição para que um governo seja legítimo. Leis e políticas não são legítimas, salvo se tiverem sido adotadas a partir de um processo democrático, e um processo não é democrático se o governante tenha coibido alguém de expressar suas convicções sobre como essas leis e políticas deveriam ser (DWORKIN, 2006, s/p, tradução nossa).

As limitações à liberdade de expressão possuem níveis diferentes a depender do país analisado, especialmente em decorrência da sua cultura, tradições, costumes e princípios. Para Winfried Brugger (2003), pode-se dividir os países em dois grupos, os quais apresentam diferentes níveis de restrições à liberdade de expressão: o primeiro grupo prioriza a liberdade de expressão em detrimento da maioria dos demais direitos fundamentais, mesmo quando se trata de discursos de ódio. Esse grupo é formado por países que adotam forte influência dos Estados Unidos da América, denominado de posição ou modelo americano. Já o segundo grupo de países atribui maior grau de proteção à dignidade humana ou à igualdade do que à liberdade de expressão, em especial quando se trata de discursos de ódio, que além de desprotegidos devem ser punidos. Essa posição é compartilhada pela Alemanha, diversos países europeus e também pelo Brasil, sendo denominada de posição ou modelo alemão.

O modelo americano, influenciado pelos ideais liberais, encontra seu fundamento no chamado “livre mercado de ideias”, que tem como fonte os escritos de John Milton, na obra *Areopagítica - discurso sobre a liberdade de expressão*, de 1644, e de John Stuart Mill, na obra *Sobre a Liberdade*, de 1859. Para Mill (2019), liberdade

significa a proteção contra a tirania dos governantes e da maioria nessa eterna luta entre liberdade e autoridade; sua obra é, portanto, uma defesa da tolerância das ideias da minoria.

Nesse livre mercado de ideias, nenhuma opinião, majoritária ou minoritária, certa ou errada, verdadeira ou falsa, pode ser silenciada. Dessa forma, a defesa da liberdade de expressão encontra dois grandes argumentos: a busca da verdade (teoria da verdade) e a presunção de falibilidade (WARBURTON, 2020).

A livre troca de ideias é que permite a emergência da verdade: mesmo a opinião falsa, incorreta, minoritária e não popular tem a sua relevância na medida em que é a partir do seu confronto que a verdade é reforçada. Se uma expressão é censurada ou calada e for verdadeira, a humanidade perdeu a oportunidade de trocar o erro pela verdade; se for falsa, o prejuízo é quase o mesmo, pois perdeu-se a oportunidade de refinar e confirmar a verdade a partir da colisão com o erro (MILL, 2019). Dessa forma, o aperfeiçoamento moral da sociedade pode ser alcançado não pela censura e restrição da liberdade de expressão, mas pela argumentação, confronto e persuasão.

A pluralidade, portanto, é fundamental para garantia da liberdade de expressão, uma vez que negar espaço para uma opinião por estar convicto da sua falsidade é presumir que a própria certeza é absoluta, arrogando-se infalibilidade. A certeza não garante a veracidade daquilo que se acredita conhecer e por isso, para Mill (2019), o pensador sério é aquele que se engaja com os que dele discordam, debatendo, argumentando e compreendendo seus pontos de vista.

A liberdade para contraditar posições ortodoxas e majoritárias é uma condição para o desenvolvimento humano, pois se a divergência for ausente seremos pensadores menos ativos e acomodados; quem perde com isso é a própria existência humana, “o mal peculiar em calar a expressão de uma opinião é que se rouba da raça humana – mais de quem discorda da opinião do que de quem a defende” (MILL, 2019, p. 75). Para Peter Häberle, o discurso proibido encoraja o ódio e a conspiração, contrapondo-se à vontade constitucional⁴¹.

Essa discussão de John Stuart Mill (2019) sobre o confronto levar à verdade pode ser exemplificada no caso específico de julgamento de difamação envolvendo os historiadores David Irving e Deborah Lipstadt. De acordo com Nigel Warburton

⁴¹ STF ADPF 187/DF.

(2020), a historiadora Deborah Lipstadt publicou, em 1994, um livro chamado *Denying the holocaust*, no qual criticava o historiador David Irving, que negava o Holocausto. Irving processou Deborah por difamação. Na Inglaterra o ônus da prova pertence ao acusado e não ao acusador, logo Deborah tinha que provar a pertinência de sua avaliação sobre Irving.

Ao final, o tribunal concluiu que era incontroverso que ele se qualificava como negador do Holocausto e afirmou que ele distorceu fatos de forma deliberada para conciliar com suas crenças políticas. Ao final, Irving ficou, mais do que nunca, desacreditado na Inglaterra. Contudo, na Áustria as leis contra o Holocausto são mais severas e, por isso, ao ingressar no país em 2006, ele foi preso e lá é considerado um mártir da liberdade de expressão (WARBURTON, 2020).

Dessa forma, a proibição do discurso, ainda que falso ou ofensivo, pode receber maior credibilidade justamente por ter sido suprimida e não abertamente refutada. Nas palavras de Warburton (2020, p. 43), “Um dos motivos pelos quais discursos falsos e ofensivos são permitidos na maioria das democracias liberais é justamente porque a melhor resposta para o discurso ruim é o discurso bom e não a censura”.

A teoria de John Stuart Mill, no entanto, não apresenta uma liberdade de expressão absoluta, ainda que bastante ampla. Na obra *Sobre a liberdade* (2019), o autor consagra o chamado “princípio do dano”, uma teoria consequencialista que justificaria a limitação da liberdade. Para ele, o Estado só poderia, de forma legítima, limitar a liberdade de um indivíduo contra a sua vontade para impedir um dano a outro indivíduo, e não a si mesmo:

O objetivo deste Ensaio é defender como indicado para orientar de forma absoluta as intervenções da sociedade no individual, um princípio muito simples, quer para o caso do uso da força física sob a forma de penalidades legais, quer para o da coerção moral da opinião pública. Consiste esse princípio em que a única finalidade justificativa da interferência dos homens, individual e coletivamente, na liberdade de ação de outrem, é a autoproteção. O único propósito com o qual se legitima o exercício do poder sobre algum membro de uma comunidade civilizada contra a sua vontade é impedir dano a outrem. O próprio bem do indivíduo, seja material seja moral, não constitui justificção suficiente. O indivíduo não pode legitimamente ser compelido a fazer ou deixar de fazer alguma coisa, porque tal seja melhor para ele, porque tal o faça mais feliz, porque na opinião dos outros tal seja sábio ou reto. Essas são boas razões para o admoestar, para com ele discutir, para o persuadir, para o aconselhar, mas não para o coagir, ou para lhe infligir um mal caso aja de outra forma. Para justificar a coação ou a penalidade, faz-se mister que a

conduta de que se quer desviá-lo tenha em mira causar dano a outrem. (MILL, 2019, p. 35).

Nessa medida, no que diz respeito unicamente a si mesmo, o indivíduo adulto e com condições de se autodeterminar é soberano de seu corpo e espírito e não deve ser protegido de suas próprias ações, afastando qualquer comportamento paternalista por parte do Estado.

O dano aqui desenhado por John Stuart Mill (2019), limitador da liberdade de expressão, é o dano físico e não apenas o psicológico ou econômico. Dessa forma, se um indivíduo se sente ofendido por aquilo que foi dito ou expressado, não está autorizada a censura enquanto silenciar-se o ofensor, porque qualquer um que se contrapõe a uma determinada ideia pode soar ofensivo para quem está a sustentá-la. Contudo, atualmente se reconhece que o dano moral, psicológico e econômico provocado por uma ofensa pode em muitos casos ser pior que o dano físico, afastando-se aqui do conceito originário de dano esboçado por John Stuart Mill.

Sendo assim, caso uma expressão possa, em dado contexto, configurar um incitamento de dano, ela deve ser limitada. Nessa medida, o contexto da expressão importa e deve determinar se ela deve ou não ser tolerada. Com bases nos ensinamentos de John Stuart Mill é que o modelo americano se desenha, sendo a teoria da verdade seu principal alicerce.

A teoria da verdade entende que a liberdade de expressão é um instrumento para potencializar e permitir o acesso à verdade e que a expressão é regulada e combatida apenas por meio de outras ideias nesse livre mercado. Essa visão pragmática da verdade surgiu a partir da evolução da jurisprudência da Suprema Corte Americana, em especial pelos votos do juiz Oliver Wendell Holmes Jr., que desenvolveu uma teoria pela qual se pudesse definir critérios claros e compreensíveis para determinar se um ato de expressão pode ou não ser limitado, o que se denominou de teoria ou teste do perigo claro e iminente (LAURENTIIS; THOMAZINI, 2020).

Em março de 1919, a Suprema Corte Americana, no caso *Schenck v. United States*⁴², condenou os socialistas Charles Schenck e Elizabeth Baer, com base na Lei de Espionagem de 1917, por distribuírem 15.000 panfletos que exortavam os jovens

⁴² *Schenck v. United States*. Disponível em: <https://www.oyez.org/cases/1900-1940/249us47>. Acesso em: 10 mai. 2023.

a desobedecer ao recrutamento militar durante a Primeira Guerra Mundial, alegando que isso era um “erro monstruoso contra a humanidade em prol dos interesses da classe seleta de Wall Street” (WARBURTON, 2020, p. 69).

Em sua defesa, os jovens invocavam a Primeira Emenda Americana⁴³, reivindicando o direito à liberdade de expressão. Em seu voto, o juiz Oliver Wendell Holmes Jr. concluiu que os tribunais deviam maior deferência ao governo durante a guerra, mesmo quando os direitos constitucionais estivessem em jogo. Apresentou, pela primeira vez, o chamado “teste do perigo claro e iminente”, ou seja, as limitações a liberdade de expressão só podem ocorrer quando o discurso produza um perigo claro e iminente.

O juiz concluiu que a disseminação generalizada dos folhetos era suficiente para interromper o processo de recrutamento e que, portanto, estaria configurado esse perigo. Nessa mesma decisão, ele defendeu que a liberdade de expressão não deveria abarcar a liberdade para gritar “fogo” em um auditório lotado quando não há fogo algum; gritar “fogo” nessa circunstância pode provocar um tumulto capaz de gerar graves danos⁴⁴.

Observa-se que o contexto em que a expressão é exercida importa para sua aceitação ou limitação, de modo que a mesma expressão pode ser tolerada ou não, a depender da situação em que está inserida. Nas palavras de Oliver Wendell Holmes Jr.:

A pergunta que paira em toda situação é se as palavras usadas foram usadas em determinadas circunstâncias e são de determinada natureza que criam um perigo claro e iminente de que provocarão os males substanciais que o Congresso tem o direito de coibir. Trata-se de uma questão de proximidade e de grau. Quando a nação está em guerra, muitas coisas que poderiam ser ditas em tempo de paz podem representar tamanho obstáculo aos seus esforços que a expressão delas não será tolerada enquanto os homens estiverem lutando, e nenhum tribunal poderá considerá-las protegidas por nenhum direito constitucional. (HOLMES JR., 1992, página 315 *apud* WARBURTON, 2020, p. 20).

⁴³ “O congresso não deverá fazer qualquer lei a respeito de um estabelecimento de religião, ou proibir o seu livre exercício; ou restringindo a liberdade de expressão, ou da imprensa; ou o direito das pessoas de se reunirem pacificamente, e de fazerem pedidos ao governo para que sejam feitas reparações de queixas.” Disponível em:

<http://www.uel.br/pessoal/jneto/gradua/historia/recdida/ConstituicaoEUARecDidaPESSOALJNETO.pdf>

⁴⁴ *Whitney v. California*. Disponível em: <https://www.oyez.org/cases/1900-1940/274us357>. Acesso em: 12 mai. 2023.

Dessa forma, até aquele momento a jurisprudência da Suprema Corte considerava que a legalidade do exercício da liberdade de expressão dependia da situação na qual ela é realizada, podendo sofrer limitações se caracterizada a existência de perigo claro e iminente.

Contudo, naquele mesmo ano, no caso *Abrams v. United States*⁴⁵, Holmes Jr. amplia a sua teoria indicando que o perigo deve, além de ser claro e iminente, ser muito grave, para justificar a restrição da liberdade de expressão. Na ocasião, Jacob Abrams e outros quatro imigrantes russos foram presos e condenados a 20 anos de prisão por violarem o chamado Ato de Sedição de 1918⁴⁶ ao distribuírem dois panfletos pelas ruas de Nova York: um que criticava o envio das tropas americanas para a Rússia e outro que incentivava os trabalhadores da indústria bélica a cessarem a produção de armas em protesto contra a Primeira Guerra.

Diversamente da sua decisão anterior, Holmes Jr. divergiu dos demais membros da corte, argumentando que a Primeira Emenda protege o direito de discordar dos pontos de vista e objetivos do governo e que, nesse caso, não foi configurado o perigo muito grave (LAURENTIIS; THOMAZINI, 2020)⁴⁷.

Assim, foi apenas no caso *Brandenburg vs. Ohio*⁴⁸ que a Suprema Corte consolidou a jurisprudência em vigor até hoje, definindo parâmetros para determinar se uma manifestação é meramente uma advocacia de ideias e, portanto, protegida pela liberdade de expressão, ou se configuraria um perigo claro e iminente, capaz de gerar a sua restrição, por meio da adoção do chamado “teste de Brandenburg” (MARTINS, 2022).

Um líder da Ku Klux Klan, Clarence Brandenburg, foi condenado pela Justiça de Ohio por divulgar filmes de natureza discriminatória. Os vídeos⁴⁹ retratavam

⁴⁵ *Abrams v. United States*. Disponível em: <https://www.oyez.org/cases/1900-1940/250us616>. Acesso em: 10 mai. 2023.

⁴⁶ O Ato de Sedição (*Seditious Act*) foi uma alteração na lei de espionagem e proibia críticas ao governo e sua participação na Guerra.

⁴⁷ Em sentido contrário, no caso *Whitney vs. California* (1927), a Suprema Corte Americana manteve a condenação de Charlotte Anita Whitney, membro fundadora do Partido Trabalhista Comunista da Califórnia, por ajudar a organizar um grupo que buscava efetuar mudanças econômicas e políticas por meio do uso ilegal da violência.

⁴⁸ *Brandenburg v. Ohio* (EUA, 1969). Disponível em: <https://www.oyez.org/cases/1968/492>. Acesso em: 10 mai. 2023.

⁴⁹ “Um filme mostrou 12 figuras encapuzadas, algumas das quais carregavam armas de fogo. Eles foram reunidos em torno de uma grande cruz de madeira, que queimaram. ninguém estava presente além dos participantes e dos jornalistas que fizeram o filme. A maioria das palavras ditas durante a cena eram incompreensíveis quando o filme foi projetado, mas frases esparsas podiam ser

comícios do clã que exigiam vingança contra judeus, negros e uma suposta opressão da raça branca. A decisão, contudo, foi revertida pela Suprema Corte Americana, que assim afirmou:

As liberdades de expressão e de imprensa não permitem que um Estado proíba a defesa do uso da força ou da violação da lei, exceto quando tal defesa for direcionada para incitar ou produzir ação ilegal iminente e for provável que incite ou produza tal ação (EUA, 1969, s/p, tradução nossa).

Nessa perspectiva, a partir do teste da ação ilegal iminente, ou teste de Brandenburg, o Estado não poderia restringir a liberdade de expressão apenas pela natureza do conteúdo da expressão – ainda que fosse um discurso de ódio (*hate speech*) ou falso (*fake news*) –, mas sim pelas consequências dele, desde que presentes os seguintes requisitos: a) o discurso fosse capaz de incitar ou produzir uma ação ilegal iminente (dano); b) que houvesse real probabilidade que a incitação fosse capaz de produzir essa ação, como uma espécie de nexo causal.

Nesse modelo, entende-se que a repressão ao discurso não traz estabilidade pública e paz social, pois o discurso proibido semeia o ódio e a reação e incentiva a conspiração. Somente um livre mercado de ideias poderia definir qual ideia deve prevalecer.

O Supremo Tribunal Federal já utilizou a expressão “livre mercado de ideias” em diversas decisões, contudo sem o alcance pretendido pela Corte Americana. O sentido empregado pelas decisões nacionais se refere à pluralidade de ideias e à diversidade, não à ausência de interferência estatal no conteúdo dos discursos, ainda

compreendidas depreciativas de negros e, em um caso, de judeus. Outra cena do mesmo filme mostrava o apelante, em trajes da Klan, fazendo um discurso. O discurso, na íntegra, foi o seguinte: ‘Esta é uma reunião de organizadores. Tivemos alguns membros aqui hoje que são - temos centenas, centenas de membros em todo o estado de Ohio. Posso citar um recorte de jornal do Columbus, Ohio, Dispatch, cinco semanas atrás, no domingo de manhã. A Klan tem mais membros no estado de Ohio do que qualquer outra organização. Não somos uma organização vingativa, mas se nosso presidente, nosso Congresso, nossa Suprema Corte continuar a reprimir a raça branca caucasiana, é possível que tenha que haver alguma vingança.’ ‘Estamos marchando no Congresso em 4 de julho, quatrocentos mil pessoas. A partir daí, estamos nos dividindo em dois grupos, um grupo para marchar em St. Augustine, Flórida, o outro grupo para marchar para o Mississippi. Obrigado.’ O segundo filme mostrava seis figuras encapuzadas, uma das quais, posteriormente identificada como o recorrente, repetia um discurso muito semelhante ao registrado no primeiro filme. A referência à possibilidade de ‘vingança’ foi omitida e uma frase foi acrescentada: ‘Pessoalmente, acredito que o negro deveria ser devolvido à África, o judeu deveria ser devolvido a Israel’. Embora algumas das figuras dos filmes carregassem armas, o orador não.” (EUA, 1969, s/p, tradução nossa).

que fundamentada na ideia de colisão de opiniões conflitantes numa busca pela verdade:

[...] 7. A colisão de opiniões conflitantes, sob o prisma jus-filosófico, reclama como premissa o denominado *free speech*, que amplia o esclarecimento público e as chances de atingimento da verdade decorrente da competição da ideia no livre mercado do pensamento (*marketplace of ideas*), consoante consagrado pela Suprema Corte Americana no memorável voto do Justice OLIVER HOLMES no caso *Abrams v. United States*, referido por JOHN NOWAK e RONALD ROTUNDA (In *Constitucional Law*. Fourth Edition. Saint Paul, West Publishing Co, 1991, págs 940 e seguintes). (BRASIL, 2011, p. 3).

Dessa forma, no modelo americano, a liberdade de expressão se apresenta em uma posição de superioridade em relação aos demais direitos fundamentais, sendo a neutralidade estatal um requisito indispensável para o seu exercício e não cabendo ao Estado valorar o conteúdo dos discursos. Assim, o discurso de ódio proferido pelo líder da Ku Klux Klan tem a mesma posição e deferência de um discurso pacífico, somente cabendo limitações em situações extraordinárias em que se analisa o contexto e as consequências reais do discurso, não o seu conteúdo.

Diferentemente do modelo americano, o modelo alemão e o brasileiro⁵⁰ permitem o controle, a valoração do discurso e uma maior restrição da liberdade de expressão quando em conflito com outros direitos fundamentais, especialmente os da personalidade, realizando um juízo de ponderação e empregando o princípio da proporcionalidade. Nesses modelos, a dignidade da pessoa humana é o valor máximo da ordem jurídica, e não a liberdade de expressão (SARMENTO, 2006).

No Brasil, dada a sua relevância para o exercício da democracia, a liberdade de expressão assume uma posição de preferência⁵¹, ainda que mitigada, se comparada ao modelo americano, quanto à resolução de conflitos com outros direitos

⁵⁰ Apesar de semelhantes, o modelo alemão e o brasileiro não são idênticos quanto as restrições à liberdade de expressão. Por exemplo, o Tribunal Constitucional Alemão, no caso *BVerfGE 35, 202 – Lebach* reconheceu o chamado “direito ao esquecimento” ao proibir a transmissão em rede de televisão de documentário sobre cidadão preso às vésperas de ser solto. Considerou-se que a divulgação poderia comprometer a ressocialização do indivíduo e que, em razão do transcurso do tempo, não havia interesse público significativo em divulgar os fatos. Nesse caso, o direito à informação foi limitado em face da privacidade do indivíduo. Já o Supremo Tribunal Federal brasileiro entendeu que o direito ao esquecimento é incompatível com a Constituição Federal e que eventuais excessos ou abusos no exercício da liberdade de expressão e de informação devem ser analisados caso a caso a partir dos parâmetros constitucionais, especialmente os relativos à proteção da honra, da imagem, da privacidade e da personalidade em geral, e as expressas e específicas previsões legais nos âmbitos penal e cível.

⁵¹ Essa tem sido a regra, em especial a partir do julgamento da ADPF 130, que não recepcionou a lei de imprensa editada durante o período da ditadura militar. Essa posição preferencial é reduzida quando comparada ao modelo americano. (ADPF 130).

fundamentais, sendo a vedação da censura (princípio da incensurabilidade) um dos seus pontos centrais (FARIAS, 2001).

A Constituição Federal consagra o princípio da incensurabilidade ao vedar toda e qualquer forma de censura, seja ela de natureza política, ideológica ou artística⁵², o que não impede a existência de limitações e restrições que justifiquem o controle do abuso da liberdade de expressão. Mas em qual medida uma restrição à liberdade de expressão pode ser considerada legítima ou configurar censura?

A conceituação de censura não é uma tarefa fácil. A censura pode ser descrita metaforicamente como a remoção da voz de um indivíduo ou grupo, o ato de silenciar ou impedir a expressão, criando um ambiente no qual as pessoas se sintam inibidas ou sejam efetivamente proibidas de manifestarem seus pontos de vista. (WARBURTON, 2020). É natural em regimes autoritários a inclinação a silenciar a difusão de ideias e informações contrários aos interesses dos governantes, criando-se um ambiente intolerante.

De forma mais pragmática, a censura pode ser definida em um sentido restrito e em um sentido amplo. Em sentido estrito, ela é a restrição prévia da liberdade de expressão, realizada por uma autoridade administrativa, que implica na proibição de veiculação de um determinado conteúdo – é um controle preventivo da expressão (MACHADO; BRITO, 2020). Para Daniel Sarmiento (2013), a censura prévia é o mais grave atentado à liberdade de expressão que se pode conceber porque é totalmente incompatível com um regime democrático, e eventuais abusos ou lesões devem ser punidos e compensados posteriormente⁵³.

A proibição à censura prévia encontra fundamento no art. 13.2 da Convenção Americana sobre Direitos Humanos (Pacto de San José da Costa Rica), com a ressalva de que a lei poderá realizar classificação etária para espetáculos públicos com o objetivo de proteger a infância e a adolescência:

2. O exercício do direito previsto no inciso precedente não pode estar sujeito a censura prévia, mas a responsabilidades ulteriores, que devem ser expressamente fixadas pela lei a ser necessárias para assegurar:
 - a) o respeito aos direitos ou à reputação das demais pessoas; ou

⁵² “Art. 5º, IX c/c art. 220 § 2º - é livre a expressão da atividade intelectual, artística, científica e de comunicação, independentemente de censura ou licença; Art. 220 § 2º É vedada toda e qualquer censura de natureza política, ideológica e artística.” (CF/88).

⁵³ Como por exemplo por meio do direito de resposta e das responsabilidades na esfera civil e penal.

b) a proteção da segurança nacional, da ordem pública, ou da saúde ou da moral públicas.

3. Não se pode restringir o direito de expressão por vias ou meios indiretos, tais como o abuso de controles oficiais ou particulares de papel de imprensa, de frequências radioelétricas ou de equipamentos e aparelhos usados na difusão de informação, nem por quaisquer outros meios destinados a obstar a comunicação e a circulação de ideias e opiniões.

4. A lei pode submeter os espetáculos públicos a censura prévia, com o objetivo exclusivo de regular o acesso a eles, para proteção moral da infância e da adolescência, sem prejuízo do disposto no inciso 2. (CIDH, 1969, s/p).

A ausência de censura é o que assegura o direito fundamental de não ser impedido de se expressar (status negativo da liberdade de expressão) e tem relevância especial para a liberdade de imprensa. O Supremo Tribunal Federal, na Rcl 18638/CE, assentou que “[...] não cabe ao Estado, por qualquer dos seus órgãos, definir previamente o que pode ou o que não pode ser dito por indivíduos e jornalistas.” Nesse sentido é a doutrina de William Blackstone:

"A liberdade de imprensa é, sem dúvida, essencial à natureza de um Estado livre; mas ela consiste em não impor restrições prévias às publicações e não na imunidade à censura por matérias criminosas quando publicadas. Todo cidadão tem o indubitável direito de expor os seus sentimentos que entender perante o público; proibir isto é destruir a liberdade de imprensa: mas se ele publicar o que é impróprio, nocivo ou ilegal, deve arrostar as consequências de sua própria temeridade" (BLACKSTONE, s/d, s/p *apud* FARIAS, 2001, p. 68).

Em sentido mais amplo, a censura abrangeria outras hipóteses de restrição indevida da liberdade de expressão: não apenas por autoridade administrativa, mas também por um particular, por decisões judiciais e até mesmo por atos legislativos, bem como restrições indevidas ocorridas *a posteriori*, após a manifestação (SARMENTO, 2013). Um exemplo de censura judicial destacada por Sarmento (2013) são as decisões judiciais que proibiam a publicação de biografias de personalidades sem autorização do biografado ou de seus herdeiros, motivo pelo qual o Supremo Tribunal Federal⁵⁴ entendeu tal exigência como inconstitucional.

Observa-se que a proibição da censura também opera efeitos de forma horizontal, vinculando particulares na medida em que não podem se valer do seu poder social para impedir a veiculação de ideias e informações (SARMENTO, 2013).

⁵⁴ “[...] Autorização prévia para biografia constitui censura prévia particular. O recolhimento de obras é censura judicial, a substituir a administrativa. O risco é próprio do viver. Erros corrigem-se segundo o direito, não se coartando liberdades conquistadas. A reparação de danos e o direito de resposta devem ser exercidos nos termos da lei.” (ADI 4815/DF).

Essa definição é especialmente válida para se definir um limite à atuação das plataformas de redes sociais, que ficam proibidas de realizar atos de censura por meio da moderação de conteúdo.

O princípio da incensurabilidade, contudo, não se apresenta como uma imunidade absoluta da liberdade de expressão, devendo ser considerado como censura a restrição arbitrária da liberdade de expressão, seja ela imposta pelo poder público ou privado, prévia ou posterior. Dessa forma, não devem ser consideradas como arbitrárias, logo censuras, aquelas limitações externas ou internas da liberdade de expressão quando autorizadas direta ou indiretamente pelo texto constitucional, que devem ser analisadas casuisticamente à luz dos princípios da razoabilidade e proporcionalidade.

Por limites externos, considera-se as restrições decorrentes das normas jurídicas quando em confronto com outros direitos fundamentais, por exemplo os crimes contra a honra; por limites internos ou imanentes consideram-se aqueles decorrentes do controle do conteúdo do discurso, tais como os discursos de ódio (*hate speech*) e as notícias falsas ou desinformações (*fake news*) (MARTINS, 2022).

Em outras palavras, a liberdade de expressão assegurada no cenário nacional é a liberdade com responsabilidade, na medida em que não está protegida a expressão que viola a privacidade e divulga conteúdos sigilosos do Estado, conteúdo erótico para crianças, propagandas falsas, discursos de ódio ou desinformação, entre inúmeras outras limitações.

Dentre os limites apresentados, dois fenômenos, que não são recentes, mas tomaram dimensões exponenciais com as redes sociais, merecem maior atenção: a desinformação (*fake news*) e a propagação de discursos de ódio⁵⁵ (*hate speech*). Optou-se por aprofundar esses fenômenos pois são os que estão em pauta na problemática envolvendo a regulação da liberdade de expressão nas redes sociais.

Apesar da neutralidade estatal ser um dos elementos mais importantes para a defesa da liberdade de expressão, discursos de ódio não podem ser tolerados. A Convenção Americana autoriza a limitação legal do discurso de ódio em seu art. 13.5: “A lei deve proibir toda propaganda a favor da guerra, bem como toda apologia ao

⁵⁵ O discurso de ódio, pelo seu caráter de incitar a discriminação e a violência, na maioria das vezes pode ser enquadrado como crime (racismo, por exemplo), mas ainda que não caracterizado o crime pode produzir consequências civis em decorrência da violação da honra, imagem e privacidade. (MARTINS, 2022).

ódio nacional, racial ou religioso que constitua incitação à discriminação, à hostilidade, ao crime ou à violência.” (CIDH, 1969, s/p).

O discurso de ódio é aquele que provoca extrema ofensa apresentada com o intuito de vilipendiar ou de incitar o ódio, a violência e a discriminação contra pessoas ou grupos que normalmente possuem certa vulnerabilidade e fazem parte de uma minoria. Ele não se confunde com o discurso preconceituoso, porque não se trata de uma manifestação de uma ideia detestável ou repulsiva, ainda que de forma pedante ou prepotente, mas de atos de extremo insulto apresentados de forma intencional, visando à discriminação.

Norberto Bobbio (2011) apresenta uma distinção entre preconceito⁵⁶ e discriminação: o preconceito é aquela opinião ou um conjunto de opiniões que se aceita sem discussão, de forma acrítica, porque ele se encontraria na esfera do não racional, de modo que não consegue ser simplesmente refutado com base em argumentos racionais. Justamente por não se submeter ao controle da razão, não se confunde com uma opinião meramente equivocada, que pode ser corrigida por um melhor conhecimento. Dessa forma, “o preconceito é uma opinião errônea tomada fortemente por verdadeira, mas nem toda opinião errônea pode ser considerada um preconceito.” (BOBBIO, 2011, p. 103).

A principal consequência do preconceito é a discriminação, que pressupõe uma distinção entre indivíduos de forma pejorativa. A discriminação se caracteriza, na visão de Bobbio (2011), quando preenchidas, de forma cumulativa, três etapas: a primeira, de natureza cognitiva; a segunda, com caráter valorativo; e a terceira, que estabelece relação de dominância.

A primeira etapa, de caráter cognitivo, atesta a desigualdade entre grupos ou indivíduos e se trata de um juízo de fato, pois a mera constatação que indivíduos são diferentes não gera ainda um juízo discriminante. A segunda etapa tem caráter valorativo e trata-se de um juízo de valor, implicando numa relação de superioridade de um grupo em detrimento de outro (um grupo é bom e o outro é mau). Já a terceira

⁵⁶ “Entende-se por ‘preconceito’ uma opinião ou um conjunto de opiniões, às vezes até mesmo uma doutrina completa, que é acolhida acrítica e passivamente pela tradição, pelo costume ou por uma autoridade de quem aceitamos as ordens sem discussão ‘acriticamente’ e ‘passivamente’, na medida em que a aceitamos sem verificá-la, por inércia, respeito ou temor, e a aceitamos com tanta força que resiste a qualquer refutação racional, vale dizer, a qualquer refutação feita com base em argumentos racionais.” (BOBBIO, 2011, p. 103).

etapa, decisiva para a caracterização da discriminação, implica na concepção de que o superior tem o direito de dominar, explorar, escravizar, eliminar, suprimir ou reduzir direitos fundamentais do diferente, que ele compreende como inferior.

Para Bobbio (2011), apenas quando preenchidas essas três etapas⁵⁷ – relação de diversidade, superioridade e dominação – é que se pode conceber uma verdadeira discriminação com todas as consequências jurídicas que se possam atribuí-la. Pode-se dar como exemplo a discriminação contra judeus pelos nazistas: primeiro, atestaram que judeus eram diferentes de arianos; segundo, que arianos eram uma raça superior; e terceiro, que os arianos (superiores) deveriam dominar e eliminar os judeus (inferiores).

Nessa linha, o Supremo Tribunal Federal, no julgamento da ADO 26, que criminalizou a homofobia, destacou que o discurso religioso preconceituoso não se enquadra no conceito de discurso de ódio, assegurando aos fiéis e ministros o direito de pregar e de externar as suas convicções de acordo com os seus livros e códigos sagrados, ainda que possam ter conotação preconceituosa ou segregacionista, “desde que tais manifestações não configurem discurso de ódio, assim entendidas aquelas exteriorizações que incitem a discriminação, a hostilidade ou a violência contra pessoas em razão de sua orientação sexual ou de sua identidade de gênero.” (BRASIL, 2019a).

Pela decisão destacada, observa-se que, enquanto o discurso religioso se mantiver apenas nas primeiras etapas da discriminação elencadas por Bobbio, ele deve ser tolerado em respeito à liberdade religiosa. Contudo, quando o discurso preencher a terceira etapa, implicando na dominação, supressão, exclusão ou violação de direitos, estará configurado o discurso de ódio, que não deve ser tolerado, mas sim repellido e, se for o caso, criminalizado⁵⁸.

Um caso paradigmático que serve de parâmetro para as questões envolvendo o discurso de ódio é o caso *Elwanger*⁵⁹. Na ocasião, o Supremo Tribunal Federal

⁵⁷ Essas etapas foram utilizadas pelo Supremo Tribunal Federal no julgamento do RHC 146303/RJ, de relatoria do Ministro Edson Fachin, para caracterização do crime de racismo cometido por líder religioso. Na ocasião, o Supremo entendeu que estariam preenchidas as três etapas da discriminação de Bobbio e, portanto, configurado o crime de racismo. (RHC 146303/RJ).

⁵⁸ Destaca-se a decisão do Supremo Tribunal Federal no HC 134.682, que não considerou como crime a manifestação de um padre católico que pregava contra o espiritismo, uma vez que, apesar de comprovada a relação de diversidade e superioridade, não foi caracterizada a dominação, violência ou exclusão. (HC 134.682).

⁵⁹ HC 82424 / RS.

entendeu que a publicação de um livro antissemita configuraria crime de racismo e que a liberdade de expressão não compreende a incitação ao racismo.

Como visto, o ordenamento jurídico brasileiro reconhece a possibilidade de controle sobre o conteúdo dos discursos, na linha do modelo alemão, afastando do âmbito de proteção da liberdade de expressão os discursos de ódio, que devem ser entendidos como aqueles que promovem cumulativamente a diferença, superioridade e dominação entre indivíduos.

Outro limite à liberdade de expressão que é objeto de debate na doutrina e na jurisprudência nacional e internacional⁶⁰ são as chamadas *fake news* (notícias falsas ou desinformação), que acabam por extrapolar os limites da liberdade de expressão, em especial no que tange ao direito à informação.

Não há um consenso técnico sobre a definição do termo *fake news*, o qual vem sendo usado indistintamente de inúmeras formas como sinônimo de sátira, propaganda, manipulação, imitação, publicidade, fabricação (RIEMSDIJK *et al.*, 2020). Nesse estudo, adotou-se por definição um conceito mais amplo sobre o tema, proposto por Wardle e Derakhshan (2017).

Sabe-se que a convivência harmônica em sociedade depende muito do que se diz, mas especialmente de como se diz. Isso porque uma simples estória pode apresentar inúmeros vieses e nesse caminho, entre aquele que fala e o indivíduo que recebe a informação, há possíveis falhas ou distúrbios de comunicação que podem provocar desentendimentos, conflitos, crises familiares e até mesmo definir uma eleição presidencial.

Para Wardle e Derakhshan (2017), há três tipos de distúrbios de informação: a informação incorreta; a desinformação; e a má-informação. Essa classificação leva em consideração duas variáveis: a veracidade (se a informação propagada é verdadeira ou não) e a intencionalidade (se a sua propagação tem a intenção de causar danos ou não).

A informação incorreta (*mis-information*) é aquela falsa, que contém erros não intencionais. A falsidade na informação não foi criada com a intenção de causar danos,

⁶⁰ A expressão *fake news* foi escolhida em 2017 como a palavra do ano pelo dicionário Collins English. Além disso, a Declaração Conjunta da Relatoria Especial da OEA de 2021 sobre políticos e autoridades públicas e liberdade de expressão estabelece recomendação aos Estados, partidos e lideranças políticas para que adotem medidas para combater a disseminação de desinformação e notícias falsas. (OEA, 2021).

ainda que possa⁶¹. Por outro lado, a desinformação (*dis-information*) é aquela informação falsa deliberadamente criada por uma pessoa ou grupo para causar danos. Nesse caso, a informação é manipulada, fabricada e deturpada com a clara intenção de prejudicar alguém ou obter vantagens. Aqui têm-se as *fake news* propriamente ditas. Já a má-informação (*mal-information*) é aquela informação verdadeira, baseada na realidade, mas privada e compartilhada de forma não autorizada, com a intenção de causar danos: informações “vazadas” (WARDLE; DERAKHSHAN, 2017).

Independentemente da intencionalidade ou veracidade, qualquer distúrbio de informação pode gerar consequências gravíssimas. Destaca-se o famoso caso Orson Welles: no dia 30 de outubro de 1938, véspera de Halloween, um programa de rádio interrompeu sua programação musical para noticiar uma suposta invasão de marcianos. Tratava-se, na verdade, de uma peça de radioteatro do livro de ficção científica *A Guerra dos Mundos*, do escritor inglês Herbert George Wells, produzida então pelo diretor de cinema Orson Welles – que não é parente do escritor inglês.

Apesar de ter sido informado no início da peça que se tratava de uma dramatização, muitos ouvintes acompanharam a transmissão após o seu início, o que causou um pânico generalizado. Estima-se que 1,2 milhão de pessoas acreditaram ser real a invasão alienígena; houve fuga em massa e reações desesperadas dos moradores⁶². A “falha” na comunicação provocou consequências desastrosas e que poderiam ter sido evitadas⁶³.

⁶¹ Um bom exemplo para ilustrar a informação incorreta é a brincadeira de infância chamada “telefone sem fio” (os mais vetustos devem se recordar). Nessa brincadeira, uma pessoa fala uma palavra ou frase, às vezes um segredo, ao ouvido de outra pessoa ao seu lado. Aquele que ouviu deve então repeti-lo para o próximo participante e assim por diante até chegar ao último, que deve contar o segredo em voz alta. Por muitas vezes, a informação chega deturpada no final. Essa informação falsa, que não corresponde com a primeira palavra ou frase dita, não fora produzida com a intenção de causar mal, mas por algum motivo apresentou essa dissonância com a realidade.

⁶² DW. Disponível em: <https://www.dw.com/pt-br/1938-p%C3%A2nico-ap%C3%B3s-transmiss%C3%A3o-de-guerra-dos-mundos/a-956037>. Acesso em: 13 mai. 2023.

⁶³ Outro caso que merece destaque é a desinformação proliferada no âmbito da vacinação. Em 1997, o médico londrino Andrew Wakefield publicou um artigo no prestigiado jornal médico *The Lancet* defendendo uma relação causal entre as vacinas do sarampo e rubéola com o autismo. Após a divulgação da pesquisa, as taxas de cobertura da vacina tríplice viral caíram expressivamente, o que desencadeou um surto de sarampo (BUTANTAN, 2021). Desde então, diversos estudos já foram publicados refutando a existência de qualquer relação entre a vacina e o autismo, contudo os argumentos defendidos por Wakefield ainda fundamentam movimentos antivacinas (POLÍGRAFO, 2019).

O debate é mais atual do que nunca, pois a epidemia de desinformação sobre o coronavírus, que inclusive já foi batizada pela Organização Mundial de Saúde (OMS) como uma infodemia⁶⁴ (PIERRO, 2020), é um potencial risco à saúde pública, além do próprio vírus. Além das informações sabidamente falsas, o volume de informações vagas e imprecisas sobre a COVID-19 têm se espalhado mais do que o próprio vírus, gerando confusão, medo e instabilidade, sendo um campo fértil para a proliferação da desinformação, em especial para os movimentos antivacina.

Assim, adota-se aqui o conceito de *fake news* como sinônimo de “desinformação”, ou seja, aquela informação falsa deliberadamente criada por uma pessoa ou grupo para causar danos ou obter vantagens, não se confundindo com a informação incorreta ou com a má informação. Por “desinformação” também se considera aquela informação verdadeira, mas intencionalmente utilizada fora de contexto e de forma tendenciosa, tornando-se então falsa pois gera interpretações equivocadas. Ademais, são essas as informações com maior capacidade de proliferação, pois as que possuam certo lastro de verdade têm mais probabilidade de serem consideradas verdadeiras e, por consequência, compartilhadas (WARDLE, 2020).

E é nesse cenário de desinformação e necessidade de identificar se uma informação é falsa ou verdadeira que vem se construindo, na doutrina e na jurisprudência, a verdade como um limite interno à liberdade de expressão, na perspectiva de um direito à informação verídica (MENDES; BRANCO, 2022). As perguntas que este trabalho pretende sucintamente responder são as seguintes: a disseminação de informações falsas está compreendida no âmbito da liberdade de expressão? Caso não esteja, como preservar os fatos da falsidade sem prejuízo à liberdade de expressão?

A primeira pergunta aparenta já ter uma resposta consolidada no âmbito da doutrina e da jurisprudência nacional. O Supremo Tribunal Federal⁶⁵, na TPA 39 MC-

⁶⁴ Infodemia se refere ao excesso de informações, algumas precisas e outras não, que tornam difícil encontrar fontes idôneas e orientações confiáveis quando se precisa (OMS, 2020).

⁶⁵ Destaca-se a decisão do Tribunal Superior Eleitoral – TSE que determinou a retirada de vídeos sobre o “kit gay”, por estar gerando desinformação especialmente no período eleitoral. Na ocasião, o ex-presidente Jair Bolsonaro – na época candidato à presidência –, apresentava em suas redes sociais um vídeo que afirmava estar sendo distribuído nas escolas públicas. A Corte entendeu que se tratava de fato sabidamente inverídico, uma vez que o livro mencionado não integra a base de livros didáticos e recomendados pelo Governo Federal. (TSE - 0601699-41.2018.6.00.0000 – Relator: Ministro Carlos Horbach, 16.10.2018).

Ref/DF⁶⁶, julgada em 07 de junho de 2022, reforçou seu entendimento de que não há liberdade de expressão, nem mesmo a imunidade parlamentar, que ampare a disseminação de informações falsas por redes sociais e na internet. A questão discutia a veiculação de notícias falsas a respeito da lisura do sistema eletrônico de votação.

A qualidade do diálogo em um ambiente de propagação de notícias falsas é deteriorada na medida em que os indivíduos formam as suas convicções a partir de ilusões e de engano. A dificuldade de discernir o real do irreal gera um espaço de desconfiança e incerteza, amplia a intolerância, prejudica a capacidade de autonomia e o livre desenvolvimento do indivíduo e coloca em risco o pluralismo e a própria democracia. Assim, não restam dúvidas quanto à possibilidade de limitar a liberdade de expressão quando se trata de informações falsas. Contudo, resposta não tão objetiva é a de se definir em qual perspectiva a verdade deve ser analisada sem escorregar em um ato de censura.

Aqui é importante fazer uma distinção entre fato e opinião, ou entre juízo de fato e juízo sobre o fato. Para Bruno Dantas, o juízo de fato ou o fato é objetivo, descritivo, relacionado a um processo ontológico-cognitivo. Ele é a descrição factual e é sobre esse juízo que repousa a busca da verdade, relacionado diretamente com o direito à informação. Já o juízo sobre o fato ou opinião é interpretativo e decorre de um juízo de valor; é nesse campo que deve se enquadrar a liberdade de expressão (DANTAS; SANTOS, 2020). Segundo José Joaquim Gomes Canotilho,

Enquanto os fatos são susceptíveis de prova da verdade, as opiniões ou juízos de valor, devido à sua própria natureza abstrata, não podem ser submetidos à comprovação. Resulta que a liberdade de expressão tem o âmbito de proteção mais amplo do que o direito à informação, uma vez que aquela não está sujeita, no seu exercício, ao limite interno da veracidade, aplicável a este último. (CANOTILHO, 1998, s/p *apud* TESTA JUNIOR, 2009, p. 8541).

Nesse sentido, o Tribunal Constitucional Alemão⁶⁷ já assentou que não se pode confundir opinião e fato e destacou que afirmar que a Alemanha não deu causa à Segunda Guerra Mundial seria uma opinião controversa, porém negar ou minimizar o holocausto seria negar uma afirmação sobre os fatos, o que é inconcebível (CAVALCANTE FILHO, 2014).

⁶⁶ Referendo na medida cautelar na tutela provisória antecedente 39/DF.

⁶⁷ BVerfGE 90, 1, conhecido como “Caso Fabricação Histórica” 158.

Observa-se que a negação do holocausto não se encontra sob a proteção da liberdade de expressão, pois não se trata de uma opinião, e sim a negativa de um fato já exaustivamente comprovado. É sobre esse juízo de fato que é autorizado falar a respeito da existência de uma verdade e a necessidade de impedir a manipulação sobre a realidade dos fatos.

Nesse contexto, as opiniões propriamente ditas devem ser livres, uma vez que não se pode afirmar sem dúvidas a veracidade ou falsidade do juízo interpretativo, diferentemente do que ocorre com a afirmação dos fatos, que se submetem a um juízo sobre a veracidade ou não do fato afirmado.

A verdade, objeto da epistemologia, assume papel de destaque no Direito porque a busca pela verdade, de forma dogmática, em um conflito judicial é o que legitima a justiça das decisões, de modo que a verdade deve ser a inspiração da justiça. Como destaca Michele Taruffo,

Si además se considera, como se ha dicho hace un momento, que la justicia de la decisión final depende también —necesariamente— de la determinación de la verdad de los hechos, se infiere que un proceso no es justo si no está estructuralmente orientado a la búsqueda y al descubrimiento de la verdad. (TARUFFO, 2019, p. 287⁶⁸).

Para Byung-Chul Han (2022), o desaparecimento da verdade factual implica no desaparecimento do mundo comum em si, uma vez que a verdade é um regulador social que mantém a estrutura da sociedade. Já para Hannah Arendt (1967), não há liberdade de expressão se não houver um consenso mínimo sobre os fatos, uma verdade factual, pois os fatos é que formam opiniões, enquanto afirmações inverídicas em nada contribuem para a formação da opinião pública, antes afastam e alienam o receptor da informação:

Ainda que se deva distingui-los, os factos e as opiniões não se opõem uns aos outros, pertencem ao mesmo domínio. Os factos são a matéria das opiniões, e as opiniões, inspiradas por diferentes interesses e diferentes paixões, podem diferir largamente e permanecer legítimas enquanto respeitarem a verdade de facto. A liberdade de opinião é uma farsa se a informação sobre os factos não estiver garantida e se não forem os próprios factos o objecto do debate. Por outras palavras, a verdade de facto fornece

⁶⁸ “Se se considerar também, como se disse há pouco, que a justiça da decisão final também depende —necessariamente— da determinação da verdade dos fatos, segue-se que um processo não é justo se não for estruturalmente orientado para a busca e a descoberta da verdade.” (TARUFFO, 2008, p. 388, tradução nossa).

informações ao pensamento político tal como a verdade racional fornece as suas à especulação filosófica (ARENDR, 1967, p. 11).

A busca da verdade factual, no que tange a liberdade de expressão, deve ser compreendida, portanto, como a adequação racional entre uma determinada expressão e a realidade dos fatos políticos, econômicos e sociais (MACHADO; BRITO, 2020). Dessa forma, não há qualquer incompatibilidade entre a liberdade de expressão e o combate à desinformação, porque a verdade como limitadora da expressão não está no campo do juízo de opinião, mas antes no plano da descrição, do necessário consenso sobre os fatos, como condição político-existencial para garantir a genuína formação e exercício da opinião.

Sendo assim, no cenário nacional, observa-se que a liberdade de expressão apresenta limites que podem ser extraídos do próprio texto constitucional, especialmente quando em conflito com outros direitos fundamentais, não se apresentando como uma licença absoluta para expor a intimidade ou privacidade das pessoas, difamar, propagar conteúdo falso ou inapropriado para crianças ou revelar sigilos do Estado. Ademais, é possível o controle não apenas das consequências do discurso, mas também do seu conteúdo, a fim de evitar a violação da dignidade humana.

O futuro da liberdade de expressão e, por consequência, da própria democracia, está diretamente atrelado à maneira como os estados autorizam os indivíduos a usarem a internet, em especial as redes sociais, com limites definidos aos usuários, às plataformas e aos próprios governantes por meio da regulação.

2 LIBERDADE DE EXPRESSÃO NAS REDES SOCIAIS

Apesar de o surgimento da internet e das plataformas de redes sociais não ter tido o condão de alterar o âmbito, a titularidade, os sentidos e os limites da liberdade de expressão que foram analisados, isso provocou uma mudança no contexto social em que essa liberdade é exercida. Compreender essa mudança social é fundamental para que se possa debater a necessidade de uma mudança regulatória da liberdade de expressão.

Por isso, este capítulo tem como objetivo apresentar essa transição do espaço de expressão do ambiente físico para o virtual, do espaço considerado privado que se apresenta como uma nova esfera pública. Isso será feito conceituando as redes sociais, bem como pontuando alguns desafios que surgiram ao exercício da liberdade de expressão no meio digital, demonstrando, ao final, como essa mudança social aponta para a necessidade de atualizações e ajustes do modelo regulatório.

2.1 TRANSIÇÃO DA LIBERDADE DE EXPRESSÃO DO MEIO FÍSICO PARA O VIRTUAL: A ESFERA PÚBLICA DIGITAL

As novas tecnologias, em especial a internet, ofereceram aos seres humanos maneiras novas de realizar diversas tarefas, o que acabou gerando um impacto significativo, senão disruptivo, na forma como a sociedade passou a pensar e se relacionar em diversos sentidos, sejam eles sociais, culturais, econômicos ou políticos. Esse impacto foi acelerado pelo surgimento da inteligência artificial, baseada no aprendizado de máquina. A chamada Quarta Revolução Industrial, em que se vive atualmente, é marcada pela convergência entre o mundo real e o tecnológico (artificial) por meio de inovações de tecnologias digitais, físicas e biológicas. (SCHWAB, 2016).

Para Wolfgang Hoffmann-Riem (2022), a digitalização e a transformação digital provocarão mudanças sociais tão sérias ou ainda maiores que a invenção tipográfica e a industrialização. O termo “digitalização” aqui empregado se refere tanto à criação das tecnologias da informação quanto às mudanças estruturais nas condições de vida desencadeadas pela sua utilização.

O uso da internet tem se mostrado um instrumento indispensável para a realização de direitos fundamentais básicos, como a liberdade de expressão, a liberdade religiosa e o direito à educação, além de atingir o exercício de inúmeros outros direitos, em maior ou menor grau. Em especial, as redes sociais têm sido utilizadas como verdadeiros espaços públicos para o exercício da liberdade de expressão, como uma nova ágora mundial.

Se no passado os espaços de manifestação do cidadão comum eram predominantemente públicos (ruas, praças), hoje são majoritariamente privados (Facebook, Instagram, Twitter, Youtube). Essa migração, do espaço físico público para o virtual privado, se deu especialmente pelo surgimento de uma nova geração da internet denominada *web 2.0*.

A *World Wide Web*, como a maior construção de informação, teve muito progresso desde o seu advento. No seu início, a chamada *web 1.0* (1993 – 2003) surgiu como um espaço de compartilhamento de informações e no qual as empresas transmitiam informações para as pessoas a qualquer momento. Contudo, a produção de conteúdo, logo a expressão, ainda permanecia concentrada nas mãos daqueles que detinham a infraestrutura da internet, ou seja, dos programadores da web, à semelhança das mídias tradicionais. Os indivíduos em geral utilizavam a *web 1.0* apenas para leitura, os sites não eram interativos e sim estáticos, e não havia possibilidade de interação e colaboração pelos usuários (AGHAEI; NEMATBAKHSH; FARSANI, 2012).

Essa realidade mudou à medida em que a *web 1.0* transitou para a *web 2.0*, que surgiu como um espaço de conexão entre as pessoas e não apenas de transmissão de informação. Essa segunda geração de serviços baseados na estrutura da internet emerge como plataforma que cria um espaço colaborativo e empodera os usuários, que passam a poder também produzir conteúdo por meio da interação de outros usuários que contribuem em uma inteligência coletiva.

Para Tim O'Reilly, a *web 2.0* se diferenciou por ter criado conexões em rede, que evoluem à medida em que são utilizadas:

Web 2.0 is the business revolution in the computer industry caused by the move to the internet as platform, and an attempt to understand the rules for success on that new platform. Chief among those rules is this: Build

applications that harness network effects to get better the more people use them. (O'REILLY, 2006, s/p⁶⁹).

É nesse contexto colaborativo, participativo e de abertura que surgem as redes sociais, e a partir da análise do seu surgimento pode-se traçar o seguinte conceito: as redes sociais são plataformas da internet, empresas privadas que oferecem serviços online. Elas permitem que os usuários tenham um perfil pessoal ou profissional, gerem conteúdo que seja acessível aos seus seguidores ou a terceiros, e ainda permitem a formação de conexões entre diversos perfis para compartilhamento dos conteúdos produzidos, por meio de diferentes formatos (textos, imagem, vídeos, som), a depender de estruturas pré-definidas pela plataforma. Essas plataformas ainda hospedam, organizam, controlam, moderam e fazem circular os conteúdos produzidos e compartilhados, os quais podem ser monetizados por anúncios.

Esse conceito está alinhado⁷⁰ com o previsto no Projeto de Lei 2630/2020, que institui a Lei Brasileira de Liberdade, Responsabilidade e Transparência na Internet:

X - rede social: aplicação de internet que realiza a conexão entre si de usuários permitindo a comunicação, o compartilhamento e a disseminação de conteúdo em um mesmo sistema de informação, através de contas conectadas ou acessíveis entre si de forma articulada. (BRASIL, 2020, s/p).

Delimitando ainda mais o espectro da pesquisa, não serão analisadas aquelas plataformas que não realizam moderação de conteúdo, entendido como o controle do que pode ser dito e expressado de forma pública. Dessa maneira, serviços de troca de mensagens privadas, ainda que sujeitas a regulamentações internas, não serão objeto deste trabalho, na medida em que não oferecem uma visibilidade pública e, portanto, um impacto direto no exercício da liberdade de expressão.

⁶⁹ “Web 2.0 é a revolução de negócios na indústria de computadores causada pela mudança para a internet como plataforma e uma tentativa de entender as regras para o sucesso nessa nova plataforma. A principal dessas regras é esta: crie aplicativos que aproveitem os efeitos de rede para melhorar à medida que mais pessoas os usam.” (O'REILLY, 2006, s/p, tradução nossa).

⁷⁰ Esse conceito está alinhado também com o definido no documento *Padrões para uma regulação democrática das grandes plataformas*, que assim estabelece: “os provedores de serviços online quando atuam como intermediários ou plataformas de armazenamento, busca ou troca de informações, opiniões, expressões e outros conteúdos gerados por seus usuários e que realizam algum tipo de curadoria ou moderação desses conteúdos (“plataformas de conteúdo”). Entre eles: redes sociais, mecanismos de busca e outras plataformas para troca de textos, imagens e vídeos.” (PALLERO *et al.*, 2019, p. 7).

É importante ressaltar que as redes sociais, apesar de criarem espaços para a circulação de informações e conexão entre os usuários, amadureceram principalmente pela criação de um novo modelo de negócio exercido por empresas privadas que visam o lucro e tem como finalidade precípua manter os indivíduos conectados, chamando a atenção dos seus usuários com a finalidade de publicidade. Este novo modelo acabou prevalecendo diante de uma internet que passou a produzir mais conteúdo do que a capacidade de seus usuários de consumi-los. Afinal, se todos podem se tornar produtores de conteúdo, quem terá capacidade para consumi-los?

Isso porque, inobstante o acesso às redes sociais ser gratuito, as plataformas, enquanto empresas privadas que visam lucro, o obtém a partir de anúncios e impulsionamento de contas e publicações, de modo que a atenção dos usuários é vendida aos anunciantes⁷¹. Logo, em verdade, os anunciantes, e não os usuários, são os principais clientes dessas plataformas. Essa foi a solução encontrada para manter a rentabilidade e lucro do negócio e ao mesmo tempo organizar a priorização dos conteúdos que devem ser exibidos para cada usuário: exibe-se o conteúdo com a melhor capacidade de reter a atenção e de gerar engajamento.

É certo que os veículos de comunicação social tradicionais também se utilizam de anúncios, ou seja, da atenção dos telespectadores, ouvintes e leitores para a manutenção dos seus empreendimentos. Contudo, nesse novo modelo de negócio, as redes sociais se mostram muito mais eficientes no seu intento, por possuírem um potencial viciante muito maior e em especial pela facilidade de acesso e pela geração de uma gigantesca quantidade de dados que revelam costumes, desejos, intenções e interesses de seus usuários.

A portabilidade do acesso, por meio dos *smartphones*, e o avanço tecnológico que provocou sua redução de custos ampliaram a inclusão digital e revolucionaram a produção de conteúdo. A comunicação na *web 2.0* passou a se dar mais por manifestações instantâneas, como fotos e vídeos, do que textos. Os *blogs* foram substituídos pelos poucos caracteres do Twitter e os *foto blogs* foram centralizados no Instagram (PIAIA, RITTER; SANGOI, 2018).

Assim, as redes sociais revolucionaram o mercado publicitário, uma vez que o acesso aos milhões de dados, bem como a funcionalidade dos algoritmos, permite o

⁷¹ Nas palavras do jornalista Andrew Lewis, no documentário *O Dilema das Redes* (2020), “se você não paga pelo produto, o produto é você”.

direcionamento da publicidade de forma muito mais eficiente, especialmente pela visibilidade que os produtos alcançam e pela disponibilidade de tempo que os usuários despendem nas redes.

Estima-se que em janeiro de 2023 havia 5,16 bilhões de usuários da internet em todo o mundo, sendo que 4,76 bilhões deles são usuários das redes sociais – o equivalente a 59,4% da população mundial. E mais, 9 a cada 10 pessoas conectadas à internet usam alguma rede social⁷².

Já o Brasil, também em janeiro de 2023, contava com 152,4 milhões de usuários nas redes sociais, o equivalente a 70,6% de sua população total⁷³. De acordo com um estudo que analisou hábitos digitais, cada usuário fica em média 2 horas e 30 minutos por dia nas redes sociais, sendo que em 2020 o brasileiro passava em média 3 horas e 31 minutos por dia online⁷⁴.

Para Anna Lembke (2021), há uma necessidade biológica do ser humano de se conectar com outras pessoas. Quando isso ocorre a contento, desencadeia a liberação de dopamina, de modo que as redes sociais, ao manterem os indivíduos conectados, oferece dopamina digital 24 horas por dia. Assim, o cérebro compreende como recompensa cada *like* ou comentário recebido, o que desencadeia um círculo vicioso, prendendo a atenção nas redes.

Essa disponibilidade de tempo que se dispende nas redes, associada ao fornecimento dos dados, permite que as plataformas tracem previsões muito mais assertivas, com maior certeza de que o anúncio vai atrair a atenção do destinatário a partir da definição de um perfil de usuário que é vendido para os anunciantes, e é por isso que o modelo se apresenta tão lucrativo.

Isso é o que se tem denominado de Capitalismo de Vigilância e se refere à nova ordem econômica que “reivindica de maneira unilateral a experiência humana como matéria-prima gratuita para a tradução em dados comportamentais” (ZUBOFF, 2021, p. 18), os quais serão usados para práticas comerciais dissimuladas de extração,

⁷² O estudo ainda mostra um aumento significativo no uso das redes sociais durante a pandemia da COVID-19. Disponível em: <https://datareportal.com/social-media-users>. Acesso em: 12 mai. 2023.

⁷³ Disponível em: <https://datareportal.com/reports/digital-2023-brazil>. Acesso em: 12 mai. 2023.

⁷⁴ Disponível em: https://datareportal.com/reports/digital-2023-deep-dive-time-spent-on-social-media?utm_source=Global_Digital_Reports&utm_medium=Partner_Article&utm_campaign=Digital_2023. Acesso em: 12 mai. 2023.

previsão e venda desse comportamento, uma vez que tudo o que se faz na internet está sendo analisado, vigiado e monitorado.

O conhecimento gerado pelos algoritmos e pelo processamento do *Big Data* alteraram muito o ambiente da comunicação social e aquilo que se considera como esfera pública (VESTING, 2020). Antes da internet, a opinião pública era formada, além das universidades, quase que exclusivamente pelas mídias de massa tradicionais – imprensa escrita, rádio e televisão. Logo, a informação estava concentrada em um conglomerado midiático que acompanhava e controlava a informação a partir de critérios jornalísticos, prezando pela neutralidade e imparcialidade (DANTAS; SANTOS, 2020).

Essa concentração foi radicalmente alterada pelo surgimento da internet e especialmente pelas redes sociais, cujo alcance foi se tornando maior do que a de veículos tradicionais de comunicação social, aos quais se assemelham pela massificação e se diferem por serem mais individuais, diversificados e interativos (MCQUAIL, 2003).

Dessa forma, importa analisar se as redes sociais se enquadram ou não no conceito de veículo de comunicação social. Essa análise é relevante sob a perspectiva acadêmica e também prática de sua conceituação, posto que permite estabelecer e delimitar o tratamento jurídico que se deve dispensar às redes sociais.

Para Marcos Sabino (2022), as redes sociais não podem ser entendidas como veículos de comunicação social pois nem todo veículo de comunicação é social. Isso se dá por dois motivos: o primeiro, referente à sua origem, diz que para um veículo ser considerado de comunicação social ele deve ter um controle total sobre a mensagem divulgada e responsabilidade editorial, o que não ocorre com as redes sociais, uma vez que a plataformas não têm ingerência total sobre o conteúdo produzido pelos seus usuários⁷⁵. O segundo, referente à consequência, põe que o veículo de comunicação social se define pelos destinatários da mensagem, que devem ser uma massa indeterminada e heterogênea de pessoas, de forma difusa, enquanto as redes sociais se destinam a pessoas determinadas, escolhidas pelo emissor numa noção de comunidade e não de difusão.

⁷⁵ O autor destaca que a moderação de conteúdo não se enquadra como controle editorial.

Nessa mesma toada, para Flávio Martins (2022), não cabe ação de direito de resposta em face de manifestações individuais realizadas nas redes sociais, o que não impediria a invocação de medidas civis, como indenização por dano moral, material ou à imagem, bem como as de natureza penal. Para ele, a Lei 13.188/2015⁷⁶, que dispõe a respeito do direito de resposta, exclui da sua proteção ofensas individuais preferidas por usuários da internet, assegurando o referido direito apenas àquelas matérias divulgadas por veículo tradicionais, ainda que nos meios digitais (rádio, televisão, portais da internet) (MARTINS, 2022).

Contudo, esse não parece ser o entendimento normativo e jurisprudencial que se tem dado no cenário brasileiro, em especial no contexto eleitoral. O Tribunal Superior Eleitoral, nas eleições de 2018, equiparou as redes sociais a veículos de comunicação social, para fins de abertura de investigação judicial para apurar a utilização indevida das redes sociais, conforme o Art. 22 da Lei complementar 64/90⁷⁷.

A questão envolvia a seguinte hipótese: se ataques ao sistema eletrônico de votação e à democracia, que disseminam fatos inverídicos e geram incertezas acerca da lisura do pleito, podem configurar abuso de poder político ou de autoridade e uso indevido dos meios de comunicação quando propagado por meio das redes sociais. Na oportunidade, destacaram os Ministros:

13. A internet e as redes sociais enquadram-se no conceito de “veículos ou meios de comunicação social” a que alude o art. 22 da LC 64/90. Além de o dispositivo conter tipo aberto, a Justiça Eleitoral não pode ignorar a realidade: é notório que as Eleições 2018 representaram novo marco na forma de realizar campanhas, com claras vantagens no uso da internet pelos atores do processo eleitoral, que podem se comunicar e angariar votos de forma mais

⁷⁶ “Art. 2º Ao ofendido em matéria divulgada, publicada ou transmitida por veículo de comunicação social é assegurado o direito de resposta ou retificação, gratuito e proporcional ao agravo. (Vide ADIN 5436). § 1º Para os efeitos desta Lei, considera-se matéria qualquer reportagem, nota ou notícia divulgada por veículo de comunicação social, independentemente do meio ou da plataforma de distribuição, publicação ou transmissão que utilize, cujo conteúdo atente, ainda que por equívoco de informação, contra a honra, a intimidade, a reputação, o conceito, o nome, a marca ou a imagem de pessoa física ou jurídica identificada ou passível de identificação. § 2º São excluídos da definição de matéria estabelecida no § 1º deste artigo os comentários realizados por usuários da internet nas páginas eletrônicas dos veículos de comunicação social.” (Lei Nº 13.188/2015).

⁷⁷ Lei Complementar Nº 64, de 18 de maio de 1990: “Art. 22. Qualquer partido político, coligação, candidato ou Ministério Público Eleitoral poderá representar à Justiça Eleitoral, diretamente ao Corregedor-Geral ou Regional, relatando fatos e indicando provas, indícios e circunstâncias e pedir abertura de investigação judicial para apurar uso indevido, desvio ou abuso do poder econômico ou do poder de autoridade, ou utilização indevida de veículos ou meios de comunicação social, em benefício de candidato ou de partido político, obedecido o seguinte rito [...]”. (Lei complementar 64/90).

econômica, com amplo alcance e de modo personalizado mediante interação direta com os eleitores (BRASIL, 2018, p. 3)⁷⁸.

O Supremo Tribunal Federal, na PET 9068/DF⁷⁹, entendeu que, ao contrário do sustentado por Marcos Sabino, as redes sociais abertas atingem público indeterminado por tempo indeterminado, o que justificaria a sua equiparação, por possuírem caráter de difusão⁸⁰.

Reforçando o caráter de difusão das redes, destaca-se a figura dos *bots*⁸¹, os robôs, que são programas de software que executam tarefas automatizadas e repetitivas e atualmente são responsáveis por mais da metade de todo o tráfego na internet (GARRETT, 2018). Essas ferramentas são utilizadas para a disseminação de informação como se fossem indivíduos, tornando as redes sociais em verdadeiros espaços de difusão, com alcance superior ao das mídias tradicionais (DE PAULA; MICHALSKI, 2019).

Assim, apesar das divergências, as redes sociais vêm sendo equiparadas aos veículos de comunicação social, dado o seu alcance (difusão) e o impacto que os conteúdos produzidos têm gerado na formação da opinião pública. O ciberespaço criado pela internet, e aqui delimitado nas redes sociais, se tornou um dos principais meios de comunicação e está diretamente envolvido na produção e disseminação de sentido acerca dos acontecimentos e contextos da vida social, pública e privada, sem limite de tempo e espaço, e se tornou uma das principais fonte de informação que forma opiniões, além de ser um espaço para a expressão.

⁷⁸ No mesmo sentido destaca-se também a decisão do TSE na AIJE 0600814-85/DF, julgada em 08 de dezembro de 2022.

⁷⁹ “8. Qualquer manifestação na internet, especialmente em redes sociais abertas, tem potencial para atingir o mundo todo e permanecer disponível para acesso, em tese, por tempo indeterminado.” (Pet 9068/DF).

⁸⁰ Nesse mesmo sentido, a resolução do TSE Nº 23.608/2019 assegura o direito de resposta, no período eleitoral, para aquelas ofensas proferidas por meio dos veículos de comunicação social, inclusive as redes sociais: “Art. 31. A partir da escolha de candidatas ou candidatos em convenção, é assegurado o exercício do direito de resposta à candidata, ao candidato, ao partido político, à federação de partidos ou à coligação atingidos, ainda que de forma indireta, por conceito, imagem ou afirmação caluniosa, difamatória, injuriosa ou sabidamente inverídica, difundidos por qualquer veículo de comunicação social, inclusive provedores de aplicativos de internet e redes sociais (Redação dada pela Resolução nº 23.672/2021)”. (TSE - Resolução Nº 23.608/2019).

⁸¹ Sobre o tema, recomenda-se o artigo *Os bots de disseminação de informação na conjuntura das campanhas presidenciais de 2018 no Brasil* (DE PAULA; MICHALSKI, 2019).

A esfera pública da internet pode ser definida, de acordo com Peter Dahlgren (2005), como “uma constelação de espaços comunicativos na sociedade que permitem a circulação de informações, ideias, debates – idealmente de forma irrestrita – e também a formação da vontade política” (p. 148, tradução nossa⁸²). Essa esfera pública deve ser capaz de permitir o diálogo e o acesso à informação.

Observa-se então que as redes sociais não são apenas espaços de conexão entre os usuários para compartilhamento de conteúdo, exercendo liberdades públicas, mas sim são especialmente espaços privados nos quais é possível influenciar as pessoas a adotarem comportamentos que sejam de interesse dos anunciantes, em um novo nível de influência e poder muito superior ao das mídias tradicionais.

Em conclusão, verifica-se que o surgimento da *web 2.0* e das plataformas de redes sociais propiciou a criação de um novo espaço de colaboração e conexão entre as pessoas, sendo que as redes se transformaram em um dos principais palcos de argumentação e deliberação disponíveis para os indivíduos. Ainda, elas podem funcionar como espaço público de debate e difusão da liberdade de expressão por amplificarem o espaço comunicacional, contudo a sua configuração pode apresentar desafios sensíveis que podem colocar em risco a democracia e a própria liberdade de expressão.

2.2 OS DESAFIOS DA LIBERDADE DE EXPRESSÃO NAS REDES SOCIAIS

O advento da *web 2.0* acabou por criar um espaço híbrido ou diferenciado, na medida em que ele é privado por natureza e regido por regras de mercado, mas permite o exercício de liberdades públicas ao autorizar a conexão e a interação entre os usuários, com uma promessa de ser um espaço livre para o exercício da liberdade de expressão. Contudo, esse espaço livre, inicialmente ideal, tem apresentado inúmeros desafios para o exercício da liberdade de expressão.

Para Richard Posner (2018), há quatro motivos de preocupação a respeito do exercício da liberdade de expressão na internet, que são diretamente observados nas

⁸² “a constellation of communicative spaces in society that permit the circulation of information, ideas, debates – ideally in an unfettered – and also the formation of political will” (DAHLGREN, 2005, p. 148).

redes sociais: (i) o anonimato; (ii) a falta de controle de qualidade sobre a informação; (iii) a audiência potencialmente enorme; (iv) o fomento ao comportamento antissocial.

Sabe-se que na internet, e especificamente nas redes sociais, é possível a criação de perfis anônimos, sem um controle imediato quanto à identidade do usuário, o que acaba por facilitar a criação, difusão e consumo de conteúdos ilícitos ou falsos, além da propagação de discursos de ódio. Isso porque os usuários se sentem à vontade atrás de uma tela, na medida em que, de forma prática, é mais difícil rastrear os criadores e usuários desses materiais, aplicando-se responsabilidades (POSNER, 2018).

Por outro lado, há quem defenda que o anonimato é uma técnica legítima de gestão da identidade, especialmente quando voltada ao exercício da liberdade de expressão, pois atua como instrumento viabilizador de uma participação mais democrática (WIMMER; CARVALHO, 2022). Isso se dá especialmente em países nos quais a expressão de forma anônima é o único meio das pessoas expressarem seus sentimentos e opiniões sem censura, contribuindo para a pluralidade de “vozes” no livre mercado de ideias.

Além disso, o anonimato ainda pode trazer segurança para denunciar crimes e abusos, sem medo de retaliações. Em meados de 2017, no Twitter, observou-se um crescimento dos chamados *#exposeds*, que são postagens nas quais meninas e mulheres se sentem seguras para expor casos de assédio. As hashtags “#metoo” e “#meuprimeiroassedio” criaram um encorajamento coletivo que ajudou na conscientização e prevenção dessa violência e isso foi possível pelo uso de perfis considerados anônimos⁸³. Diante desse dilema, a definição adequada de anonimato e o seu enquadramento jurídico-constitucional é um debate relevante para essa pesquisa e será enfrentado no item 3.3.

O segundo motivo de preocupação destacado por Posner (2018) é a falta de controle de qualidade do discurso ou da informação, uma vez que, em princípio, qualquer pessoa pode publicar qualquer coisa na internet e nas redes sociais⁸⁴. Diferente das mídias tradicionais, isso ocorre sem controle prévio ou filtro quanto à

⁸³ G1. Disponível em: <https://g1.globo.com/pop-arte/noticia/a-verdadeira-origem-da-hashtag-me-too-usada-no-twitter-por-mulheres-que-sofreram-violencia-sexual.ghtml>. Acesso em: 12 mai. 2023.

⁸⁴ Como se verá adiante, esse controle muitas vezes é feito pelos algoritmos, mas com diversos problemas.

veracidade das informações, sendo um campo fértil para a propagação de desinformação (*fake news*), como já analisado no item 1.3.

A verdade fica ainda mais em risco no ciberespaço pela chamada *deepfake*, que é uma técnica desenvolvida a partir de um algoritmo criado com aprendizagem profunda (*deep learning*), capaz de se autocorriger e criar vídeos adulterados e realistas. A partir de um vídeo verdadeiro, é possível alterar o rosto e a voz da pessoa, simulando movimento e falas de uma forma muito real, criando um vídeo falso, numa espécie de ilusionismo digital (ROBLES-LESSA; CABRAL; SILVESTRE, 2020). Michael K. Spencer explica que

Deepfakes são, essencialmente, identidades falsas criadas com o Deep Learning [aprendizagem profunda, por meio de uso maciço de dados], por meio de uma técnica de síntese de imagem humana baseada na inteligência artificial. É usada para combinar e sobrepor imagens e vídeos preexistentes e transformá-los em imagens ou vídeos “originais” [...] Essa combinação de vídeos existentes e “originais” resulta em vídeos falsos, que mostram uma ou algumas pessoas realizando ações ou fazendo coisas que nunca aconteceram na realidade. Em 2019, também estamos vendo uma explosão de faces fake, através das quais a IA é capaz de conjurar pessoas que não existem na realidade, e que têm um certo fator de fluência. (SPENCER, 2019, s/p).

Apesar dos vídeos⁸⁵ não serem perfeitos, podem ser bem convincentes e induzir milhares de pessoas ao engano, além de suscitar dúvida quanto ao que é verdadeiro e o que é uma montagem dissimulada, levando a desinformação a uma dimensão muito maior e mais perigosa.

A ausência de controle de qualidade, além da propagação de desinformação, permitiu a criação dos chamados *gripe sites*⁸⁶, que são sites nos quais os usuários expressam seu desprezo por determinada pessoa, organização ou grupo, ampliando a disseminação de discursos de ódio (WARBURTON, 2020).

Para Posner (2018), o primeiro e o segundo problemas (o anonimato e o controle do conteúdo) possuem um aspecto transitório, pois ele acredita que com o uso de tecnologias adequadas e com técnicas regulatórias é possível coibir a

⁸⁵ Compilado de vídeos famosos de *deepfakes* envolvendo artistas e políticos. Disponível em: <https://www.youtube.com/watch?v=8vBj8neXSSE>. Acesso em: 11 mai. 2023.

⁸⁶ Como exemplo cita-se o site www.webgripesites.com, que reúne diversos links com outros *gripe sites*.

disseminação de conteúdos ilegais, protegendo os usuários, e também definir critérios de controle de qualidade tão eficazes quanto os da mídia tradicional.

Contínuo ao problema da qualidade do discurso, tem-se a amplitude da audiência: uma simples postagem em uma rede social pode, em questão de horas, alcançar milhares de pessoas. Por um viés, a amplitude da audiência pode ser considerada como um benefício da internet e das redes, mas o custo social de um discurso ruim, ou seja, o seu prejuízo ou dano, está relacionado com o seu conteúdo e também com a sua extensão, ou o tamanho do público alcançado (POSNER, 2018).

Nesse sentido, pior seria gritar “fogo” em um teatro grande e lotado do que em um pequeno, ou seja, um discurso de ódio ou uma *fake news* em uma rede social terá alcance e impacto muito maiores do que ocorreria em um ambiente apenas físico. O perigo torna-se muito mais claro e iminente no meio virtual: discursos de ódio e desinformação sempre existiram, mas a amplitude dessas expressões na internet e nas redes sociais toma proporções extremamente danosas e não imaginadas por John Stuart Mill em meados do século XIX.

Como já visto, o argumento fundamental para a defesa da liberdade de expressão, para Mill, é a falsidade ou a veracidade dos fatos. Contudo, o espaço de discussão e expressão em 1850 não era o mesmo dos dias atuais, de modo que naquela época a exposição de um discurso falso ou intolerante permitiria um adequado confronto na mesma medida e proporção, podendo se extrair a verdade. Entretanto, essa visão não reflete o que se observa nas redes sociais, dada a amplitude e difusão que os discursos podem alcançar. Em outras palavras, o discurso danoso pode ser muito mais prejudicial nas redes sociais do que já foi em qualquer outro momento da história.

De acordo com um estudo realizado pelo *Massachusetts Institute of Technology* (MIT)⁸⁷, constatou-se que as notícias falsas têm 70% (setenta por cento) mais chances de serem retuitadas do que histórias verdadeiras, e que histórias verdadeiras levam seis vezes mais tempo para atingirem 1500 pessoas do que uma história falsa. Dessa forma, as contribuições de John Stuart Mill, apesar de relevantes, precisam ser compreendidas nesse novo local de discussão, porque o ambiente de discussão de

⁸⁷ Disponível em: <https://news.mit.edu/2018/study-twitter-false-news-travels-faster-true-stories-0308>. Acesso em: 12 mai. 2023.

sua época favorecia um embate igualitário entre verdade e mentira. Já nas redes sociais, a mentira tem seis vezes mais vantagem do que a verdade.

Por outro lado, segundo Posner (2018), com a redução dos custos da comunicação pelas redes sociais, uma vez que o seu acesso é gratuito, houve um aumento na fala, de modo que a internet e as redes sociais se tornaram um método eficaz de se contornar a censura privada feita pelas mídias tradicionais. Essas, por sua vez, muitas vezes relutavam em compartilhar ideias não politicamente corretas, o que se mostrava um obstáculo ao chamado livre mercado de ideias. Nesse ponto reside um grande paradoxo da liberdade de expressão nas redes sociais: sem qualquer censura privada, o volume do discurso aumentaria “ao nível de cacofonia ininteligível”, mas, com ela, discursos, informações ou ideias importantes poderão ser suprimidos (POSNER, 2018, p. 150).

Nessa mesma linha, Denis McQuail (2003) entende que o surgimento das novas mídias tornou mais difícil a comunicação, pois o aumento do volume de informação não necessariamente amplifica o acesso a essa informação, visto que transfere aos indivíduos o dever de escolher, dentre uma avalanche de publicações, a informação que pretende integrar e dar sentido. A publicação de um texto de Machado de Assis em um periódico no passado dificilmente não seria lida, diferentemente se fosse publicada na internet ou redes sociais, pois estaria perdida em uma vastidão de quantidade de informações (DANTAS; SANTOS, 2020).

Em conclusão a este item, o mercado livre de ideias nas redes sociais paradoxalmente depende da seletividade e dos *gatekeepers*⁸⁸, mas também pode ser prejudicado por esses instrumentos (POSNER, 2018).

O quarto problema é o fomento ao comportamento antissocial, porque as redes sociais permitem, de forma muito mais rápida e eficiente, o encontro entre pessoas que pensam de forma semelhante, ao começar a sugerir influenciadores e conteúdos similares por meio dos algoritmos. Isso não seria um problema se essa ferramenta não permitisse também o encontro entre pessoas que possuem comportamentos extremistas, violentos e antissociais.

⁸⁸ *Gatekeepers*, guardiões nos portões, se referem aqui aos responsáveis pela filtragem de conteúdo. No meio jornalístico, seria aquele responsável por definir, de acordo com critérios editoriais, o que vai ser veiculado. Sobre o tema e a mudança no conceito de *gatekeepers* na era digital, recomenda-se o artigo de Gil Baptista Ferreira, *Gatekeeping Changes in the New Media Age: The Internet, Values and Practices of Journalism* (2018).

Pessoas com comportamentos antissociais se tornam encorajadas a expressar suas ideias, aumentando sua autoconfiança, quando encontram outras que pensam da mesma maneira (POSNER, 2018). Em um ambiente físico, esse encontro seria possível, mas em uma amplitude reduzida, o que geraria um impacto também reduzido do problema. Contudo, por intermédio das redes sociais, extremistas podem se comunicar a qualquer hora e em qualquer lugar.

Nesse sentido, relembra-se o episódio ocorrido em 2019, no município de Suzano, São Paulo, no qual dois criminosos mataram oito pessoas na Escola Estadual Professor Raul Brasil⁸⁹. Os criminosos frequentavam o fórum extremista *Dogolachan*, celeiro de atos criminosos de extremismo e ódio. Nesse fórum, os assassinos pediram dicas de como realizar o atentado e, dias antes do episódio, agradeceram o administrador pelas orientações e conselhos para a ação. Os assassinos compartilhavam, em seus perfis no Twitter, Youtube e Facebook, vídeos e mensagens com discursos de ódio (MIRANDA; SANTOS, 2021).

Na maioria das situações, essas atividades são ilegais e merecem ser combatidas de imediato. Contudo, pode-se exemplificar comportamentos não criminosos com consequências gravíssimas nas redes sociais, como por exemplo grupos que promovem a anorexia e bulimia nas redes; a hashtag “borboletana” é usada por jovens no Twitter para compartilhar livremente seus transtornos alimentares e como fazem para emagrecer⁹⁰.

Esses encontros acabam sendo aprimorados pela atuação dos algoritmos, porque conteúdos extremistas e sensacionalistas geram um maior engajamento (cliques, comentários e tempo assistido) e acabam por ser mais recomendados. Desse modo, os algoritmos, ao serem programados para buscar uma sequência de conteúdos que sejam de interesse do usuário, mantendo-o conectado, recomendam conteúdo dessa natureza⁹¹.

⁸⁹ G1. Disponível em: <https://g1.globo.com/sp/mogi-das-cruzes-suzano/noticia/2019/03/14/quilherme-taucci-monteiro-e-luiz-henrique-de-castro-saiba-quem-sao-os-assassinos-de-suzano.ghtml>. Acesso em: 12 mai. 2023.

⁹⁰ Entre maio e junho de 2017, foram identificadas 28.000 menções com o código no Twitter, feitas por 11.000 meninas que contaram com um alcance de 1,7 milhão de pessoas. Disponível em: <https://veja.abril.com.br/saude/estudo-codigos-secretos-sobre-anorexia-twitter/>. Acesso em: 11 mai. 2023.

⁹¹ Disponível em: <https://www.algotransparency.org/>. Acesso em: 11 mai. 2023.

Para Richard Posner (2018), o melhor caminho, nesses casos, é a vigilância por parte do Estado, e não simplesmente a censura de perfis ou sites, pois isso pode gerar um efeito colateral de censurar comunicações, mesmo as não convencionais. Isso porque a liberdade de expressão se presta tanto aos fanáticos quanto aos intelectuais educados (WARBURTON, 2020).

A weak government can be undermined by free speech, but a strong government can be strengthened by it because it enables government to keep itself better informed of potential threats, assuming the cost of surveillance is not prohibitive. (POSNER, 2018, p. 150⁹²).

Seguindo essa linha de pensamento, recomenda-se, ao invés do simples bloqueio das contas, a criação de instrumentos tecnológicos por meio, por exemplo, de algoritmos, que, ao identificar comportamentos antissociais, rebatessem essas expressões com informações verdadeiras e oferta de ajuda.

Cass R. Sustein (2020) apresenta uma preocupação diferente em relação à internet, às redes sociais e ao exercício da liberdade de expressão e sua conexão com a democracia. Para ele, um dos principais riscos da internet e das redes sociais é a filtragem de conteúdo feita pelo próprio usuário, que tem o “poder” de selecionar o que quer ver, adequando as notícias, o entretenimento e todo material que recebe e consome aos seus gostos pessoais. Assim, pode facilmente encontrar pessoas com os mesmos interesses e comportamentos e evitar assuntos dos quais discorda, atuando como *gatekeeper* de si próprio, o que foi denominado por Nicholas Negroponte⁹³ como *The daily me*, ou seja, o seu próprio noticiário, numa individualização do seu próprio universo comunicacional, escolhendo as categorias e ordenando-as de acordo com as suas preferências.

Essa filtragem se apresenta como um risco à democracia, na medida em que uma sociedade plural e democrática depende de dois requisitos: (i) as pessoas devem ser expostas a conteúdos e informações que não teriam escolhido com antecedência; (ii) os indivíduos devem ter uma série de experiências compartilhadas. Sem esses

⁹² “Um governo fraco pode ser prejudicado pela liberdade de expressão, mas um governo forte pode ser fortalecido por ela porque permite que o governo se mantenha mais bem informado sobre ameaças potenciais, desde que o custo da vigilância não seja proibitivo.” (POSNER, 2018, p. 150, tradução nossa).

⁹³ Nicholas Negroponte é um professor do MIT que cunhou a expressão *The daily me* em 1995. (Disponível em: <http://assets.press.princeton.edu/chapters/s8468.pdf>. Acesso em: 10 mai. 2023).

requisitos, a sociedade terá dificuldades em lidar com problemas sociais, inclusive uns com os outros, ampliando a intolerância (SUSTEIN, 2020)⁹⁴.

A filtragem de conteúdo não é uma exclusividade da internet e das redes sociais; as pessoas sempre filtraram os materiais que gostariam ou não de consumir, comprando jornais com editoriais das mesmas ideologias ou revistas de seus interesses. Contudo, o ponto destacado por Sustain (2020) é o de que na internet há um aumento no grau da individualização, enquanto nas mídias tradicionais, em algum momento, o indivíduo poderia se deparar, ainda que casualmente, com um conteúdo que não teria escolhido previamente, cruzado ou antagônico à sua ideologia.

Essa individualização é amplificada a patamares não antes conhecidos com o incremento dos algoritmos, que selecionam conteúdos a serem exibidos aos usuários conforme os seus comportamentos e as suas escolhas, criando as chamadas “bolhas” ou *echo chambers* (QUATTROCIOCCI; SCALA; SUNSTEIN, 2016).

Para Piaia, Ritter e Sangoi (2018), *echo chambers* “seriam ambientes semelhantes a câmaras de eco acústicas, em que informações, ideias ou crenças seriam amplificadas pela comunicação, sendo repetidas dentro de um sistema definido para aquele propósito” (p. 302). Nessas câmaras o conteúdo compartilhado é apenas aquele que reforça uma narrativa específica, ampliando a radicalização dos usuários pela deterioração das capacidades racionais em uma autolimitação temática e desvirtuada da realidade, aumentando a desinformação numa espécie de “patologia social”⁹⁵, uma vez que opiniões divergentes são geralmente censuradas ou desautorizadas⁹⁶.

Por vezes os usuários nessas câmaras são expostos em algum nível a um conteúdo “cruzado”, antagônico, em decorrência das informações eventualmente compartilhadas pelas pessoas que os usuários estão seguindo, mas em proporções muito inferiores daqueles que não as integram (RECUERO; SOARES; ZAGO, 2020).

⁹⁴ Em sentido contrário, Brundidge entende que a internet contribuiu para o aumento da heterogeneidade das redes de discussão política, em decorrência da facilidade de acesso aos espaços de exposição e confronto das opiniões políticas. (BRUNDIDGE, 2010 *apud* COLLEONI; ROZZA; ARVIDSSON, 2014).

⁹⁵ Expressão do sociólogo Axel Honneth em *El derecho de la libertad*. (HONNETH *apud* PIAIA; RITTER; SANGOI, 2018).

⁹⁶ “Aceitamos a realidade do mundo como nos é mostrada”, em trecho do filme *O Show de Truman* (1998).

Essas câmaras de eco, inicialmente consideradas como inofensivas e um produto de algoritmos de publicidade, ganharam repercussão mundial ao se tornarem alvo de investigação criminal nas eleições americanas de 2016, com a polêmica eleição de Donald Trump e a contratação da *Cambridge Analytica* (VESTING, 2020). Como noticiado à época, em 2018, a empresa teria se utilizado, sem consentimento, de 87 milhões de perfis no Facebook a fim de direcionar propagandas políticas em prol do candidato, influenciando a decisão dos eleitores.

A Revista *Science* publicou um artigo em 2015⁹⁷ no qual foram analisados 10,1 milhões de usuários norte-americanos no Facebook. Observou-se que dentro do domínio de notícias políticas encontradas nas mídias sociais houve uma significativa exposição seletiva, de modo que uma notícia com ideologias republicanas quase nunca apareceria no *feed* de notícias de um democrata e vice-versa. O estudo constatou que o conteúdo que um indivíduo consome no Facebook depende: do que os amigos compartilham; de como o algoritmo do *feed* separa os conteúdos; e o que os usuários decidem ler (PIAIA; RITTER; SANGOI, 2018).

Em resumo, o algoritmo do Facebook reduz em 1 ponto percentual a exposição ao conteúdo que poderia ser considerado antagônico às crenças de seus usuários, mas os cliques causam uma redução de 4 pontos percentuais na exposição aos artigos cruzados apresentados pela rede social. Ademais, os autores identificaram que um usuário médio tem em torno de 20% da lista de amigos composta por pessoas de ideologias distintas, de modo que as escolhas individuais e a composição das redes sociais são fatores relevantes para a formação das câmaras de eco⁹⁸ (PIAIA; RITTER; SANGOI, 2018).

Ainda de acordo com o estudo, após ranking, chegou-se a uma taxa de exposição de conteúdo cruzado de 5% para conservadores (direita) e 8% para liberais (esquerda), ou seja, apenas 5% de todo o conteúdo visto por um conservador norte americano era divergente da sua posição ideológica (PIAIA; RITTER; SANGOI, 2018).

Para Colleoni, Rozza e Arvidsson (2014), em estudo sobre *echo chambers* no Twitter, essa fragmentação do discurso político é chamada de homofilia, definida

⁹⁷ *Exposure to ideologically diverse news and opinion on Facebook* (2015). Disponível em: <https://www.science.org/doi/10.1126/science.aaa1160>. Acesso em: 13 mai. 2023.

⁹⁸ Disponível em: https://www.unicamp.br/unicamp/sites/default/files/jornal/paginas/ju_624_paginacor_02_web.pdf. Acesso em: 11 mai. 2023.

como uma tendência de indivíduos semelhantes formarem grupos homogêneos e laços uns com os outros, de modo que as câmaras de eco são fenômenos que atestam a homofilia política em certo ambiente.

A questão da homofilia política é importante porque diz respeito à capacidade dos meios digitais de apoiar a formação de uma esfera pública, onde a diversidade de opiniões e informações possam interagir ou, ao contrário, funcionar como uma câmara de eco que reforça perspectivas e convicções estabelecidas (COLLEONI; ROZZA; ARVIDSSON, 2014, p. 3).

Dentre as suas explicações para esse fenômeno estão a dissonância cognitiva e as teorias da exposição seletiva, que atestam que os indivíduos experimentam sentimentos positivos quando recebem informações que confirmam suas opiniões e se sentem estressadas diante de opiniões divergentes. Assim, as escolhas feitas pelos usuários seguem a tendência de reduzir sua dissonância cognitiva, optando por receber conteúdos que reforcem suas crenças (COLLEONI; ROZZA; ARVIDSSON, 2014).

Essa homofilia, quando trazida para o universo político, pode resultar na polarização política, empiricamente observada nas redes sociais, especialmente em períodos eleitorais, pois as informações recebidas é que vão formar as opiniões e as atitudes dos eleitores. Assim, quanto maior a heterogeneidade da rede que o indivíduo está inserido, maior será a sua busca por conteúdos diferentes, bem como há um aumento da consciência e tolerância a posições políticas divergentes. Em contrapartida, a exposição exclusiva do indivíduo a ideias homogêneas está associada à adoção de posições extremistas e intolerantes (SUSTEIN, 2018).

Relembra-se o caso dos discursos de ódio propagados no Facebook, que levaram à morte de mais de dez mil muçulmanos rohingya em Myanmar; como o país é predominantemente budista, os muçulmanos têm sido alvos de uma limpeza étnica na região. Diversos refugiados na Inglaterra e Estados Unidos estão processando o Facebook e exigindo indenização por permitir que esses discursos se propagassem, uma vez que naquele país há mais de 20 milhões de usuários e a rede social é considerada a principal fonte de notícias⁹⁹.

⁹⁹ G1. Disponível em: <https://g1.globo.com/tecnologia/noticia/2021/12/07/refugiados-rohingya-pedem-r-850-bilhoes-de-indenizacao-ao-facebook-por-discurso-de-odio-em-mianmar.ghtml>. Acesso em: 13 mai. 2023.

Especificamente em relação ao Twitter, Colleoni, Rozza e Arvidsson (2014) ressaltam que, como as postagens e os perfis são abertos para todos, salvo se o usuário expressamente alterar, pode-se seguir uma conta independentemente da permissão do dono e o conteúdo pode ser facilmente difundido (retuite). Com o sistema de hashtag e menção é possível criar espaços de discussão sem a necessidade de criação de grupos, de modo que, teoricamente, o Twitter pode ser um ambiente propício para a formação de uma esfera pública, o que não afasta a possibilidade de formação das câmaras de eco.

Em conclusão ao estudo, o nível de homofilia no Twitter variou de acordo com a finalidade para o qual ele é usado:

Se olharmos para o Twitter como um meio social, veremos níveis mais altos de homofilia e uma estrutura de comunicação mais parecida com uma câmara de eco. Mas se, em vez disso, focarmos no twitter como um meio de notícias, olhando para a difusão de informações independentemente das redes sociais, veremos níveis mais baixos de homofilia e um cenário mais parecido com a esfera pública. (COLLEONI; ROZZA; ARVIDSSON, 2014, p. 21).

Dessa forma, observou-se que todas as redes sociais são propícias à formação das câmaras de eco, sendo que o seu grau de homofilia é mais elevado nas redes interpessoais que exigem reciprocidade, como seguir a conta para receber informações, como é o caso do Facebook, do que nas redes que permitem relação independentemente de reciprocidade, como ocorre no Twitter.

Os desafios aqui analisados não são os únicos que colocam em risco o exercício da liberdade de expressão, mas os que foram pontuados devem ser objeto de reflexão, debate e, eventualmente, regulação.

2.3 DA MUDANÇA SOCIAL PARA A MUDANÇA REGULATÓRIA

Observou-se que as redes sociais tiveram o condão de alterar o espaço social em que a liberdade de expressão é exercida, criando um novo espaço de expressão, e que esse espaço desencadeou inúmeros desafios. Contudo, é preciso analisar mais a fundo essa mudança social para então se compreender a necessidade ou não de uma mudança regulatória.

Para Jack M. Balkin (2014), a mudança social provocada pelo surgimento das novas tecnologias, aqui em especial as redes sociais, implica na necessidade de uma mudança regulatória. Compreender essa mudança social, de um período pré-digital para um período digital, é útil para o debate sobre a necessidade de uma mudança regulatória, no que diz respeito à liberdade de expressão.

O exercício da liberdade de expressão não se resume apenas ao direito de expressar-se; é um direito que pressupõe uma estrutura material para ser exercido, o que se denominou de “infraestrutura da liberdade de expressão”¹⁰⁰ (BALKIN, 2014). Identificar essa infraestrutura é uma das formas de compreender como as regulações afetam o exercício da liberdade de expressão, pois ao regulá-la se estará também regulando essa liberdade. Nesse sentido, a principal característica da era digital, no que diz respeito à liberdade de expressão, é a revolução provocada na sua infraestrutura, e nela reside a mudança social (BALKIN, 2014)¹⁰¹.

Assim, passa-se a comparar a infraestrutura da liberdade de expressão nos períodos antes e depois da era digital, aqui tratando especificamente do surgimento da internet e das plataformas de redes sociais.

Como visto, antes da era digital as tecnologias usadas para o exercício da liberdade de expressão, para além dos espaços públicos, eram predominantemente as mídias impressas e a televisão (CNN, Rede Globo, The New York Times etc.), às quais pertenciam os conglomerados de mídias. A disponibilização desses espaços de fala era feita àqueles que detinham essa infraestrutura (RAMOS, 2021).

Atualmente, a tecnologia utilizada é a internet e em especial as redes sociais (Facebook, Twitter, Instagram, Tiktok e Youtube), das quais são detentores os

¹⁰⁰ Definição de Jack Balkin sobre a infraestrutura da liberdade de expressão: “What is the infrastructure of free expression? (...) These elements, in turn, are surrounded by an even larger network of supporting institutions. Similarly, the New York Times of the midtwentieth century featured in Sullivan and Pentagon Papers was not simply a set of pages with ink. It was the cumulative product of editors, reporters, newsrooms, bureaus, wire services, printing machines, labor unions, delivery trucks, and subscription services; and it too depended on a larger set of businesses, contractual arrangements, customs, and conventions to produce “[a]ll the news that’s fit to print.” From an even broader perspective, we can see that the democratic model of free expression celebrated in Sullivan and Pentagon Papers, and the public sphere of knowledge and opinion that legitimates democracy, depend on a variety of institutions like telephone companies, public libraries, bookstores, schools, universities, post offices, subsidized postal rates, broadband services, and so on. Once we shift our focus from the moment of expression to the technological, economic, and social infrastructure that supports and enables expression, we can understand how crucial infrastructure is to the freedoms of speech and press.” (BALKIN, 2014, p. 2301-2302).

¹⁰¹ “The hallmark of the digital age is a revolution in the infrastructure of free expression.” (BALKIN, 2014, p. 2301)

conglomerados de tecnologia (Google, Facebook etc.). Contudo, a disponibilização do espaço de manifestação pertence não mais àqueles que detêm a infraestrutura, mas aos usuários dos serviços das plataformas na *web 2.0*. (RAMOS, 2021).

Observa-se, assim, que a grande mudança social não está apenas na mudança tecnológica, mas sim na abertura de espaços de manifestação para divulgação de conteúdos não confeccionados por aqueles que detêm a estrutura, e sim pelos usuários consumidores e anunciantes da plataforma. Ademais, essa mudança social não implicou na extinção dos intermediários da liberdade de expressão, mas na criação de novos, que são as plataformas, as quais se tornam alvos diretos da regulação por deterem a nova estrutura da liberdade de expressão (RAMOS, 2021). Contudo, de que forma essa mudança social, que é uma mudança na infraestrutura, impõe uma mudança regulatória da liberdade de expressão?

Jack Balkin (2014) analisou essa mudança regulatória a partir de uma comparação entre o modelo de regulação das mídias tradicionais, que chamou de *old school speech regulation*, e o modelo de regulação das mídias digitais, que chamou de *new school speech regulation*. Ressalta-se que esses modelos de regulação levam em consideração apenas a regulação estatal.

No modelo *old school*, a regulação estatal da liberdade de expressão é direcionada para: (i) as pessoas, que, como já visto no item 1.2, podem ter uma amplitude de expressão maior ou menor; (ii) os espaços de expressão; e (iii) os veículos de comunicação social tradicionais (BALKIN, 2014).

Essa regulação pode ser facilmente identificada no texto constitucional quando, por exemplo, ao regular pessoas, assegura-se a liberdade de manifestação do pensamento, mas veda-se o anonimato ou garante-se o sigilo da fonte ao jornalista; ao regular espaços de expressão, garante-se o direito de reunião em espaços abertos ao público, mas controla-se pela exigência do prévio aviso¹⁰²; regula-se os veículos de comunicação social ao estabelecer que têm a sua produção e programação delimitada a princípios constitucionais (art. 221 CF/88), além da responsabilidade editorial (art. 222 CF/88) e que dependem de concessão pública para a prestação de

¹⁰² A respeito do alcance do prévio aviso, o Supremo Tribunal Federal firmou a seguinte tese: “A exigência constitucional de aviso prévio relativamente ao direito de reunião é satisfeita com a veiculação de informação que permita ao poder público zelar para que seu exercício se dê de forma pacífica ou para que não frustre outra reunião no mesmo local”. (STF, RE 806339). Logo, a ausência de aviso formal perante as autoridades competentes, não tornam, por si só, ilegítimas as manifestações.

serviços de radiodifusão sonora e de sons e imagens (art. 223 CF/88), controlando, assim, o monopólio sobre esses serviços.

Já na *new school*, a regulação estatal da liberdade de expressão passa a ser prioritariamente do intermediário e não do produtor do conteúdo, ou seja, das plataformas. Isso porque o aumento do número de pessoas com lugar de fala dificulta a ação direta do Estado sobre os indivíduos, de modo que é mais fácil controlar as plataformas por meio de estímulos regulatórios do que controlar aqueles que realmente produziram o conteúdo (BALKIN, 2014).

A mudança social que acabou por alterar o principal destinatário da regulação, do indivíduo para as plataformas, também alterou o modelo tradicional de regulação da liberdade de expressão – antes dualista ou diádico, em que se tem o Estado de um lado e aquele que se expressa do outro – para um modelo plural com múltiplos atores. Ele pode ser exemplificado, de acordo com Jack Balkin (2018), por um triângulo, pois é formado em um dos lados pelo Estado, do outro pelas empresas de infraestrutura (plataformas) e, por fim, por aqueles que se expressam nesses ambientes (usuários).

Nessa estrutura triangular observa-se a coexistência de três modelos regulatórios: a regulação tradicional (*old school*), direcionada diretamente àqueles que se expressam (indivíduos e meios de comunicação tradicionais); a nova regulação (*new school*), direcionada à infraestrutura da internet, que indiretamente implica no controle do discurso dos indivíduos e se apresenta de forma cooperativa; e a regulação privada, feita diretamente pelas plataformas, com técnicas de governança privada, numa autorregulação por meio de instrumentos próprios e decisórios de controle e moderação de conteúdo.

Dessa forma, as empresas privadas não devem ser consideradas como instituições neutras, meras intermediárias e delegatárias da ação estatal, que somente asseguram o espaço de expressão sem qualquer ingerência sobre o conteúdo ou procedência, mas devem ser compreendidas como verdadeiras *publishers*, ou seja, instituições capazes de formar políticas de conteúdo, editando e controlando as expressões realizadas pelos usuários e, portanto, com responsabilidades (VENTURI, 2022).

Assim, a partir do diagnóstico feito por Jack Balkin (2018), pode-se compreender que o surgimento das novas tecnologias não provocou uma mudança no conceito da liberdade de expressão. A mudança social provocada pelo surgimento

das redes sociais está na estrutura utilizada para o exercício dessa liberdade, uma vez que passa a ser disponibilizada para os usuários e não apenas para os detentores da infraestrutura. Isso desencadeou a necessidade de uma mudança regulatória estatal (*old school* para a *new school*), a partir do surgimento de um novo destinatário da regulação: as plataformas, que além de intermediárias de conteúdo, criaram também um novo espaço de decisão, controlando a liberdade de expressão por meio de instrumentos de regulação próprios.

Desse modo, se a regulação da liberdade de expressão nas redes sociais não fica a cargo apenas do Estado, mas também é feita pelas plataformas, compreender apenas as mudanças regulatórias estatais (de *old school* para *new school*) não é suficiente para abranger todos os impactos que a era digital provocou na regulação da liberdade de expressão; é preciso ir além e compreender o processo de decisão feito pelas empresas.

Sendo assim, a regulação da liberdade de expressão nas redes sociais deve ser analisada a partir dessa lógica regulatória triangular: regulação privada por meio da moderação de conteúdo (autorregulação), regulação estatal direcionada aos indivíduos diretamente (*old school*) e regulação estatal direcionada às plataformas (*new school*).

3 A REGULAÇÃO DA LIBERDADE DE EXPRESSÃO NAS REDES SOCIAIS

As redes sociais, além de criarem um amplo espaço de manifestação, também criaram um espaço de decisão, uma vez que as plataformas, ao operarem como intermediárias de conteúdo, acabam por decidir o que pode ou não ser dito nesse ambiente. Isso ocorre por meio da moderação de conteúdo, com amplas competências e possibilidade de intervenção, numa espécie de poder normativo (termos e condições) e decisório, pois forma um fundamento institucional de ação e processo ao qual são submetidos os usuários, caso queiram ingressar na plataforma.

Esse novo ambiente, unilateralmente regulado pelas plataformas, não pode estar distante da função de proteção e garantia do Direito. É tarefa do Estado assumir uma postura de responsabilidade garantidora para assegurar o exercício socialmente aceitável da liberdade de expressão no meio digital, limitando o poder privado e transnacional das plataformas, sem, contudo, ampliar o risco de violações desse direito (HOFFMANN-RIEM, 2022).

Com efeito, para muitos, o ambiente online ainda é visto como uma verdadeira terra sem lei, baseada em uma arquitetura técnica descentralizada e complexa, que impede ou ao menos estabelece restrições significativas à efetiva regulação e imposição coercitiva de normas jurídicas pelos Estados nacionais (WIMMER; CARVALHO, 2022).

Essa dificuldade decorre centralmente do caráter dinâmico do mundo digital, o que denota uma certa incerteza sobre a eficiência da regulação puramente estatal, que talvez não seja capaz de acompanhar esse dinamismo. Aliás, em um primeiro momento, o espaço virtual era considerado como um ambiente que não precisaria de regulamentação e nele as pessoas estariam livres do poder estatal; relembra-se a Declaração de Independência do Ciberespaço, de John Perry Barlow (1996):

Governos do Mundo Industrial, vocês gigantes aborrecidos de carne e aço, eu venho do espaço cibernético, o novo lar da Mente. Em nome do futuro, eu peço a vocês do passado que nos deixem em paz. Vocês não são bem-vindos entre nós. Vocês não têm a independência que nos une.

Os governos derivam seu justo poder a partir do consenso dos governados. Vocês não solicitaram ou receberam os nossos. Não convidamos vocês. Vocês não vêm do espaço cibernético, o novo lar da Mente.

Não temos governos eleitos, nem mesmo é provável que tenhamos um, então eu me dirijo a vocês sem autoridade maior do que aquela com a qual a liberdade por si só sempre se manifesta.

Eu declaro o espaço social global aquele que estamos construindo para ser naturalmente independente das tiranias que vocês tentam nos impor. Vocês não têm direito moral de nos impor regras, nem ao menos de possuir métodos de coação a que tenhamos real razão para temer. (s/p).

Contudo, esse ambiente autorregulado, no qual não haveria espaço para a soberania estatal, não tem se apresentado como um ambiente tão livre e ideal. No que toca especialmente à liberdade de expressão, a instrumentalização das plataformas para propagação de discursos de ódio e desinformação, as consequências negativas suportadas pela sociedade e as acusações de censura privada pelas plataformas têm desencadeado um debate mundial a respeito da eficiência dos modelos de regulação atual das redes.

Com esses aspectos em análise, o presente capítulo visa lançar luzes sobre esse debate para o fim de: (3.1) apresentar a nomenclatura regulatória que será aqui utilizada e delimitar o espectro regulatório da liberdade de expressão, analisando seu objeto, que são as relações jurídicas existentes nas redes e limites; (3.2) analisar a autorregulação ou regulação privada feita pelas plataformas, tendo como norte a moderação de conteúdo; (3.3) apresentar o modelo de autorregulação regulada no cenário regulatório nacional a partir de uma análise dos principais aspectos dos marcos regulatórios nacionais, a lei 12.965/2014 (Marco Civil da Internet) e o projeto de lei 2.630/2020 (Lei de Transparência e Responsabilidade na Internet).

3.1 A REGULAÇÃO DAS REDES SOCIAIS: MODELOS, OBJETO E LIMITES NO CENÁRIO NACIONAL

Para compreender a dinâmica regulatória das redes sociais, é preciso analisar, de forma mais específica, o fenômeno da regulação, tanto econômica quanto social, e os modelos regulatórios existentes, delimitando o seu objeto e traçando os seus limites na ordem jurídica brasileira.

A regulação aqui deve ser compreendida como uma intervenção pública na atividade econômica, por meio da produção e utilização de normas que visam conduzir e até mesmo limitar a atuação de determinados sujeitos jurídicos, para assegurar que a sua atividade seja desenvolvida com a ponderação de interesses para além daqueles que decorreriam apenas das regras de mercado (FARINHO, 2020).

A doutrina ainda tem reconhecido a chamada regulação social, que é também uma forma de intervenção, mas que não tem como objetivo preliminar regular um mercado e sim delimitar aspectos comportamentais dos sujeitos em uma relação social sensível (FARINHO, 2020).

Essa distinção entre regulação econômica (mercado) e social (comportamental) se mostra pertinente pois, apesar da regulação das redes sociais operar em mercados – tanto o publicitário quanto o de interação dos usuários –, os principais problemas enfrentados nas redes são sociais. Por exemplo, quando uma plataforma impede a postagem de um determinado conteúdo, violando a liberdade de expressão do usuário, tal conflito demanda uma regulamentação que não possui diretamente um conteúdo econômico, mas sim social. Isso porque o conflito poderá até avançar para um problema de qualidade e eficiência do serviço prestado pela plataforma, mas inicialmente não é econômico, e sim de violação de um direito fundamental que se pretende assegurar.

O conceito de regulação também não deve ser confundido com legislação. A produção legislativa, que também ordena comportamentos e regulamenta atividades, deve ser considerada como uma espécie de conceito mais amplo, que, para além da legislação, inclui as normas administrativas, as decisões dos tribunais e as normas privadas, de modo que a regulação legislativa é apenas uma modalidade de regulação a par da regulação administrativa, privada e judicial (FARINHO, 2020).

Considerando as diferentes tipologias regulatórias e as diversas nomenclaturas utilizadas, primeiramente apresenta-se aquela que será aqui adotada, a qual é mais adequada à análise do tema: a) autorregulação ou regulação privada; b) heterorregulação ou regulação pública e c) corregulação ou autorregulação regulada.

A autorregulação ou regulação privada é a capacidade de um sujeito de auto submeter-se às regras determinadas na esfera de sua autonomia privada, e se refere aqui à possibilidade de as plataformas de redes sociais configurarem o seu próprio modelo de negócio, bem como determinarem a formação de relações com e entre os usuários da plataforma por meio de normas de conduta auto vinculativas, como os termos e condições gerais de uso (TCe) (HOFFMANN-RIEM, 2022).

Inclui-se nessa esfera de autonomia a capacidade das empresas definirem os seus projetos técnicos e controle das estruturas e serviços, como por exemplo: o perfil da plataforma; se poderão ser postados fotos, textos, vídeos; qual será o formato e a

extensão; e definir o desenvolvimento dos algoritmos, uma vez que são eles que também controlam os serviços processados nas plataformas, os quais devem ser utilizados sob a responsabilidade da empresa.

A autorregulação tem como vantagem a “[...] eficiência pela disposição do conhecimento interno e dinâmica de constante revisão de conceitos” (MARANHÃO; CAMPOS, 2020, p. 344), mas tem como desvantagem o fato de que nem sempre está preocupada em atender o interesse público, como se verá na sequência.

A heterorregulação ou regulação pública, chamada por Wolfgang Hoffmann-Riem apenas de regulação (2022), ocorre quando há uma intervenção pública em processos sociais, impondo determinadas diretrizes para que se atenda interesses públicos. Como já destacado, ela pode se apresentar como regulação legislativa, administrativa (poder regulamentar) e, de forma suplementar, a judicial.

A regulação pública tem a vantagem de ser facilmente implementada pela coercitividade, dado o poder de império do Estado, mas tem a desvantagem de não conseguir, no ambiente digital, acompanhar de forma eficiente a dinâmica das redes. A complexa estrutura organizacional das empresas (cada vez mais especializadas) e a velocidade com que a tecnologia avança fazem com que o Estado apresente um déficit cognitivo em relação às empresas privadas, se mostrando ineficaz no papel exclusivo de regular, supervisionar e sancionar essas instituições.

Nesse cenário de insuficiência dos dois modelos regulatórios anteriores é que desponta um modelo promissor, híbrido, que visa conciliar os dois modelos anteriores, com prevalência da autorregulação, que é a autorregulação regulada ou corregulação, que será melhor analisada posteriormente.

A atividade regulatória, independente da sua modalidade, deve estar enquadrada e conformada de acordo com os direitos fundamentais que aqui se apresentam como limites à regulação das redes sociais.

Os direitos fundamentais, na sua dimensão horizontal, se apresentam como o principal limite à autorregulação das plataformas que, num viés negativo, devem, na moderação e recomendação de conteúdo, assegurar a liberdade de expressão e se abster de práticas de censura. Num viés positivo, devem combater a divulgação de discursos de ódio, notícias falsas ou conteúdos ilegais, como pornografia infantil e violação de direitos autorais, além de definir termos e condições de uso compatíveis com a ordem constitucional brasileira.

Ademais, esses mesmos direitos devem ser garantidos quando da heterorregulação, cabendo ao Estado ponderar as liberdades fundamentais tanto dos usuários quanto da própria plataforma. Assim, para a plataforma devem ser asseguradas a liberdade econômica, aqui considerada a livre iniciativa; a livre concorrência; e o direito de propriedade, nos termos do art. 170, *caput*, II, IV e parágrafo único da CF/88. Para Domingos Soares Farinho (2020), há que se considerar, inclusive, o direito à liberdade de expressão das plataformas, que consistiria no direito das redes, como verdadeiras *social medias*, possuírem uma liberdade editorial, para o fim de escolherem e moderarem “conteúdos destinados a garantir que os utilizadores vêem aquilo que as redes sociais entendem que eles querem ver.” (p. 59).

Contudo, esse conceito de liberdade de expressão das plataformas deve ser visto com certa cautela, na medida em que a recomendação de conteúdos tem seguido uma lógica apenas de mercado e não de proteção dos direitos fundamentais, o que tem contribuído para a formação das câmaras de eco e polarização, como já analisado.

Para os usuários, devem ser assegurados direitos para que possam se defender de eventuais abusos cometidos pelas plataformas no exercício de sua autorregulamentação, como por exemplo maior transparência e informação sobre o consentimento dos termos e condições de uso; devido processo legal e ampla defesa, quando da remoção de conteúdo ou inadmissibilidade de impulsionamento; e direitos à liberdade de expressão, informação e privacidade.

Além disso, é importante que se reconheça que nas relações entre os usuários e as plataformas há uma evidente assimetria de poder, de modo que a regulação estatal, dado o seu dever de proteção dos direitos fundamentais, pode impor comportamentos positivos e não apenas negativos às plataformas, comandos regulatórios, propondo inclusive formas de ponderar, em abstrato, direitos em conflito, propondo soluções (FARINHO, 2020).

Esses limites podem ser mais bem visualizados após se definir o objeto da regulação, que são as relações jurídicas existentes nas redes. Para Farinho (2020), observa-se três relações jurídicas distintas nas redes sociais: (i) relação entre rede social e outras redes sociais concorrentes (relações de mercado); (ii) relação entre

rede social e usuários ou um sujeito externo; (iii) relação entre os usuários da rede social entre si, ou um usuário e um sujeito externo.

A primeira relação jurídica, de ordem concorrencial, é uma regulação de mercado, externa, que gira em torno do equilíbrio entre a eficiência do mercado e a proteção dos consumidores. Nessa relação, a regulação deve buscar seus fundamentos e limites nas regras de mercado previstas na Constituição e em normas infraconstitucionais, como vedação ao monopólio¹⁰³, aqui se considerando as redes sociais como veículos de comunicação social, bem como coibir práticas anticoncorrenciais¹⁰⁴. Como essa situação jurídica não guarda relação direta com a liberdade de expressão, objeto desse trabalho, não será alvo de maior análise.

A segunda relação jurídica é entre a rede social e os usuários. Ela deve ser tratada como uma relação de consumo, da qual se extraem dois tipos de consumidores: o usuário-utilizador, que é aquele que utiliza os serviços da plataforma para compartilhamento de conteúdo no seu perfil, e o usuário-anunciante, que se utiliza das ferramentas de publicidade para anunciar e vender seus produtos na plataforma. Aliás, nesse modelo de negócio é o primeiro consumidor que alimenta o último (FARINHO, 2020).

Como já visto, os limites da regulação dessa segunda relação jurídica devem ser os direitos e liberdades consagrados na Constituição Federal, agregados pelos direitos decorrentes da sua posição de consumidor¹⁰⁵, incluindo-se o reconhecimento da sua hipossuficiência na relação com a plataforma.

A relação de consumo pode ser vislumbrada, por exemplo, nas seguintes situações: imagine-se que uma determinada rede social tenha excluído a conta de um usuário. Apesar das redes terem prerrogativas quanto ao controle e moderação do conteúdo, essa exclusão não pode ser feita de forma arbitrária. Devem ser assegurados ao usuário, de forma transparente, os motivos da exclusão (qual infração teria cometido), bem como garantido o direito de defesa, sob pena de estar violando o seu direito à liberdade de expressão. Outro exemplo comum são as contas invadidas

¹⁰³ A rede social Facebook, atual meta, está sendo processada por monopólio nas redes sociais e adoção de condutas anticoncorrenciais. (G1. Disponível em: <https://g1.globo.com/economia/tecnologia/noticia/2020/12/09/facebook-e-alvo-de-acao-antitruste-nos-estados-unidos-diz-agencia.ghtml>. Acesso em: 11 mai. 2023).

¹⁰⁴ Vide art. 220 § 5º da CF/88 e art. 9º IV da Lei 12.965/2014 (Marco Civil da Internet).

¹⁰⁵ Ressalta-se que a defesa do consumidor é um dos princípios da ordem econômica, elencado no art. 170, V da CF/88.

(*hackeadas*), ou quando há criação de perfil falso de um usuário, que mesmo denunciado não é excluído pela rede social; nessas hipóteses há uma evidente falha na prestação do serviço, que deve ser resolvida com a aplicação das regras consumeristas.

O Supremo Tribunal Federal, no RE 1037396¹⁰⁶, ao reconhecer a repercussão geral do tema 987, que discute a responsabilidade civil de provedores de internet, websites e gestores de aplicativos das redes sociais (art. 19 do Marco Civil da Internet), entendeu que, em situações como as exemplificadas, o usuário, na condição de vítima, equipara-se à figura do consumidor, nos termos do art. 17 do Código de Defesa do Consumidor¹⁰⁷.

A terceira relação jurídica se dá entre os usuários ou ainda entre um usuário e um sujeito externo, não usuário da plataforma. Observa-se que nesse caso não há uma relação direta de consumo, mas uma relação entre indivíduos mediada pela plataforma, que deverá, na hipótese de conflito, realizar uma ponderação entre os direitos. É essa mediação, feita primeiramente no âmbito privado, que difere essa relação jurídica das relações jurídicas “tradicionais” (FARINHO, 2020).

Destaca-se ainda a relação jurídica possível entre usuário e um sujeito externo, ou que envolva interesses jurídicos externos à rede, como é o caso da violação de direitos autorais, pornografia infantil ou disseminação de *fake news*, que por vezes se utilizam de anúncios na plataforma para essa propagação.

Dessa forma, a regulação das redes sociais deve considerar as peculiaridades das diferentes relações jurídicas aqui delimitadas, encontrando limites nos direitos

¹⁰⁶ “Inegável que na relação entre as litigantes a autora, diante de sua notória condição de vítima, equipara-se à figura do consumidor (art. 17 do Código de Defesa do Consumidor).” (RE 1037396).

¹⁰⁷ Nesse mesmo sentido: “APELAÇÃO. AÇÃO DE OBRIGAÇÃO DE FAZER C/C DANOS MORAIS. INSTAGRAM. CONTA DESATIVADA. AUSÊNCIA DE PROVA DE VIOLAÇÃO AOS TERMOS DE USO. Músico renomado teve sua conta do Instagram desativada. Notificada a plataforma, sobreveio pedido de desculpas com a informação de reativação imediata. Entretanto, até a atualidade, a conta não foi reativada, não obstante ordem judicial e aplicação de multa. Plataforma que se valeu de argumentação genérica, sem qualquer supedâneo probatório. Ilícito contratual praticado pelo consumidor não comprovado. Aplicação do CDC à casuística. Desbloqueio da conta confirmado, com majoração das astreintes. DANOS MORAIS. Ocorrência. Situação que superou o mero dissabor. O apelado foi tolhido do uso da rede social Instagram, utilizada para divulgação da atividade profissional. Afronta aos direitos de personalidade. Exagerada mora no restabelecimento do serviço, que não pode ser admitida como mero transtorno. Quantia de R\$ 13.265,00 fixada em primeiro grau, que é mantida. SUCUMBÊNCIA. Manutenção dos honorários recursais. Fixação no patamar máximo legal. RECURSO IMPROVIDO.” (TJSP; Apelação Cível 1109557-32.2019.8.26.0100; Relator (a): Rosangela Telles; Órgão Julgador: 31ª Câmara de Direito Privado; Foro Central Cível - 38ª Vara Cível; Data do Julgamento: 11/05/2021; Data de Registro: 11/05/2021)

fundamentais e nas regras aplicáveis a cada uma das relações, a depender de sua natureza.

Como visto, tanto na relação jurídica usuário-plataforma quanto usuário-usuário, o exercício da liberdade de expressão se submete a um controle primário feito diretamente pela plataforma por meio da moderação de conteúdo, que se passa a analisar na sequência.

3.2 A AUTORREGULAÇÃO: MODERAÇÃO DE CONTEÚDO

Como visto, a mudança social provocada pelas redes sociais pela transformação na infraestrutura da liberdade de expressão colocou as empresas privadas em uma posição não apenas de meras intermediárias de conteúdo, mas de controladoras dele. As empresas decidem, de forma independente, o que pode ou não ser divulgado e quais conteúdos devem ser ou não impulsionados, por meio da moderação de conteúdo.

As plataformas de redes sociais se encontram em um espaço híbrido entre o mero condutor e o produtor de conteúdo. Isso porque elas não apenas transmitem o conteúdo, como por exemplo em uma empresa de telefonia ou e-mail, mas também não são responsáveis pela produção direta dele, como nas mídias tradicionais. Desse modo, as plataformas estão nesse meio termo porque, como intermediárias, determinam onde, como (alcance) e quais conteúdos irão veicular. É aqui que reside a moderação de conteúdo.

Essas decisões tomadas pelas plataformas não decorrem apenas das pressões diretas e indiretas do Estado, como por exemplo uma ordem judicial de restrição, de modo que elas não podem ser vistas unicamente como instrumentos estatais para atingir fins públicos, mas como uma nova esfera de poder que toma decisões. Essas decisões nem sempre estão alinhadas com a ordem constitucional de um determinado Estado, e sim com as políticas internas da empresa.

A motivação para a realização da moderação de conteúdo pode até decorrer de pressões estatais de natureza regulatória, como a responsabilidade civil, contudo, dentro do seu modelo de negócio, a redução de conteúdos danosos se mostra necessária para evitar um ambiente hostil e violador de direitos, pois os anunciantes podem considerá-lo prejudicial aos seus negócios (BATTISTI, 2023).

Assim, pode-se conceituar a moderação de conteúdo de uma perspectiva geral e de uma perspectiva específica. De forma geral, é “o instrumento pelo qual as plataformas da internet exercem algum tipo de controle sobre as manifestações de seus usuários, dispondo quanto à forma como elas virão à luz na rede sem nunca, entretanto, tomá-las como sua.” (RAMOS, 2021, p. 137).

Essa moderação pode se dar por meio de escolhas estruturantes que definem o perfil da plataforma, por exemplo: quanto ao formato (se a expressão será apenas em texto, imagens ou vídeos); quanto à extensão (se haverá ou não limite de caracteres e tempo de vídeo); quanto ao acesso (se a rede será aberta, podendo o conteúdo ser visto por qualquer pessoa, ou se depende de um cadastro prévio, obrigando o indivíduo a criar um perfil caso queira acessá-lo), dentre inúmeros outros detalhes que irão balizar e limitar o modo como as pessoas poderão se expressar.

Já de uma perspectiva específica, a moderação é “a prática de estabelecer e aplicar um conjunto de regras que determinam aquilo que os usuários de uma plataforma podem ou não dizer”, por meio da elaboração de padrões prévios de comportamento e da aplicação desses padrões. (RAMOS, 2021, p. 141).

Essas regras impõem limites ao exercício da liberdade de expressão na arena digital, atingindo o debate público. Contudo, dada a transnacionalidade das plataformas, as regras são aplicadas indistintamente em centenas de países com culturas, línguas, tradições e ordenamentos jurídicos diferentes, que asseguram a liberdade de expressão sob perspectivas diversas, como analisado no item 1.3. Essa capacidade das plataformas de moderar conteúdos, numa perspectiva específica, determinando o que pode ou não ser dito, as coloca no centro do dilema entre a liberdade de expressão e a censura.

Esse dilema se dá, em geral, pela manifestação de um usuário que desencadeia uma discussão sobre o exercício legítimo ou não da liberdade de expressão dentro de um contexto de uma determinada sociedade, além da consequente decisão pela manutenção ou retirada do conteúdo, feita pela plataforma. Essa decisão parte de critérios elaborados pela própria plataforma, que em regra são aplicados de forma indistinta em todo o mundo e não de acordo com a ordem jurídica do país em que a manifestação foi proferida.

Um exemplo de decisão do Facebook que foi considerada pela imprensa mundial como ato de censura foi a postagem feita pelo jornalista norueguês Tom

Egeland no perfil do Jornal *Aftenposten*. Em 2016, ele publicou um artigo sobre o impacto que a fotografia tem sobre a história das guerras e incluiu a foto histórica da menina Kim Phuc correndo nua para fugir de um ataque de Napalm na Guerra do Vietnã. Logo após a publicação, a imagem foi excluída pelo Facebook sob o argumento de que violava os padrões da comunidade¹⁰⁸. Egeland reagiu, escrevendo um artigo criticando a remoção da fotografia e publicando-a novamente, e foi suspenso da rede, ficando proibido de publicar no Facebook por um período. Após grande repercussão, o Facebook informou que quaisquer fotografias expondo nudez total seriam removidas.

O editor chefe do censurado Jornal *Aftenposten*, Espen Egil Hansen, publicou uma carta aberta¹⁰⁹ criticando a atuação do Facebook, que após intensas críticas reconheceu a sua incapacidade de diferenciar uma imagem de nudez infantil, ou mesmo pornografia, de uma imagem histórica. A empresa decidiu abrir uma exceção ao caso, pelo seu caráter icônico, se comprometendo a implementar melhorias em sua programação (RAMOS, 2021). Relembra-se, conforme analisado no item 1.2, que a fotografia, como expressão artística, possui um maior lastro de proteção da liberdade de expressão, que nesse caso foi desconsiderada pelo algoritmo da plataforma.

¹⁰⁸ Disponível em: <https://oglobo.globo.com/economia/facebook-volta-atras-republica-foto-historica-de-menina-de-napalm-20077188>. Acesso em: 11 mai. 2023.

¹⁰⁹ “[...] Escute, Mark, isso é sério! Primeiro, você cria regras que não distinguem entre pornografia infantil e fotografias de guerra famosas. Depois, aplica essas regras sem dar espaço para a ponderação. Finalmente, você até mesmo censura as críticas e as discussões sobre essa decisão – e pune a pessoa que ousa dar voz a essas críticas... A mídia livre e independente tem a importante tarefa de fornecer informações, o que inclui até mesmo divulgar fotos, que, às vezes, podem ser desagradáveis, e que a elite dominante e talvez até cidadãos comuns não conseguem ver ou ouvir, mas que podem ser importantes exatamente por esse motivo... A mídia tem a responsabilidade de considerar a publicação em todos e em cada um dos casos. Isso pode ser uma responsabilidade pesada. Cada editor deve pesar os prós e contras. Esse direito e dever, que todos os editores do mundo têm, não deveria ser enfraquecido por algoritmos programados em seu escritório na Califórnia. A Declaração de Missão do Facebook afirma que seu objetivo é “tornar o mundo mais aberto e conectado”. Na realidade, você está fazendo isso em um sentido totalmente superficial. Se você se recusar a distinguir entre pornografia infantil e fotografias documentais de uma guerra, isso simplesmente promoverá a estupidez e não cumprirá a missão de aproximar os seres humanos entre si. Fingir que é possível criar regras comuns e mundiais para o que pode e o que não pode ser publicado apenas joga areia nos olhos das pessoas...” (Excertos de uma carta aberta de Espen Egil Hansen a Mark Zuckerberg, publicada na primeira página do jornal *Aftenposten* em 8 de setembro de 2016. (Disponível em: <https://www.unesco.org/pt/articles/aftenposten-versus-facebook-uma-polemica-esclarecedora>. Acesso em: 11 mai. 2023). Esse caso foi objeto de painel de debate na Unesco. O vídeo do painel pode ser acessado no seguinte link: <https://www.youtube.com/watch?v=SpQjvM-t19U>. Acesso em: 11 mai. 2023.

Outro exemplo de controvérsia envolvendo as políticas das plataformas e a ordem jurídica de um Estado é a negação do Holocausto. Em 2018, Mark Zuckerberg, CEO do Facebook, afirmou que apesar de achar a negação do Holocausto uma atitude extremamente ofensiva, inclusive por ser judeu, as manifestações de usuários nesse sentido não seriam retiradas da plataforma e que o combate às *fake news* não seria impedir as pessoas de dizerem algo falso, mas impedir que elas se proliferem na plataforma.

Contudo, na Alemanha a negação ao Holocausto é uma conduta ilegal associada ao antissemitismo e, com a aprovação em 2017 da lei alemã (NetzDG) que determina a remoção de conteúdo ilegal em até 24 horas, se estabeleceu um conflito entre a ordem jurídica alemã e a política do Facebook (RAMOS, 2021). Na ocasião, a solução encontrada pelo Facebook foi o geobloqueio, que consiste na restrição de determinados conteúdos de acordo com a origem do acesso à plataforma. Dois anos depois, em 2020, o Facebook alterou sua política de discurso de ódio, proibindo qualquer conteúdo que negue ou distorça o Holocausto¹¹⁰.

A nova posição do Facebook se amolda ao modelo alemão e brasileiro de controle da liberdade de expressão, que permite um controle quanto ao conteúdo e não apenas dos efeitos da expressão, e vai de encontro com a posição americana, que, de acordo com o entendimento da sua Suprema Corte, tolera discursos de ódio quando não possam causar um dano iminente, como analisado no item 1.3 deste trabalho. Dessa forma, observa-se que as decisões de moderação de conteúdo podem entrar em conflito com a regulação dos estados nacionais, os quais inclusive podem influenciar a alteração das políticas das empresas.

Diante desses conflitos, entende-se necessário analisar mais a fundo o funcionamento da moderação de conteúdo, tendo como referência duas plataformas com perfis distintos, o Facebook (Meta)¹¹¹ e o Twitter¹¹², a partir de duas etapas da moderação específica: a definição das regras e a aplicação delas.

¹¹⁰ Disponível em: <https://g1.globo.com/economia/tecnologia/noticia/2020/10/12/facebook-removera-conteudo-que-negue-o-holocausto.ghtml>. Acesso em: 11 mai. 2023.

¹¹¹ O Instagram, por ser um serviço também da Meta Platforms, está incluído nessa análise.

¹¹² Os Termos de Serviço do Twitter estão disponíveis em: <https://twitter.com/pt/tos>. Acesso em: 11 mai. 2023.

Na primeira etapa, quanto à definição das regras, tem-se que as plataformas possuem dois principais documentos públicos¹¹³, com nomenclaturas variáveis, que norteiam a moderação de conteúdo: os “termos de serviço¹¹⁴”, com uma natureza mais jurídica contratual, e os “padrões da comunidade” ou, no caso do Twitter, as “regras do Twitter”, que são os documentos que definem as políticas de conteúdo em uma linguagem mais acessível aos usuários.

O primeiro, como dito, funciona como um contrato, uma vez que depende do aceite do usuário para poder ingressar na plataforma. Em geral, apresenta os serviços fornecidos e a forma como são financiados e define obrigações para o usuário, dentre elas o que ele pode compartilhar e fazer na plataforma, estabelecendo as consequências do seu descumprimento.

No Facebook (Meta), uma das previsões do termo de serviço é o compromisso assumido pelo usuário de não violar os termos e padrões da comunidade¹¹⁵, que é o

¹¹³ Cada plataforma pode contar com inúmeros outros termos e políticas, com regras mais específicas aplicáveis a cada uma das funcionalidades da plataforma, por exemplo, no Facebook há termos específicos dispendo sobre as políticas de pagamento, termos comerciais, sobre o compartilhamento de conteúdos com música, sobre as diretrizes específicas das páginas e grupos, dentre outros.

¹¹⁴ Os Termos de Serviço do Facebook estão disponíveis em: <https://pt-br.facebook.com/terms.php>. Os do Instagram estão disponíveis em: <https://pt-br.facebook.com/help/instagram/581066165581870>. Acesso em: 11 mai. 2023.

¹¹⁵ “2. O que você pode compartilhar e fazer nos Produtos da Meta

Queremos que as pessoas usem os Produtos da Meta para se expressar e compartilhar conteúdo que seja importante para elas, mas não às custas da segurança e do bem-estar de outras pessoas ou da integridade da nossa comunidade. Por isso, você concorda em não adotar o comportamento descrito abaixo nem facilitar ou apoiar que outras pessoas o façam:

1. Você não pode usar nossos Produtos para fazer ou compartilhar conteúdo:

- que viole estes Termos, os Padrões da Comunidade ou outros termos e políticas aplicáveis ao seu uso dos nossos Produtos;
- que seja ilegal, enganoso, discriminatório ou fraudulento (ou que ajude alguém a usar nossos Produtos dessa maneira);
- que não seja de sua propriedade ou que não tenha os direitos necessários para ser compartilhado.
- que possa infringir ou violar os direitos de outra pessoa, incluindo seus direitos de propriedade intelectual, como direitos autorais ou marcas registradas, ou a distribuição ou venda de mercadorias falsificadas ou pirateadas, a não ser que uma exceção ou limitação seja aplicável de acordo com a lei. [...] Podemos remover ou restringir o acesso a conteúdo que viole essas disposições. Podemos também suspender ou desativar sua conta por condutas que violem essas disposições, conforme previsto na Seção 4.B. [...] Se removermos conteúdo que você compartilhou por violação dos Padrões da Comunidade, avisaremos a você e explicaremos suas opções para solicitar outra análise, a menos que você viole de forma grave ou repetida estes Termos ou faça algo que possa expor a nós ou outros a responsabilidades legais; prejudicar nossa comunidade de usuários; comprometer ou interferir na integridade ou operação de qualquer um dos nossos serviços, sistemas ou Produtos; quando formos restritos devido a limitações técnicas; ou quando formos proibidos de fazê-lo por motivos legais. Para informações sobre suspensão ou extinção de conta, consulte a Seção 4.B abaixo. [...] Também podemos remover ou restringir o acesso a conteúdo, serviços ou informações se determinarmos que isso é razoavelmente necessário para evitar ou reduzir impactos jurídicos ou regulatórios adversos para a Meta.” (Disponível em: <https://pt-br.facebook.com/terms.php>. Acesso em: 11 mai. 2023).

segundo documento que merece destaque. Esse documento, que pode variar o nome a depender da plataforma, apresenta verdadeiras regras, em geral divididas em eixos temáticos. Nessa rede, os padrões da comunidade são divididos em seis eixos: “violência ou comportamento criminoso”; “segurança”; “conteúdo questionável”; “integridade e autenticidade”; “respeito à propriedade intelectual”; “solicitações e decisões relativas a conteúdo”.

Cada um desses eixos se desdobra em temáticas específicas contendo regras gerais e específicas. As regras gerais são descrições que fundamentam aquela política, justificando a sua previsão e apresentando definições. Elas são seguidas de regras específicas, divididas em dois graus de restrição: conteúdo não permitido e conteúdo que requer mais informação ou contexto para a tomada de providências, que é permitido com uma tela de aviso ou que é permitido, mas apenas maiores de 18 anos de idade podem vê-lo.

Por exemplo, no eixo temático “conteúdo questionável” tem-se um desdobramento referente ao “Discurso de ódio”. Em um primeiro momento, a plataforma define o que é discurso de ódio e justifica a sua proibição (regra geral); na sequência, nas regras específicas, estabelece aquilo que não pode ser publicado (primeiro grau de restrição) e o que depende de mais informações ou contexto (segundo grau de restrição).

“Não publique:

Nível 1

Conteúdo visando um indivíduo ou grupo de pessoas (incluindo todos os grupos, salvo os que são considerados grupos não protegidos responsabilizados pelo cometimento de crimes violentos ou ofensas sexuais ou que representem menos da metade de um grupo), nos moldes das referidas características protegidas ou status de imigração com:

- Discurso violento ou apoio de forma escrita ou visual.
- Imagem ou discurso degradante sob a forma de comparações, generalizações ou declarações de comportamento não qualificadas (de forma visual ou escrita) voltadas para ou sobre:

[...]

- Estereótipos prejudiciais historicamente ligados à intimidação, exclusão ou violência com base em uma característica protegida, como Blackface; Negação do holocausto; afirmações de que o povo judeu controla instituições financeiras, políticas ou da mídia; e referências a dalits como trabalhadores braçais

Nível 2

- Conteúdo que vise uma pessoa ou um grupo de pessoas com base em características protegidas contendo: [...]

- Expressões sobre ser melhor/pior do que outra característica protegida, incluindo, entre outras: “Eu acredito que os homens são superiores às mulheres.”¹¹⁶

No Twitter, as “regras do Twitter” são divididas em quatro eixos temáticos: “segurança”; “privacidade”; “autenticidade” e “publicidade de terceiros em conteúdo de vídeo”. Esses eixos também se desdobram em temáticas específicas, que apresentam uma regra geral, chamada de “visão geral”, e na sequência regras específicas, determinando o que viola aquela política e apresentando as sanções pelo seu descumprimento.

Por exemplo, no eixo temático “segurança” tem-se o “discurso violento”. A visão geral justifica os motivos da limitação e especifica o que viola a política, como ameaças violentas ou glorificação da violência.

“Visão geral

Você não pode ameaçar, incitar, exaltar ou expressar desejo por violência ou danos.

O Twitter é um lugar onde as pessoas podem se expressar, saber o que está acontecendo e debater sobre questões globais. Contudo, conversas saudáveis não prosperam quando é usado discurso violento para entregar uma mensagem. Como resultado, temos uma política de tolerância zero contra discurso violento, como forma de garantir a segurança dos usuários e evitar a normalização de ações violentas.

O que viola essa política?

Ameaças violentas

Você não pode ameaçar infligir danos físicos a outras pessoas, o que inclui, entre outros, ameaçar matar, torturar, atacar sexualmente ou machucar de alguma outra forma uma pessoa. Isso também inclui ameaçar danificar casas e abrigos civis, ou infraestruturas essenciais para atividades diárias, cívicas ou empresariais.

(...)

O que acontecerá se você violar esta política?

Na maioria dos casos, suspenderemos a conta de forma imediata e permanente se houver outras violações dessa política.

(...)”¹¹⁷.

Observa-se que há uma preocupação comum entre as duas plataformas analisadas de se combater discursos de ódio e comportamentos criminosos. Contudo, alguns comportamentos são tolerados em uma plataforma e em outra não (por

¹¹⁶ Padrões da comunidade disponíveis em: <https://transparency.fb.com/pt-br/policies/community-standards/hate-speech/>. Acesso em: 10 mai. 2023.

¹¹⁷ Regras do Twitter disponíveis em: <https://help.twitter.com/pt/rules-and-policies/violent-speech>. Acesso em: 11 mai. 2023.

exemplo, no Twitter nudez e conteúdos sexuais são permitidos, mas no Facebook são proibidos) (RAMOS, 2021).

Paralelamente a esses documentos públicos, há regras específicas destinadas aos moderadores de conteúdo, mas que são mantidas em sigilo. Sabe-se da sua existência a partir de vazamentos de informações, como foi o caso do *Facebook Files*, noticiado pelo *The Guardian* em 2017¹¹⁸. De acordo com o jornal, há mais de 100 manuais de treinamento interno, planilhas e fluxogramas que estabelecem regras de moderação, além dos padrões da comunidade.

A segunda etapa da moderação refere-se à aplicação das regras. Valendo-se da classificação de Carlos Eduardo Vieira Ramos (2021), essa etapa será analisada dividindo-se em três fases: a primeira é a de detectar o conteúdo que será submetido à moderação; a segunda é a de decidir manter ou não o conteúdo; e a terceira, que ocorre apenas na hipótese de violação das políticas da empresa, é a aplicação de sanções.

A detecção de um conteúdo que será submetido à moderação não é uma tarefa simples, isso porque os conteúdos produzidos na plataforma operam em escala, ou seja, há muito conteúdo se comparado aos recursos de controle existentes, que são limitados. Isso pode ocorrer por um estímulo que parte de dentro da plataforma, como por exemplo a denúncia de usuários por meio de formulários¹¹⁹ que sinalizam a violação das políticas, em colaboração com a rede social, ou a partir de um estímulo externo à plataforma, como pedidos de governos ou uma ordem judicial que determine a remoção de um conteúdo¹²⁰ (RAMOS, 2021).

A moderação ainda pode ser feita antes da sua publicação, de forma preventiva, ou depois, de forma repressiva. Para Ramos (2021), a moderação preventiva é a ideal, uma vez que permite a análise de todo o conteúdo veiculado

¹¹⁸ Disponível em: <https://www.theguardian.com/news/2017/may/21/revealed-facebook-internal-rulebook-sex-terrorism-violence>. Acesso em: 11 mai. 2023.

¹¹⁹ Atualmente no Instagram é possível denunciar uma conta pelos seguintes motivos: perfil falso; pessoa é menor de 13 anos; é *spam*; suicídio, automutilação ou distúrbios alimentares; venda de produtos regulamentados ou ilícitos; nudez ou atividade sexual; símbolos ou discursos de ódio; violência ou organizações perigosas, *bullying* ou assédio; violação de propriedade intelectual.

¹²⁰ O ministro da Justiça do governo Lula, Flávio Dino, fez 546 pedidos para o Twitter para a exclusão de conteúdos e perfis considerados indevidos que fomentam a violência nas escolas. (Disponível em: <https://g1.globo.com/politica/noticia/2023/04/12/seguranca-nas-escolas-ministerio-da-justica-fez-546-pedidos-para-o-twitter-inclusive-exclusao-de-conteudo.ghtml>). Acesso em: 11 mai. 2023).

numa espécie de “revisão editorial”¹²¹ e, por ser realizada antes da postagem, impede que eventual conteúdo indevido possa causar maiores problemas.

A moderação preventiva, em geral, é feita pela compatibilização do conteúdo com um banco de dados de proibições, incluindo fotos e vídeos. Quando se faz o *upload* de um vídeo, geralmente aparece uma mensagem dizendo que ele está sendo processado: é nesse momento que aquele conteúdo está sendo compatibilizado com o banco de dados de proibições¹²² (KLONICK, 2018). Contudo, na moderação preventiva há o risco de uma leitura indevida por parte da plataforma do conteúdo, que, ao impedir a postagem, incorrerá em censura prévia.

A moderação repressiva, por sua vez, ocorre quando a remoção é feita posteriormente à postagem, como no caso relatado da foto da guerra do Vietnã. A detecção pode ser feita pela ação humana de um moderador, ou de forma automática por um software ou algoritmo programado para essa finalidade.

A detecção automática se apresenta como uma solução ao problema da quantidade de conteúdo produzido, o que inviabiliza o controle apenas pela ação humana. Além disso, a automatização afasta a subjetividade da moderação. Contudo, nem sempre o algoritmo consegue diferenciar um conteúdo ilegítimo de um sarcasmo ou mensagem de protesto, porque as particularidades das ações humanas nem sempre conseguem ser traduzidas em simples comandos e instruções (RAMOS, 2021).

Após a detecção, o conteúdo é submetido a um processo decisório que pode ser centralizado (feito pela própria plataforma) ou descentralizado (quando a plataforma autoriza os usuários a tomarem decisões sobre o conteúdo de outros usuários). Como exemplo do segundo, tem-se a figura do administrador e moderador dos grupos do Facebook, que tem autonomia para, dentre outras atribuições, aprovar ou recusar publicações e ainda aceitar ou remover pessoas do grupo. Outro exemplo

¹²¹ Expressão de Tarleton Gillespie (2018) em *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*.

¹²² Uma das técnicas utilizadas pelo Facebook e Twitter é o PhotoDNA, que é um algoritmo criado pela Microsoft que detecta fotos de pornografia infantil a partir da técnica de *hashing*. “Funciona assim: a imagem é convertida para preto e branco, depois redimensionada e então dividida em várias células. Cada uma dessas células tem suas cores e gradientes analisados individualmente, gerando o DNA da foto. Caso um arquivo gere um hash já registrado previamente em um banco de dados associado a imagens de teor pedófilo, um alerta é gerado.” (Disponível em: <https://tecnoblog.net/arquivo/162694/photodna-pornografia-infantil-microsoft/>. Acesso em: 11 mai. 2023).

é a possibilidade de o próprio usuário bloquear outros usuários, impedindo que esses possam acessar o seu perfil e interagir¹²³ (RAMOS, 2021). Essa capacidade decisória dos usuários é objeto de discussão, especialmente em se tratando de figuras públicas, como agentes políticos, vide caso recente do ex-presidente do Estados Unidos da América, Donald Trump¹²⁴.

A decisão ainda pode ser feita de forma manual, como resultado da atuação dos moderadores de conteúdo, ou automática, pelos algoritmos. Destaca-se que o modo como o conteúdo é detectado não se confunde com o agente responsável pela decisão, isso porque um conteúdo pode ser detectado e submetido à moderação por um algoritmo, mas a decisão pode ser feita de forma manual por um moderador ou vice-versa.

Na decisão manual, uma vez detectado o conteúdo irregular, ele é submetido a uma espécie de fila, aguardando a avaliação de um moderador. Kate Klonick (2018) identificou, a partir de entrevistas, que no Facebook existem três níveis de moderadores de conteúdo: os moderadores de nível 3, que fazem a maior parte da revisão diária e geralmente trabalham em centrais distribuídas em todo o mundo, no formato de *call centers*; os moderadores de nível 2, que em geral supervisionam os revisores de nível 3 ou são moderadores especializados em julgar conteúdos e revisar eventual conteúdo priorizado; e os moderadores de nível 1, que em geral são advogados que trabalham na sede jurídica e política da plataforma, tratam de assuntos mais relevantes e são responsáveis pela elaboração das políticas da empresa.

Os moderadores de nível 2 e 3 decidem se o conteúdo é mantido ou retirado a partir de treinamentos recebidos e, como já relatado, não há informação disponibilizada pelas plataformas sobre como se dá efetivamente a moderação, o que gera um problema sensível para a regulação, que é a falta de transparência. Como destaca Kate Klonick (2018):

¹²³ Disponível em: <https://pt-br.facebook.com/help/901690736606156>. Acesso em: 11 mai. 2023.

¹²⁴ A Suprema Corte Americana proibiu o ex-presidente Donald Trump de bloquear opositores no Twitter, a decisão considerou que a rede social é utilizada como meio de comunicação oficial e, portanto, deve estar acessível a todo o público. (Disponível em: <https://veja.abril.com.br/mundo/suprema-corte-dos-eua-proibe-trump-de-bloquear-opositores-no-twitter/>. Acesso em: 11 mai. 2023).

After content has been flagged to a platform for review, the precise mechanics of the decisionmaking process become murky. The “army” of content moderators and “[t]he details of moderation practices are routinely hidden from public view,” write Catherine Buni and Soraya Chemaly.²⁸⁶ “[S]ocial media companies do not publish details of their internal content moderation guidelines; no major platform has made such guidelines public.”²⁸⁷ These internal guidelines also change much more frequently than the public Terms of Service or Community Standards. (p. 1639)¹²⁵.

Além disso, esse modelo de moderação tem sofrido críticas pela própria metodologia de trabalho, que exige que os moderadores, apenas com os treinamentos recebidos, os quais se desconhece, tomem decisões que envolvam a limitação de direitos fundamentais, em especial a liberdade de expressão. Essas limitações encontram divergência tanto na doutrina quanto na jurisprudência, além das diferentes interpretações que tais direitos podem sofrer em decorrência dos diversos ordenamentos jurídicos e do próprio relativismo cultural e linguístico (RAMOS, 2021). É de se pensar: será que um jovem na Índia saberia diferenciar um discurso preconceituoso de um discurso de ódio na realidade brasileira?

As condições de trabalho desses moderadores também têm sido objeto de denúncia e controvérsia e são relevantes na medida em que podem refletir diretamente na qualidade dos julgamentos. De acordo com o portal *The Verge*¹²⁶, os moderadores tomam em média 400 decisões por dia de trabalho e recebem \$15,00 (quinze) dólares por hora. Lembra-se que esses moderadores entram em contato diariamente com o que há de pior nas redes sociais, como atos de violência, automutilação, terrorismo e pedofilia. Tais condições de trabalho têm desencadeado inúmeros transtornos psicológicos.

Assim, observa-se que a moderação manual de conteúdo, além de não conseguir atender a demanda em escala, apresenta problemas quanto à ausência de informação sobre os treinamentos e regras decisórias, bem como às condições de

¹²⁵ “Depois que o conteúdo é sinalizado para uma plataforma para revisão, a mecânica precisa do processo de tomada de decisão torna-se obscura. O ‘exército’ de moderadores de conteúdo e ‘[os] detalhes das práticas de moderação são rotineiramente escondidos da vista do público’, escrevem Catherine Buni e Soraya Chemaly. “[as] empresas de mídia social não publicam detalhes de sua moderação interna de conteúdo orientações; nenhuma plataforma importante tornou essas diretrizes públicas.’ Essas diretrizes internas também mudam com muito mais frequência do que os Termos de serviço públicos ou os Padrões da comunidade.” (KLONICK, 2018, p. 1639, tradução nossa).

¹²⁶ De acordo com a denúncia feita pelo portal *The Verge*, os funcionários são submetidos a condições precárias de trabalho, as quais tem desencadeado inúmeros transtornos psicológicos, inclusive estresse pós-traumático. (Disponível em: <https://www.theverge.com/2019/6/19/18681845/facebook-moderator-interviews-video-trauma-ptsd-cognizant-tampa>. Acesso em: 11 mai. 2023).

trabalho às quais os moderadores são submetidos, que podem repercutir na qualidade de suas decisões.

Quanto à decisão automática, feita pelos algoritmos, o problema tem sido o mesmo, que é a falta de transparência quanto à sua programação, a presença de vieses interpretativos, bem como a falta de capacidade de compreender contextos de conteúdos.

Na página do Facebook (Meta), as informações sobre a moderação de conteúdo são superficiais, mas, em resumo, ela é feita prioritariamente pelos algoritmos, que fazem a compatibilização do conteúdo com os padrões da comunidade e, caso seja necessário um maior contexto, a moderação é transferida para os moderadores humanos:

Como a tecnologia detecta violações

As pessoas publicam bilhões de conteúdo no Facebook e no Instagram todos os dias. Mesmo com milhares de analistas pelo mundo, é impossível analisar tudo isso sem ajuda. É aí que a inteligência artificial da Meta entra em jogo. Nossa tecnologia detecta e remove de modo proativo a grande maioria do conteúdo violador antes que alguém o denuncie.

Todos os dias, removemos milhões de publicações e de contas violadoras do Facebook e do Instagram. Grande parte disso acontece de modo automático, com a tecnologia trabalhando nos bastidores para remover conteúdos violadores, geralmente, antes que alguém os veja. Outras vezes, nossa tecnologia detecta os conteúdos possivelmente violadores, mas envia para as equipes de análise verificarem e tomarem as medidas necessárias.

Esse trabalho nunca termina. As pessoas continuarão tentando escapar da nossa tecnologia. Por isso, precisamos continuar melhorando.

Quando alguém publica no Facebook ou no Instagram, nossas tecnologias verificam se o conteúdo vai contra os Padrões da Comunidade do Facebook e as Diretrizes da Comunidade do Instagram. Na maioria dos casos, a identificação é algo simples. Fica claro se uma publicação viola nossas políticas ou não.

Contudo, algumas vezes a identificação é mais difícil. Talvez o sentimento por trás da publicação não seja claro, a linguagem seja particularmente complexa ou as imagens dependam muito do contexto. Nesses casos, realizamos análises adicionais usando pessoas¹²⁷.

Uma vez decidido que o conteúdo viola as regras de uso da plataforma, passa-se à última fase, que é a sancionatória. As consequências podem ser desde a retirada do conteúdo, acompanhado da proibição para o usuário punido de se publicar algo em determinado período, até mesmo o banimento do usuário na plataforma.

¹²⁷ Disponível em: <https://transparency.fb.com/pt-br/enforcement/detecting-violations/technology-detects-violations/>. Acesso em: 11 mai. 2023.

Algumas manifestações submetidas à moderação podem não ser removidas, mas identificadas como sensíveis, como aquelas com cenas de violência, mas que não chegam a violar as políticas de conteúdo. Nesse caso, a moderação implica na identificação desse tipo de conteúdo para os usuários, que são alertados, e para que possam ser efetivamente visualizados se submetem a um duplo clique, por exemplo (RAMOS, 2021).

Em 2019, o Facebook anunciou um conjunto de medidas para combater a desinformação: assim, quando uma postagem ou um *story* contiver um conteúdo falso, ele ficará borrado e terá um rótulo de informação falsa. Contudo, o conteúdo não será removido e poderá ser compartilhado. O usuário que tentar acessá-lo terá acesso a links para a agência de checagem de notícias sobre aquele conteúdo. De acordo com o Facebook, a aferição da veracidade do conteúdo será feita por uma agência de checagem de fatos independente e parceira, mas não apresenta maiores informações sobre quais empresas efetivamente farão essa verificação¹²⁸.

Quanto às sanções, observa-se que, no Facebook, os usuários que repetidamente tiverem conteúdos removidos enfrentarão uma intensificação gradual da punição: dois posts removidos em um determinado período podem significar a suspensão da conta por vinte e quatro horas, assim como outras violações dos padrões da comunidade podem resultar em banimentos totais (KLONICK, 2018). Já o Youtube prevê a desmonetização como uma forma de sanção, ou seja, os canais que violarem as políticas de conteúdo não poderão executar anúncios em seus canais¹²⁹.

Contudo, as decisões proferidas pelas plataformas não são devidamente motivadas. Em alguns casos, o usuário sequer sabe qual o motivo da decisão de remoção do seu conteúdo e qual política foi efetivamente violada, o que inviabiliza o exercício do seu direito de defesa. Em algumas situações, no Instagram, a plataforma indica apenas a diretriz violada (por exemplo, se foi nudez ou violência).

A partir da decisão, o usuário pode solicitar uma análise para a plataforma, mas não há qualquer campo disponível para argumentar. Caso a plataforma reveja o seu parecer, o conteúdo será publicado novamente; caso contrário, a remoção é mantida. Se o usuário ainda discordar, poderá tentar se qualificar para fazer uma apelação ao

¹²⁸ Disponível em: <https://about.fb.com/br/news/2019/10/ajudando-a-protoger-as-eleicoes-dos-eua-de-2020/>. Acesso em: 11 mai. 2023.

¹²⁹ Disponível em: <https://support.google.com/youtube/answer/6162278?hl=pt-BR>. Acesso em: 11 mai. 2023.

Comitê de Supervisão (*Oversight board*). Esse comitê, em operação desde 2020, funciona como uma última instância recursal e julga apenas casos emblemáticos e controvertidos, ou seja, nem todos os conteúdos e decisões estão sujeitos à revisão por ele¹³⁰.

Assim, a moderação de conteúdo¹³¹ se apresenta como uma forma de governança privada, uma vez que governa e cria regras tanto materiais quanto procedimentais, na medida em que, respectivamente, decide aquilo que os usuários podem ou não publicar e define o procedimento para essa apuração.

Observa-se que há uma preocupação por parte das plataformas com o exercício da liberdade de expressão, mas justificada pelos interesses corporativos de se criar um ambiente digital que seja favorável ao seu modelo de negócio, uma vez que um ambiente no qual os usuários se sintam desconfortáveis para se expressar gera perda de visibilidade, confiança e, conseqüentemente, de lucro. Desse modo, a moderação de conteúdo se faz a partir de uma “filosofia corporativa da liberdade de expressão”, que visa equilibrar a liberdade de expressão com os interesses de mercado. (KLONICK, 2018, p. 1626).

Além disso, as regras processuais de moderação não asseguram um devido processo no meio digital e essas falhas processuais podem ser identificadas pelos três princípios de Santa Clara. Em 2018, em Santa Clara, Califórnia, um grupo de defensores, acadêmicos e especialistas lançou um conjunto de três princípios operacionais sobre a melhor forma de obter transparência e responsabilidade significativas das plataformas quanto a moderação de conteúdo. São eles: princípio 1, números; princípio 2, aviso; e princípio 3, recurso¹³².

Segundo o The Santa Clara Principles¹³³ (2018), o princípio 1 estabelece que “As empresas devem publicar informações sobre conteúdos e contas acionadas, discriminadas por país ou região, se disponível, e categoria de regra violada, em cada uma destas dimensões” (s/p). Esses dados devem ser fornecidos a partir de relatórios

¹³⁰ Disponível em: <https://transparency.fb.com/pt-br/oversight/appealing-to-oversight-board/>. Acesso em: 11 mai. 2023.

¹³¹ Disponível em: https://pt-br.facebook.com/help/instagram/675885993348720/?helpref=related_articles. Acesso em: 11 mai. 2023.

¹³² Em 2020/2021, esses especialistas desenvolveram uma segunda parte desses princípios (princípios de Santa Clara 2.0), incorporando aos princípios operacionais, também princípios fundamentais que devem ser levados em consideração pelas empresas.

¹³³ Disponível em: <https://santaclaraprinciples.org/>. Acesso em: 11 mai. 2023.

regulares, numa espécie de *accountability*, permitindo a realização de estudos a respeito das decisões a partir dos dados fornecidos.

O princípio 2 estabelece que

As empresas devem notificar cada usuário cujo conteúdo for removido, cuja conta for suspensa ou quando alguma outra ação for tomada devido ao descumprimento das regras e políticas do serviço, sobre o motivo da remoção, suspensão ou ação. Quaisquer exceções a esta regra, por exemplo, quando o conteúdo for *spam*, *phishing* ou *malware*, devem ser claramente definidas nas regras e políticas da empresa. (THE SANTA CLARA PRINCIPLES, 2018, s/p).

Esse princípio, além de assegurar maior transparência, permite a participação do usuário no procedimento de moderação, dando condições para que possa exercer seu direito de defesa.

O princípio 3 complementa o anterior e estabelece uma obrigação para as empresas de disponibilizar processos de explicação, revisão e recurso: “Os usuários devem ter acesso suficiente aos canais de suporte que forneçam informações sobre a decisão de ação e os processos de apelação disponíveis assim que a decisão de ação inicial for tomada” (THE SANTA CLARA PRINCIPLES, 2018, s/p). Ele ainda prevê a necessidade de agilidade das revisões (razoável duração do processo), a existência de uma revisão humana da decisão por um indivíduo que não esteja envolvido na decisão inicial, bem como que o revisor esteja familiarizado com o idioma e o contexto cultural sobre o conteúdo recorrido. Prevê ainda a possibilidade, a longo prazo, de se garantir processos de revisão independentes.

Em conclusão, tem-se que a autorregulação privada realizada pelas plataformas não é suficiente para garantir o livre fluxo de informações e ideias na sociedade, dada a ausência de neutralidade, transparência, *accountability*, devido processo e direito de defesa dos usuários, motivo pelo qual a intervenção estatal se mostra necessária.

3.3 A REGULAÇÃO ESTATAL E A AUTORREGULAÇÃO REGULADA DAS REDES SOCIAIS

A mudança social provocada pelo surgimento das plataformas de redes sociais, que passam a ser detentoras da infraestrutura do exercício da liberdade de expressão

sem produzir diretamente os conteúdos, acabou por desencadear a necessidade de uma mudança regulatória, uma vez que alterou o principal destinatário da regulação, do indivíduo para as plataformas. Isso porque, como visto (item 2.3), o aumento no número de fala acabou por dificultar uma ação estatal direta sobre os indivíduos, de modo que é mais viável para o Estado controlar as plataformas, por meio de estímulos regulatórios, do que controlar aqueles que realmente produziram o conteúdo.

Assim, a regulação estatal da liberdade de expressão no cenário digital implica na regulação das redes sociais (*new school*), posto que o modelo tradicional regulatório, antes dualista – Estado de um lado e aquele que se expressa do outro (*old school*) –, não é mais suficiente para garantir o exercício livre e saudável da liberdade de expressão nesse novo ambiente digital.

Essa regulação estatal, que pode ser legislativa, administrativa ou judicial, pode se apresentar por instrumentos diretos, como o bloqueio de sinais de celular e internet¹³⁴ e o bloqueio de aplicativos¹³⁵; e indiretos, por meio do que Jack Balkin (2014) denomina de “censura colateral”. Nesse último caso, os Estados, por intermédio de instrumentos regulatórios, criam incentivos para que as plataformas controlem o conteúdo dos seus usuários, como “censores por delegação”, ou seja, o controle enfatiza a prevenção e não a punição, como ocorre no modelo *old school*¹³⁶ (RAMOS, 2021).

Contudo, a regulação puramente estatal (heteroregulação) das plataformas, apesar da sua vantagem coercitiva, não consegue acompanhar a dinâmica do ambiente digital, apresentando um déficit cognitivo que o impede de alcançar seu objetivo. Se para cada nova tecnologia os legisladores se sentirem impulsionados a realizar alterações no ordenamento jurídico, essas alterações, que podem parecer pertinentes no momento, poderão tornar-se defasadas com a mesma velocidade que

¹³⁴ Sobre o tema, o STF decidiu que é competência privativa da União legislar sobre bloqueadores de sinal de celular e internet (ADI 5.585).

¹³⁵ Sobre o fenômeno de bloqueio de aplicativos, recomenda-se o artigo de De Souza Abreu (2018) *Interrompendo o disruptivo: dando sentido ao bloqueio de aplicativos no Brasil* (Disponível em: <https://policyreview.info/articles/analysis/disrupting-disruptive-making-sense-app-blocking-brazil>. Acesso em: 11 mai. 2023).

Também relembra-se a decisão do Ministro Alexandre de Moraes, atualmente suspensa, que determinou a suspensão do funcionamento do Telegram no Brasil. (Disponível em: <https://portal.stf.jus.br/noticias/verNoticiaDetalhe.asp?idConteudo=483659>. Acesso em: 11 mai. 2023).

¹³⁶ Exemplo disso é o fato de que na Constituição Federal – em um modelo *old school* - eventual abuso na liberdade de expressão deve ser resolvido posteriormente, através do direito de resposta e eventual indenização e não de forma prévia, preventiva.

a tecnologia avança. O Direito, como expressão das normas postas e enquanto produto centralizado no Estado, especialmente na manifestação do Legislativo, apesar de sua legitimidade, não consegue acompanhar as transformações sociais dentro do mundo digital.

Por outro lado, a autorregulação das plataformas, por meio da moderação de conteúdo, tem se mostrado insuficiente para enfrentar os desafios da expressão no meio digital, pois, como visto (item 3.2), os interesses das plataformas nem sempre estão alinhados com a defesa do interesse público ou mesmo com a ordem jurídica de determinado estado. Além disso, a instrumentalização das plataformas para propagação de discursos de ódio e desinformação e as consequências negativas suportadas pela sociedade, bem como as acusações de censura privada pela ausência de transparência dos algoritmos e dos processos de moderação, demonstram que a liberdade de expressão não pode estar à mercê exclusivamente da regulação privada.

Nesse cenário de insuficiência dos dois modelos regulatórios anteriores é que se apresenta a autorregulação regulada ou correção, modelo esse já utilizado nos programas de *compliance*.

Nesse modelo, os entes privados mantêm sua autonomia para definir regras privadas, focando na própria expertise e dinâmica. Contudo, estão sujeitos à atividade fiscalizatória e limitadora do Estado, que cria uma espécie de “moldura”, definindo parâmetros gerais dentro dos quais as próprias redes sociais devem criar soluções para cumprir os comandos estatais. Dessa forma, a autorregulação regulada consegue, por meio de estímulos, induzir as empresas privadas a atingir os interesses públicos.

Para Juliano Maranhão e Ricardo Campos (2020), a autorregulação regulada apresenta-se como a melhor opção regulatória, pois reúne duas características importantes, que são: “(1) a participação do objeto da regulação, que são as plataformas, na implementação dos objetivos públicos, visto que o Estado não possui conhecimento técnico para suprir tal demanda; 2) o estabelecimento de determinados parâmetros a serem seguidos pela instituição da autorregulação regulada, parâmetros esses advindos do interesse público.” (p. 346).

Nesse mesmo sentido, Lawrence Lessig (1999) destaca que a regulação estatal da internet deve ser por ação legislativa indireta, cabendo ao Estado

concretizar *standards* de direitos fundamentais dentro desse ambiente virtual marcado por relações privadas. Assim, uma forma de regulação indireta é a autorregulação regulada, que se coloca como um modelo de direito proceduralizado, diferente dos modelos tradicionais (ABBOUD; CAMPOS, 2020).

A crise do direito regulatório, dada a complexidade social provocada pelas novas tecnologias, coloca o Estado em uma posição de não possuir o conhecimento necessário para a tomada de decisão e leva-o a buscar formas de angariar esse conhecimento da sociedade, aqui especialmente junto às plataformas, nos *standards* normativos dessas organizações. É nessa aquisição de conhecimento que desponta o modelo de direito proceduralizado em que há uma cooperação entre Estado regulador e os atores da sociedade que estão sendo regulados. Para George Abboud e Ricardo Campos (2020):

“Regras procedurais visam incorporar no direito uma dimensão de abertura para um constante processo de aprendizagem para temas complexos, nos quais o conhecimento para decisão não se encontra nem na norma posta legitimamente pelo parlamento, nem em princípios abstratos, mas no procedimento estabelecido no direito posto. O conhecimento para decisão, que em âmbitos complexos das novas tecnologias não se encontram nem na norma posta, nem em princípios abstratos, passa a ser gerado dentro de um procedimento preestabelecido. (p. 143)

Dessa forma, a autorregulação regulada, como modelo de proceduralização do direito, tem como fundamento central a “flexibilização e capacidade de aprendizagem criadas dentro de uma temporalização do processo de tomada de decisão visando a geração de conhecimento em setores sociais de alta tecnologia” (ABBOUD; CAMPOS, 2020). Ademais, esse modelo de se regular uma autorregulação coaduna-se com o papel indutor do Estado regulador da atividade econômica, previsto no art. 174 da CF/88¹³⁷.

Assim, Estados, organizações públicas e privadas ao redor do mundo estão debruçadas sobre a tarefa de buscar soluções para esse desafio regulatório, buscando inclusive reformular as principais regulações inauguradas no primeiro momento da internet (*web 1.0*), nos anos de 1990 a 2000.

¹³⁷ “Art. 174. Como agente normativo e regulador da atividade econômica, o Estado exercerá, na forma da lei, as funções de fiscalização, incentivo e planejamento, sendo este determinante para o setor público e indicativo para o setor privado.” (CF/88).

Nos Estados Unidos da América está em discussão, no Congresso Americano, a reforma da *section 230*, promulgada como parte do *Communications Decency Act* (CDA) de 1996; um dos marcos legais mais importantes da internet, que estabelece um modelo de regulação das atividades das plataformas do tipo *content neutrality*, inspirada nos ideais liberais do modelo norte-americano de proteção da liberdade de expressão. Essa norma define a chamada “regra do bom samaritano” ou “bloqueio do bom samaritano”, e é dividida em duas seções.

A seção 1 estabelece uma imunidade para as plataformas em relação ao conteúdo produzido pelos usuários; a seção 2 concede imunidade para que as plataformas façam a moderação de conteúdos gerados pelos usuários quando considerarem, de boa-fé (daí a ideia do bom samaritano), que há conteúdo obsceno, lascivo, imundo, excessivamente violento, assediante ou de outra forma censurável, seja tal material protegido constitucionalmente ou não. Essa imunidade se dá na medida em que a parte prejudicada pela remoção de um determinado conteúdo não pode responsabilizar o provedor (SOUZA; LEMOS, 2016).

Assim, as plataformas são imunes – não podem ser responsabilizadas – tanto pelos conteúdos produzidos por terceiros quanto pelas decisões tomadas na moderação de conteúdo. Contudo, esse modelo tem sido objeto de questionamento perante o Congresso Americano, especialmente após os escândalos envolvendo a polêmica eleição de Donald Trump e a contratação da *Cambridge Analytica* (vide item 2.2), bem como o advento da *web 2.0*, no qual se atestou que as plataformas não atuam de forma neutra, mas sim influenciam diretamente a percepção, o pensamento e a própria opinião pública.

Ao contrário do modelo americano, as legislações europeias, em especial a lei alemã (NetzDG) e a Lei de serviços digitais (DAS) da União Europeia, adotam o modelo do tipo *process-based regulation*, ou seja, o modelo proceduralizado de autorregulação regulada. Isso estabelece uma regulamentação procedimental do discurso no ciberespaço, baseada em processos de identificação, avaliação e controle de riscos, que devem ser realizados, documentados e (geralmente) auditados. Para Ilton Norberto Robl Filho (2023)¹³⁸, o *process-based regulation* é um tipo de autorregulação regulada

¹³⁸ Disponível em: <https://www.conjur.com.br/dl/nota-tecnica-idp-fake-news.pdf>. Acesso em: 11 mai. 2023.

na qual são instituídas obrigações positivas para os provedores, com o intuito de criar incentivos para que eles próprios, mediante supervisão independente, aprimorem a transparência nas decisões de moderação de conteúdo e assumam compromissos de cautela no tratamento de manifestações ilícitas na internet. (p. xx).

No Brasil, a Lei 12.965/2014, apelidada de Marco Civil da Internet, é o instrumento normativo em vigor que estabelece princípios, garantias, direitos e deveres para o uso da Internet no Brasil, mas que não tem conseguido enfrentar com eficiência todos os desafios decorrentes do exercício da liberdade de expressão nas redes sociais, justamente por focar em um modelo tradicional de regulação, na medida em que não define regras concretas para a moderação de conteúdo¹³⁹.

Dessa forma, o projeto de lei 2630/2020, chamado de PL das *Fake News*, que propõe a Lei de Transparência e Responsabilidade na Internet, busca reformular o modelo regulatório do ambiente digital no cenário nacional ao incorporar o modelo proceduralizado da autoregulação regulada. O projeto inicial é de autoria do Senador Alessandro Vieira, tendo como relator na Câmara o Deputado Orlando Silva, e aguarda a votação em plenário pela Câmara, já tendo sido aprovada pelo Senado. Contudo, sofreu inúmeras alterações: num recorte de uma semana (18/04/2023 – 25/04/2023), o projeto, que conta com 60 artigos, teve a redação de 27 deles alterada¹⁴⁰.

Nesse sentido, mostra-se relevante para a pesquisa a análise pontual do referido projeto, que se faz a partir da versão final do substitutivo apresentado no dia 27 de abril de 2023 à Câmara dos Deputados¹⁴¹, e dos seguintes desafios enfrentados pela liberdade de expressão nas redes sociais e levantados durante a pesquisa: o anonimato; as *fake news* e os discursos de ódio (item 2.2), juntamente com a

¹³⁹ Lembra-se que a Medida Provisória 1068/2021 objetivava alterar o “Marco Civil da Internet”, de maneira a explicitar direitos e garantias dos usuários de redes sociais, entre os quais a necessidade de os provedores indicarem justa causa e motivarem decisões relacionadas à moderação de conteúdo. Contudo, a referida MP foi rejeitada por dispor sobre conteúdo que não pode ser objeto de Medida Provisória, bem como por tratar de matéria constante do Projeto de lei 2630/2020, que na época já havia sido aprovado no Senado. (Disponível em: https://www.planalto.gov.br/ccivil_03/_Ato2019-2022/2021/Congresso/adc-58-mpv1068.htm. Acesso em: 11 mai. 2023).

¹⁴⁰ Tabela comparativa das versões recentes do PL 2630, elaborada por Vozes da Regulação: https://docs.google.com/spreadsheets/u/0/d/1Wv5NJ3qaOAWRE_4T07pSV0F4sfXnfVeE/htmlview. Acesso em: 11 mai. 2023.

¹⁴¹ Versão consolidada, submetida à Câmara. Disponível em: https://www.camara.leg.br/proposicoesWeb/prop_mostrarintegra?codteor=2265334. Acesso em: 11 mai. 2023.

moderação de conteúdo (3.2); e a responsabilidade civil das plataformas, verificando-se o enfrentamento dado pelo projeto de lei e se ele de fato avançou na implementação do modelo de autoregulação regulada.

Como destacado no item 2.2, o anonimato tem sido considerado um dos desafios para o exercício da liberdade de expressão nas redes sociais, pois, considerando o binômio liberdade-responsabilidade, a identificação daquele que se expressa é necessária para a apuração de responsabilidades. Em geral, discute-se o anonimato a partir de um pensamento binário, ou entende-se que é possível uma completa vinculação entre a expressão e o indivíduo, permitindo a sua perfeita identificação, ou qualquer expressão desvinculada de uma identidade é considerada como anônima.

Contudo, para Miriam Wimmer e Lucas Borges de Carvalho (2022), o anonimato é apenas uma das várias formas de gestão da identidade e por isso deve ser mais bem interpretado, diante de sua vedação constitucional. Na análise desses autores, a Constituição Federal, ao vedar o anonimato, não impõe a obrigação de identificação completa dos indivíduos em todas as circunstâncias; ao contrário, ela admitiria o anonimato como instrumento de garantia da liberdade de expressão.

O ordenamento jurídico brasileiro prevê diversas formas de gestão da identidade, que variam desde a vedação total à sua identificação até a sua divulgação obrigatória. Por exemplo, em se tratando do exercício de direitos políticos, a Constituição assegura o sigilo do voto, logo, há uma vedação absoluta à identificação do eleitor e à sua manifestação política-eleitoral. Esse sigilo se justifica para assegurar a livre escolha dos candidatos pelos eleitores, sem medo de ameaças.

Em contrapartida, no processo de julgamento de *impeachment*, o Supremo Tribunal Federal entendeu que o voto dos deputados e senadores deve se dar por meio de escrutínio aberto, assegurando maior transparência e publicidade às decisões¹⁴².

O sigilo da fonte também é um instrumento de gestão constitucional da identidade, na medida em que a fonte permanece no anonimato como garantia do exercício da liberdade de informação. A lei de proteção à vítima e testemunhas

¹⁴² ADPF 378/MC DF, 2015. Disponível em: <https://redir.stf.jus.br/paginadorpub/paginador.jsp?docTP=TP&docID=10444582>. Acesso em: 11 mai. 2023.

também determina a preservação da identidade, imagem e dados pessoais da pessoa protegida.

Há ainda o uso de pseudônimos como forma de omissão da identidade, comumente usados no meio artístico ou literário (que são formas de expressão) e que inclusive têm proteção legal, desde que utilizados para atividades lícitas. Pode-se mencionar o caso de Chico Buarque, que utilizou o pseudônimo Julinho da Adelaide para driblar os censores de suas canções durante o período da ditadura militar (BRAUNER, 2005).

Dessa forma, observa-se que a não identificação direta dos indivíduos, que numa análise binária seria considerada anonimato, se apresenta como instrumento viabilizador do exercício da liberdade de expressão e da autonomia política, fortalecendo a democracia, bem como uma expressão da própria privacidade¹⁴³ dos indivíduos, na medida em que diminui a sua vulnerabilidade quanto à exposição dos dados, especialmente no ambiente digital¹⁴⁴.

De outro modo, a ideia de que o ambiente virtual é uma terra sem lei e que o anonimato deve ser prontamente coibido, presumindo-se uma postura de má-fé, não se mostra totalmente adequada atualmente, uma vez que há ferramentas tecnológicas de controle e rastreamento que reduzem o espaço daquela manifestação totalmente anônima, especialmente diante da disseminação de uma cultura de vigilância.

Para David Lyon (2018, p. 165 *apud* WIMMER; CARVALHO, 2022, p. 9), “não existe anonimato em um mundo de Big data”. Assim, o anonimato não deve ser considerado como prática ilegal *per se*, mas deve ser analisado, de forma casuística, posteriormente à expressão, se o seu exercício foi ou não legítimo (WIMMER; CARVALHO, 2022).

Nesse sentido, um dos pontos que gerou maior debate no projeto de lei das *fake news* (PL 2630/2020) foi justamente o que envolvia a identificação dos usuários. Inicialmente, o projeto de lei determinava a necessidade de identificação do usuário (conta identificada) e proibia a criação das chamadas contas inautênticas¹⁴⁵, que

¹⁴³ Outro exemplo é o sigilo da identidade dos doadores e receptores de gametas ou embriões para fins de reprodução assistida, conforme resolução do Conselho Federal de Medicina n. 2168/2017.

¹⁴⁴ Está pendente de julgamento perante o STF o tema 912: *Leading Case* ARE 905149, que analisará a possibilidade de lei proibir o uso de máscaras em manifestações públicas, analisando-se a questão do anonimato e os limites da liberdade de expressão. Até o encerramento dessa pesquisa o caso ainda não havia sido julgado pela corte.

¹⁴⁵ Texto do projeto antes do substitutivo: “Art. 5º Para os efeitos desta Lei, considera-se:

seriam aquelas criadas ou usadas para assumir ou simular a identidade de terceiros. Contudo, os dois conceitos, bem como a proibição de contas inautênticas, foram retirados do projeto na versão em debate atualmente na Câmara, com a seguinte justificativa:

No art. 5º, optamos por excluir a definição de conta identificada, uma vez que o propósito do texto é concentrar esforços no enfrentamento da desinformação, mais do que propor qualquer regime de identificação ampla e geral dos usuários da internet no Brasil. Também excluímos a definição de conta inautêntica, que pode levar a restrições no uso constitucional de pseudônimos e na liberdade de expressão dos usuários. (BRASIL, 2023, p. s/p)

O que se tem na versão atual, no art. 5º II do PL, é o conceito de conta automatizada, que é aquela “conta gerida, total ou preponderantemente, por programa de computador ou tecnologia para simular, substituir ou facilitar atividades humanas” (BRASIL, 2020, s/p). Tais contas não são proibidas, mas o seu uso para a divulgação de notícias falsas pode configurar crime, previsto no art. 50 do PL:

Art. 50. Promover ou financiar, pessoalmente ou por meio de terceiros, mediante uso de conta automatizada e outros meios ou expedientes não fornecidos diretamente pelo provedor de aplicações de internet, divulgação em massa de mensagens que contenha fato que sabe inverídico, que seja capaz de comprometer a higidez do processo eleitoral ou que possa causar dano à integridade física e seja passível de sanção criminal. Pena: reclusão, de 1(um) a 3 (três) anos e multa. (BRASIL, 2023, s/p).

Quanto à identificação do usuário, nas regras de publicidade o PL determina que as plataformas devem requerer a identidade por meio da apresentação de documento válido dos anunciantes, tanto daquele que anuncia quanto daquele que paga pela publicidade na plataforma. Além disso, determina ainda que a identidade do usuário responsável pelo impulsionamento ou o anunciante seja identificado (art. 26¹⁴⁶). Assim, os usuários do serviço devem conseguir identificar de forma clara:

II – conta inautêntica: conta criada ou usada com o propósito de assumir ou simular identidade de terceiros para enganar o público, ressalvados o direito ao uso de nome social e à pseudonímia nos termos desta Lei, bem como o explícito ânimo humorístico ou de paródia;

Art. 6º Com o objetivo de proteger a liberdade de expressão e o acesso à informação e fomentar o livre fluxo de ideias na internet, os provedores de redes sociais e de serviços de mensageria privada, no âmbito e nos limites técnicos de seu serviço, devem adotar medidas para: I – vedar o funcionamento de contas inautênticas” (BRASIL, 2020).

¹⁴⁶ “Art. 26. Os provedores que ofereçam publicidade de plataforma devem identificá-la, de modo que o usuário responsável pelo impulsionamento ou o anunciante sejam identificados.” (BRASIL, 2020, s/p).

a) quais conteúdos constituem um anúncio publicitário, através de sinalização bem visível; b) o beneficiário do anúncio, ou seja, aquele em cujo nome o anúncio publicitário é apresentado; c) a pessoa que paga o anúncio publicitário, caso seja diferente do beneficiário; d) informações pertinentes, direta e facilmente acessíveis a partir do anúncio publicitário, sobre os principais parâmetros utilizados para determinar o destinatário da exibição do anúncio publicitário e, se for o caso, sobre como alterar esses parâmetros. (BRASIL, 2023, s/p).

Dessa forma, tem-se que o projeto de lei, na versão analisada, se afasta um pouco do conceito tradicional binário de anonimato, reconhecendo as diversas possibilidades de gestão da identidade e ressaltando as contas que envolvem publicidade. Assim, espera-se que a versão final do projeto de lei considere as variações legítimas de gestão da identidade, de acordo com os riscos verificados nas diferentes circunstâncias e de modo que a garantia do anonimato, como uma limitação imediata à identificação direta do usuário, seja assegurada como prática legítima, desde que não utilizada para fins ilícitos, o que deve ser verificado *a posteriori* e, no caso concreto, resguardada a possibilidade de identificação.

O enfrentamento da desinformação e dos discursos de ódio está previsto especialmente no art. 7º do projeto de lei, que determina que os provedores (plataformas) têm o dever de “identificar, analisar e avaliar diligentemente os riscos sistêmicos decorrentes da concepção ou do funcionamento dos seus serviços e dos seus sistemas relacionados, incluindo os sistemas algorítmicos” (BRASIL, 2023, s/p). Por riscos sistêmicos, entende-se aqueles conteúdos que não devem ser tolerados nas redes sociais, os quais constam em um rol exemplificativo previsto no § 2º do mesmo art. 7º, quais sejam:

§ 2º A avaliação abrangerá especificamente em cada um dos serviços dos provedores e considerará os riscos sistêmicos, tendo em conta a sua gravidade e probabilidade de ocorrência, e incluirá, no mínimo, a análise dos seguintes riscos:

I – a difusão de conteúdos ilícitos no âmbito dos serviços de acordo com o caput do art. 11;

II – à garantia e promoção do direito à liberdade de expressão, de informação e de imprensa e ao pluralismo dos meios de comunicação social;

III – relativos à violência contra a mulher, ao racismo, à proteção da saúde pública, a crianças e adolescentes, idosos, e aqueles com consequências negativas graves para o bem-estar físico e mental da pessoa;

IV – ao Estado democrático de direito e à higidez do processo eleitoral; e

V - os efeitos de discriminação ilegal ou abusiva em decorrência do uso de dados pessoais sensíveis ou de impactos desproporcionais em razão de características pessoais. (BRASIL, 2023, s/p).

Dessa forma, o legislador elencou 5 grupos de riscos que devem ser identificados e combatidos pelas plataformas: (i) conteúdos ilícitos, especialmente aqueles crimes elencados no art. 11 do PL, que estabelece um dever de cuidado às plataformas, tais como crimes contra o Estado democrático, terrorismo, induzimento a suicídio, racismo, violência contra a mulher, crimes contra a criança e adolescente e condutas que configurem infração sanitária; (ii) riscos à liberdade de expressão (informação e imprensa) e ao pluralismo, nesse grupo podendo ser enquadradas as *fake news*, uma vez que, como já analisado (item 1.3), a proteção da liberdade de informação e expressão implica no combate à desinformação; (iii) riscos envolvendo pessoas em situação de vulnerabilidade ou minorias, aqui podendo ser enquadrados os discursos de ódio (item 1.3); (iv) riscos contra o Estado democrático de direito e o processo eleitoral; e (v) riscos envolvendo discriminação ilegal e abusiva em decorrência do uso de dados pessoais ou em razão de características pessoais (na versão anterior do texto havia um trecho que complementava essas características pessoais, especificando que se referiam a questões de “raça, cor, etnia, identidade de gênero, orientação sexual, deficiência, idade” (BRASIL, 2023, s/p), logo, nesse grupo pode-se enquadrar também os discursos de ódio).

Cabe, portanto, às plataformas identificar, analisar e avaliar os riscos mencionados, seguindo diretrizes que serão fixadas anualmente por um órgão regulador (atualmente o Comitê Gestor da Internet no Brasil – CGI.br, art. 51), e devem levar em consideração alguns fatores que são considerados potencializadores dos riscos, tais como a moderação e recomendação de conteúdo:

Art. 7º § 3º Quando da realização de avaliações de risco, os provedores terão em conta como os seguintes fatores influenciam os riscos sistêmicos referidos no § 2º:

I - a concepção dos seus sistemas de recomendação e de qualquer outro sistema algorítmico pertinente;

II - os seus sistemas de moderação de conteúdos;

III - os termos de uso e a sua aplicação;

IV - os sistemas de exibição de publicidade de plataforma; e

V - a influência da manipulação maliciosa e intencional no serviço, incluindo casos de contas criadas ou usadas com o propósito de assumir ou simular identidade de terceiros para enganar o público, ou explorar o serviço de maneira automatizada. (BRASIL, 2020, s/p).

Além do levantamento dos riscos, as plataformas devem adotar medidas razoáveis, proporcionais e eficazes para atenuá-los (art. 8º). Essa avaliação, bem como as medidas tomadas, deve constar de um relatório chamado de “avaliação de risco sistêmico”, a ser enviado para um órgão fiscalizador (art. 10º). Cabe ainda às plataformas enviarem, semestralmente, o relatório de transparência (art. 23), com a finalidade de informar os procedimentos de moderação de conteúdo, numa espécie de prestação de contas das empresas.

Há também uma outra obrigação para as plataformas, que se refere ao chamado “dever de cuidado”. De acordo com o art. 12 da PL, quando configurada a iminência de riscos descritos no art. 7º ou a negligência ou insuficiência da ação do provedor, poderá ser instaurado, na forma da regulamentação e por decisão fundamentada, protocolo de segurança pelo prazo de até 30 dias. Trata-se de um procedimento de natureza administrativa cujas etapas e objetivos não constam no projeto de lei e deverão ser objeto de regulamentação.

Esse protocolo tem como objetivo criar um mecanismo de fiscalização dos provedores, sendo que a partir de sua instauração as empresas podem ser responsabilizadas civilmente por danos decorrentes de conteúdos gerados por terceiros. É nesse ponto, quanto à responsabilidade civil das plataformas, que reside um dos principais debates da regulação das plataformas.

Um dos instrumentos do que Jack Balkin (2014) chamou de “censura colateral com maior impacto social” é a responsabilidade civil dos intermediários da internet. A depender do modelo regulatório utilizado, é possível a responsabilização das plataformas por conteúdos produzidos pelos usuários, como no modelo europeu, *process-based regulation*, ou a sua imunização, como no modelo americano do tipo *content neutrality*. No Brasil, atualmente, o Marco Civil da Internet¹⁴⁷ adotou, no art. 19, a regra da responsabilidade civil das plataformas apenas após notificação judicial, se aproximando, assim, do modelo europeu.

¹⁴⁷ “Art. 19. Com o intuito de assegurar a liberdade de expressão e impedir a censura, o provedor de aplicações de internet somente poderá ser responsabilizado civilmente por danos decorrentes de conteúdo gerado por terceiros se, após ordem judicial específica, não tomar as providências para, no âmbito e nos limites técnicos do seu serviço e dentro do prazo assinalado, tornar indisponível o conteúdo apontado como infringente, ressalvadas as disposições legais em contrário.” (Lei Nº 12.965/2014).

Para Martins Hartmann e Ingo Sarlet (2019), a adoção da responsabilidade civil das plataformas após notificação judicial é uma opção melhor do que a aplicação de responsabilidade já após notificação extrajudicial. Isso porque, para eles, a responsabilidade após notificação extrajudicial (denúncia de usuário, por exemplo) acabaria criando incentivos econômicos para que as plataformas adotem políticas restritivas à liberdade de expressão como forma de reduzir o risco de indenizar.

Contudo, esse entendimento não é unânime. Para Thaís G. Pascoaloto Venturi (2022), a responsabilidade das plataformas apenas após decisão judicial impõe um ônus econômico e temporal aos usuários lesados, incompatível com a tutela dos direitos fundamentais, que devem ser preservados preventivamente. Ela entende, também, que o modelo atual acarreta às vítimas maior tempo de exposição aos danos e o seu agravamento.

Na Alemanha, adota-se a responsabilidade já após notificação extrajudicial. A Lei de Fiscalização da Rede (NetzDG)¹⁴⁸, aprovada em 2017, determina que as redes sociais devem remover ou bloquear o acesso a qualquer conteúdo manifestamente ilícito no prazo de 24 horas após recebimento de reclamação (§ 3º 2), aqui considerada como qualquer denúncia do conteúdo.

Para David Kaye¹⁴⁹, leis como a alemã, que determinam que as plataformas retirem determinados conteúdos por violarem conceitos vagos ou complexos, como “extremismo”, “discurso de ódio” ou “*fake news*”, acabam por permitir práticas de censura de discursos legítimos.

Diante dessa divergência, está pendente julgamento, perante o Supremo Tribunal Federal, o RE 1037396, com repercussão geral reconhecida, tema 987. Ele discute justamente a constitucionalidade do art. 19 da Lei n. 12.965/2014 (Marco Civil da Internet), que determina a necessidade de prévia e específica ordem judicial de exclusão de conteúdo para a responsabilização civil de provedores de internet, websites e gestores de aplicativos de redes sociais por danos decorrentes de atos ilícitos praticados por terceiros. Trata-se de recurso extraordinário apresentado pelo Facebook Serviços Online do Brasil Ltda. contra acórdão prolatado pela Segunda Turma Recursal Cível do Colégio Recursal de Piracicaba/SP, que reconheceu o dever

¹⁴⁸ Disponível em: <https://www.gesetze-im-internet.de/netzdg/BJNR335210017.html>. Acesso em: 12 mai. 2023.

¹⁴⁹ Disponível em: <https://digitallibrary.un.org/record/1304394#record-files-collapse-header>. Acesso em: 12 mai. 2023.

da plataforma de indenizar pela não retirada de perfil falso da autora, mesmo sem decisão judicial prévia e específica¹⁵⁰.

Nesse mesmo sentido, o Supremo Tribunal Federal, no RE 1057258, tema 533¹⁵¹, deve se manifestar sobre o dever das plataformas de fiscalizar o conteúdo publicado em seus domínios eletrônicos e de retirar do ar informações consideradas ofensivas, sem necessidade de intervenção do Poder Judiciário.

Diferentemente do Marco Civil da Internet, o projeto de lei estabelece a possibilidade de responsabilidade civil e solidária das plataformas por conteúdos gerados por terceiros, independentemente de notificação judicial, em duas situações distintas: (i) quando houver dano a terceiro decorrente de conteúdo que foi objeto de publicidade na plataforma, ou seja, aquele conteúdo que foi patrocinado ou impulsionado no seu alcance em troca de pagamento pecuniário ou estimável em dinheiro (art. 5º, XI); (ii) quando houver dano a terceiro decorrente do descumprimento de um dever de cuidado, durante o protocolo de segurança, definido no art. 12 da PL.

Art. 6º Os provedores podem ser responsabilizados civilmente, de forma solidária: I – pela reparação dos danos causados por conteúdos gerados por terceiros cuja distribuição tenha sido realizada por meio de publicidade de plataforma; e II – por danos decorrentes de conteúdos gerados por terceiros quando houver descumprimento das obrigações de dever de cuidado, na duração do protocolo de segurança de que trata a Seção IV. (BRASIL, 2023, p. s/p).

Assim, as plataformas respondem na esfera civil por danos, de forma solidária, quando de alguma forma lucrarem com o conteúdo indevido ou ainda quando não cumprirem com o dever de cuidado durante o protocolo de segurança, sem prejuízo das sanções administrativas previstas no art. 47, que variam desde a advertência até a suspensão das atividades.

Para Ilton Norberto Robl Filho (2023), esse modelo de responsabilização se aproxima ainda mais do modelo europeu (*process-based regulation*) e se adequa melhor aos problemas do controle dos discursos. Ademais, ele entende que esse

¹⁵⁰ Destaca-se trecho da decisão da turma recursal: “Destarte, condicionar a responsabilização da ré à prévia tomada de medida judicial pela autora, na conformidade do art. 19 do Marco Civil da Internet, fulminaria seu direito básico de efetiva prevenção e reparação de danos patrimoniais e morais, individuais, coletivos e difusos (art. 6º, inciso VI, do Código de Defesa do Consumidor). Logo, a indenização pelos danos morais é medida que se impõe (...)” (STF - RE 1037396 RG/SP).

¹⁵¹ RE 1057258, tema 533. Disponível em:

<https://portal.stf.jus.br/processos/detalhe.asp?incidente=5217273>. Acesso em: 12 mai. 2023.

modelo, que envolve uma conduta preventiva de um dever de cuidado, já encontra previsão no ordenamento jurídico nacional, que define responsabilidade pelo risco da atividade (art. 927, parágrafo único do Código Civil), a exemplo do que ocorre no direito ambiental.

A responsabilidade pelo risco da atividade é muito comum no direito ambiental, buscando-se instrumentos eficientes de alocação dos riscos e para evitar a sua externalização na forma de custos sociais. No referido ramo do direito, os princípios da precaução e da prevenção constituem deveres de internalização dos riscos criados pela atividade no processo produtivo. Caso venha ocorrer algum dano, haverá a presunção da causalidade entre tais riscos e o dano.

No caso do art. 11, do texto substitutivo do PL 2630, a relação entre o dever de cuidado e a responsabilização por danos tem semelhanças com o direito acima citado, mas inova ao inicialmente obrigar os provedores ao desenvolvimento e à aplicação de procedimentos de moderação como forma de comprovar que se desincumbiram previamente do dever primário de *neminem leadere* e de não contribuir, com sua atividade de moderação, para a realização de danos causados por terceiros por intermédio do ambiente online de propriedade dos provedores. (ROBL FILHO, 2023, s/p).

Dessa forma, para os defensores do projeto, a responsabilização das plataformas, quando não cumprirem com o dever de cuidado durante o protocolo de segurança, é uma característica da sociedade de risco contemporânea. Isso aproxima a legislação nacional dos modelos regulatórios europeus, mais compatíveis com a defesa da liberdade de expressão, bem como não implica em uma inovação substancial da ordem jurídica, posto que a legislação civil já prevê responsabilidade por risco da atividade (ROBL FILHO, 2023).

Contudo, em sentido contrário, de acordo com o Google, esse modelo de responsabilização

cria um incentivo para que plataformas, como o YouTube, que hospedam conteúdos de usuários, adotem uma moderação excessiva, levando a um maior número de remoções e comprometendo a liberdade de expressão de milhares de criadores que desejam manifestar suas opiniões de forma legítima” (LACERDA, 2023, s/p).

Em manifestação sobre o projeto de lei, no texto *Como o PL 2630 pode piorar a sua internet*, o Diretor de Relações Governamentais e Políticas Públicas do Google Brasil afirmou:

4) Traz sérias ameaças à liberdade de expressão

A proposta atual traz várias disposições que determinam um “dever de cuidado” a ser executado preventivamente pelas plataformas, principalmente, no que se refere aos conteúdos considerados ilegais pela proposta. Se o texto avançar assim, empresas de tecnologia terão que filtrar e moderar conteúdos considerando uma análise legal e assumindo uma função exercida tradicionalmente pelo Poder Judiciário.

Sem os parâmetros de proteção do Marco Civil da Internet e com as novas ameaças de multas, as empresas seriam estimuladas a remover discursos legítimos, resultando em um bloqueio excessivo e uma nova forma de censura.

Quando pensamos no YouTube ou na Busca do Google, que já têm mecanismos de denúncia disponíveis para usuários, a redação atual do PL 2630 cria um sistema que pode incentivar abusos, permitindo que pessoas e grupos mal-intencionados inundem nossos sistemas com requerimentos para remover conteúdos sem nenhuma proteção legal.

A incerteza do que pode ou não ser disponibilizado na internet levaria as empresas a restringir a quantidade de informações disponíveis, reduzindo a representatividade de vozes que existem nas plataformas. Isso violaria diretamente o princípio do acesso livre à informação, o que seria uma grande retrocesso na guerra contra conteúdos enganosos. (LACERDA, 2023, s/p).

Dessa manifestação do Google, apenas no que toca à liberdade de expressão, pode-se extrair três grandes preocupações das plataformas: (i) que caberá a elas filtrar e moderar conteúdos que demandariam análise legal, substituindo o poder judiciário; (ii) que há uma incerteza quanto aos conteúdos que podem ou não ser disponibilizados; e (iii) que a responsabilização das plataformas, mesmo sem notificação judicial, pode estimular a remoção de discursos legítimos, especialmente diante da possibilidade de denúncias mal-intencionadas por usuários a determinadas contas.

Quanto à primeira e segunda preocupações, sabe-se que as plataformas já realizam o controle dos discursos nas redes sociais por meio da moderação de conteúdo, com fundamento em conceitos abertos previstos em suas normas internas, como discursos de ódio e *fake news*. Pela ausência de transparência na moderação de conteúdo, já há uma incerteza, pelo menos para os usuários, quanto aos conteúdos que podem ou não ser disponibilizados, de modo que não há nesses dois primeiros pontos qualquer inovação às práticas já realizadas. A terceira preocupação, contudo, ainda suscita debates e divergências no cenário acadêmico, como demonstrado.

A delegação do controle do discurso para as plataformas é uma medida que se mostra necessária nesse novo modelo regulatório (*new school*) pela impossibilidade do Estado de controlar diretamente os usuários. Contudo, cabe ao Estado estabelecer um espectro regulatório, delimitando como as plataformas devem realizar esse controle (moderação) e atendendo os interesses públicos.

Nesse sentido, o PL define, no capítulo III, regras de notificação de conteúdo infringente, bem como estabelece um devido processo para a moderação de conteúdo, com mecanismos e canais para a revisão das decisões. Além disso, o capítulo IV define uma série de regras e deveres de transparência nos termos de uso e dos algoritmos de recomendação; tais disposições estão alinhadas com os princípios de Santa Clara, que devem nortear a moderação de conteúdo (item 3.2).

Observa-se, entretanto, que não há disposições no projeto de lei sobre a moderação manual de conteúdo, obrigatoriedade de revisão humana das decisões por indivíduos alinhados ao contexto da expressão, nem informações sobre as agências de checagem.

Sobre a definição dos conteúdos que podem ou não ser disponibilizados nas redes, o projeto inicialmente previa a criação de uma entidade autônoma de supervisão, órgão da administração pública indireta, de natureza de autarquia federal especial, responsável por zelar, implementar e fiscalizar o cumprimento da Lei em todo o território nacional, em parceria com o Comitê Gestor da Internet (CGIbr), que ficaria responsável pela emissão de diretrizes. Contudo, após pressões políticas, a previsão dessa entidade foi retirada do texto.

Assim, na versão analisada, caberá ao Comitê Gestor da Internet no Brasil (CGIbr)¹⁵², órgão já existente e vinculado ao Ministério das Comunicações, o papel regulatório das plataformas, que, dentre outras atribuições, deverá realizar estudos e debates para aprofundar o entendimento sobre desinformação, propor diretrizes para o seu combate no contexto da internet e das redes sociais, e também apresentar diretrizes para a elaboração de um código de conduta e validá-lo após sua construção pelos provedores.

Dessa forma, observa-se que o enfrentamento estratégico e objetivo dos conteúdos a serem disponibilizados nas redes, em especial o combate à desinformação e aos discursos de ódio, dependerá da elaboração dos códigos de conduta das plataformas (autorregulação), que deverão seguir as diretrizes elaboradas pela CGIbr e por ela serem homologados, o que se coaduna com o modelo de autorregulação regulada.

¹⁵² Portaria Interministerial N° 147, de 31 de maio de 1995. Disponível em: <https://cgi.br/portarias/numero/147/>. Acesso em: 12 mai. 2023.

Em contrapartida, para o Instituto Legal Grounds¹⁵³, a autorregulação regulada nos moldes do projeto apresenta um problema, uma vez que está ausente um mecanismo de incentivo para que as plataformas possam aderir à instituição de autorregulação. Para o instituto, esse mecanismo seria a isenção de punição às plataformas que aderirem à instituição.

Isso porque, como a proposta legislativa traz uma série de obrigações procedimentais com as melhores práticas de governança na moderação de conteúdo, uma instituição de autorregulação homologada assegura a adequação a tais práticas de modo que seus membros, ao seguirem seu código e procedimentos, estariam conforme à lei, portanto se legitimaria a isenção de punição (MARANHÃO; CAMPOS, 2020).

Destaca-se, por fim, que na versão oriunda do Senado havia um capítulo inteiro dedicado à autorregulação regulada, que foi retirado da versão final ora analisada. Contudo, inobstante a retirada dos dispositivos específicos da autorregulação regulada, o projeto ainda se coaduna com o instituto na medida em que define diretrizes concretas para a moderação de conteúdo, bem como estabelece formas de supervisão e responsabilização.

Conclui-se que o modelo regulatório mais adequado para enfrentar os desafios do exercício da liberdade de expressão nas redes sociais, que é compatível com a ordem jurídica brasileira e que encontra melhor respaldo na experiência internacional do tema, é a chamada autorregulação regulada. Ela visa compatibilizar e submeter a autorregulação privada a uma supervisão e fiscalização estatal, atendendo aos interesses públicos, posto que consegue responder ao déficit de conhecimento desencadeado pela complexidade social das dinâmicas do ambiente digital. No Brasil, caminha-se, ainda que com dificuldade, para um avanço normativo que visa implementar esse modelo regulatório.

¹⁵³ O Instituto Legal Grounds é uma organização brasileira voltada à estudos sobre políticas públicas para a comunicação social, novas mídias, tecnologias digitais da informação e proteção de dados pessoais. O instituto ainda fundou o Observatório Nacional de Regulação Digital criado em 2021 objetiva promover o debate nacional sobre propostas legislativas no Brasil e iniciativas legislativas internacionais relevantes sobre as regulações de mídias. Com isso, o Legal Grounds Institute ambiciona criar um espaço público e de referência para o debate aberto entre os diversos setores da sociedade brasileira, em ambiente digital democrático e plural.

CONCLUSÃO

A partir de um arcabouço teórico e conceitual apresentado, do qual se extraiu e definiu o alcance, a titularidade e os limites à liberdade de expressão na realidade nacional, pôde-se concluir que o surgimento das novas tecnologias, aqui consideradas as redes sociais, não teve o condão de alterar o conceito de liberdade de expressão. A mudança por elas provocada está nas condições sociais com as quais esse direito passou a ser exercido, ou seja, na alteração da infraestrutura da liberdade de expressão, uma vez que a disponibilização do espaço de manifestação pertence não mais àqueles que detêm a estrutura (conglomerados de mídias tradicionais), mas aos usuários dos serviços das plataformas na *web 2.0*.

Assim, as plataformas de redes sociais, mantidas por empresas privadas com fins lucrativos e regidas por regras de mercado, passaram a exercer uma função de extrema relevância para as democracias, que é a de mediadora do debate público, transformando-se no principal palco de argumentação e deliberação disponível para os indivíduos e emergindo como uma nova esfera pública.

Contudo, esse novo ambiente, apesar de ampliar o “livre mercado de ideias” e amplificar também a audiência, de forma paradoxal acabou por desencadear inúmeros desafios que podem colocar em risco a democracia e a própria liberdade de expressão. A ausência de um controle qualitativo do discurso e da informação se apresentou como um campo fértil para a propagação de discursos de ódio e *fake news*, os quais, associados às novas formas de inteligência artificial (*bots* e *deepfake*), colocam em risco a própria noção de verdade. A formação das câmaras de eco, a partir da filtragem e recomendação de conteúdo, tem fomentado comportamentos extremistas e intolerantes, se afastando de um ambiente saudável e adequado para a formação de uma esfera pública.

Ademais, as redes sociais, além de ampliarem o espaço comunicacional, também criaram um espaço de decisão. Isso porque as plataformas, ao operarem como intermediárias de conteúdo, acabam por decidir o que pode ou não ser dito nesse ambiente, por meio da moderação de conteúdo, com amplas competências e possibilidade de intervenção, numa espécie de poder normativo (termos e condições) e decisório, pois formam um fundamento institucional de ação e processo ao qual são submetidos os usuários, caso queiram ingressar na plataforma.

Nesse sentido, tais empresas privadas não devem ser consideradas como instituições neutras, meras intermediárias, delegatárias da ação estatal, que somente asseguram o espaço de expressão sem qualquer ingerência sobre o conteúdo ou procedência, mas devem ser compreendidas como instituições capazes de formar políticas de conteúdo, editando e controlando as expressões realizadas pelos usuários. Ao participarem da comunicação coletiva, formando a opinião pública e impactando as estruturas democráticas, devem se submeter a um regime de obrigações diferenciado por meio da regulação.

Assim, observou-se que mudança social provocada pelo surgimento das redes sociais desencadeou a necessidade de uma mudança regulatória estatal, a partir do surgimento de um novo destinatário da regulação, que são as plataformas, uma vez que é tarefa do Estado assumir uma postura de responsabilidade garantidora para assegurar o exercício socialmente aceitável da liberdade de expressão no meio digital, limitando o poder privado e transnacional das plataformas, sem, contudo, ampliar o risco de violações desse direito.

Entretanto, verificou-se que a regulação puramente estatal (heteroregulação) das plataformas, apesar da sua vantagem coercitiva, não consegue acompanhar a dinâmica do ambiente digital, apresentando um déficit cognitivo que o impede de alcançar seu objetivo. Isso porque se para cada nova tecnologia os legisladores se sentem impulsionados a realizar alterações no ordenamento jurídico, essas alterações, que podem parecer pertinentes no momento, poderão tornar-se defasadas com a mesma velocidade que a tecnologia avança.

Por outro lado, a autorregulação das plataformas, por meio da moderação de conteúdo, tem se mostrado insuficiente para enfrentar os desafios da expressão no meio digital, pois, como visto, os interesses das plataformas nem sempre estão alinhados com a defesa do interesse público ou mesmo com a ordem jurídica de determinado Estado.

Assim, nesse cenário de insuficiência dos dois modelos regulatórios tradicionais é que desponta um modelo promissor, híbrido, de direito proceduralizado, que visa conciliar os dois modelos anteriores, com prevalência da autorregulação, que é a autorregulação regulada ou correção.

Concluiu-se que esse modelo se apresenta como a melhor opção regulatória, compatível com o papel indutor do Estado regulador da atividade econômica, pois

permite a participação do objeto da regulação, que são as plataformas. Elas mantêm sua autonomia para definir regras privadas, focando na própria expertise e dinâmica, contudo são sujeitas ao estabelecimento de determinados parâmetros e *standards* de direitos fundamentais a serem seguidos, definidos pelo Estado e advindos do interesse público.

REFERÊNCIAS BIBLIOGRÁFICAS

ABBOUD, Georges; CAMPOS, Ricardo. A autorregulação regulada como modelo do Direito proceduralizado. *In*: ABBUOD, Georges; NERY JÚNIOR, Nelson; CAMPOS, Ricardo. **Fake News e Regulação**. 3. ed. São Paulo: Thomson Reuters Revista dos Tribunais, 2020.

AGHAEI, Sareh; NEMATBAKHSH, Mohammad Ali; FARSANI, Hadi Khosravi. Evolution of the World Wide Web: from web 1.0 to web 4.0. **International Journal Of Web & Semantic Technology**. [S.L.], v. 3, n. 1, p. 1-10, 31 jan. 2012. Academy and Industry Research Collaboration Center (AIRCC).
<http://dx.doi.org/10.5121/ijwest.2012.3101>. Disponível em:
<https://airccse.org/journal/ijwest/papers/3112ijwest01.pdf>. Acesso em: 11 maio 2023.

ALLEMAND, Luiz Cláudio; MAGRO, Américo Ribeiro. **A natureza e proteção das obras geradas pelo Chat Gpt e outros sistemas de Inteligência Artificial**. 2023. Disponível em: <https://www.iabnacional.org.br/opiniao/a-natureza-e-protECAo-das-obras-geradas-pelo-chat-gpt-e-outros-sistemas-de-inteligencia-artificial>. Acesso em: 11 maio 2023.

ARENDDT, Hannah. **Verdade e Política**. Tradutor: Manuel Alberto, 1967. [S.I.]: E-Disciplinas, USP. Disponível em:
https://edisciplinas.usp.br/pluginfile.php/5144219/mod_resource/content/0/ARENDDT%2C%20Hannah%20%281967%29%20Verdade%20e%20pol%C3%ADtica.pdf. Acesso em: 12 mai. 2023.

BALKIN, Jack M. FREE SPEECH IS A TRIANGLE. **Columbia Law Review**, New York, v. 118, n. 7, p. 2011-2056, nov. 2018. Disponível em:
<https://www.jstor.org/stable/26524953>. Acesso em: 12 maio 2023.

BALKIN, Jack M. OLD-SCHOOL/NEW-SCHOOL SPEECH REGULATION. **The Harvard Law Review Association**, [S.L.], v. 127, n. 8, p. 2296-2342, jun. 2014. Disponível em: https://harvardlawreview.org/wp-content/uploads/2014/06/vol127_balkin.pdf. Acesso em: 11 maio 2023.

BARLOW, John Perry. **Declaração de Independência do Ciberespaço**. 1996. Disponível em: <http://www.dhnet.org.br/ciber/textos/barlow.htm>. Acesso em: 11 maio 2023.

BATTISTI, Roberta. **Regulação das Big Techs**. São Paulo: Almedina, 2023.

BERLIN, Isaiah. **Quatro Ensaios Sobre a Liberdade**. 2. Ed. Brasília: Editora Universidade de Brasília, 1969.

BOBBIO, Norberto. **Elogio da serenidade e outros escritos morais**. 2. ed. São Paulo: UNESP, 2011.

BRASIL. **AÇÃO DIRETA DE INCONSTITUCIONALIDADE POR OMISSÃO 26 DISTRITO FEDERAL**. Brasília/DF. Supremo Tribunal Federal, 2019a. Disponível em:

<https://redir.stf.jus.br/paginadorpub/paginador.jsp?docTP=TP&docID=754019240>.

Acesso em: 10 mai. 2023.

BRASIL. **ARGÜIÇÃO DE DESCUMPRIMENTO DE PRECEITO FUNDAMENTAL 187 DISTRITO FEDERAL**, de 15 de julho de 2011. Brasília/DF. Supremo Tribunal Federal, 2011. Disponível em:

<https://redir.stf.jus.br/paginadorpub/paginador.jsp?docTP=TP&docID=5956195>.

Acesso em: 10 mai. 2023.

BRASIL. **PARECER PROFERIDO EM PLENÁRIO AO PROJETO DE LEI Nº 2.630, DE 2020, E APENSADOS**. Brasília/DF. Câmara dos Deputados, 2023. Disponível em:

https://www.camara.leg.br/proposicoesWeb/prop_mostrarintegra?codteor=2265334.

Acesso em: 06 maio. 2023.

BRASIL. **Projeto de Lei 2630/2020**. Institui a Lei Brasileira de Liberdade, Responsabilidade e Transparência na Internet. Brasília/DF. Senado Federal, 2020. Disponível em:

https://www.camara.leg.br/proposicoesWeb/prop_mostrarintegra?codteor=1909983.

Acesso em: 10 mai. 2023.

BRASIL. **RECURSO ORDINÁRIO ELEITORAL Nº 0603975-98 PARANÁ**.

Curitiba/PR. Tribunal Superior Eleitoral, 2018. Disponível em:

<https://www.conjur.com.br/dl/voto-salomao-tse-gustavo-francischini.pdf>. Acesso em:

11 mai. 2023.

BRASIL. **Resolução Nº 305 de 17/12/2019**. Brasília/DF. Conselho Nacional de

Justiça, 2019b. Disponível em: <https://atos.cnj.jus.br/atos/detalhar/3124>. Acesso em:

10 mai. 2023.

BRAUNER, Eugenio. Julinho da Adelaide, um pseudônimo que driblou a Censura - em processo. **Nau Literária**, Porto Alegre, v. 1, n. 1, p. 1-10, 20 jun. 2005.

Universidade Federal do Rio Grande do Sul. Disponível em:

<https://seer.ufrgs.br/index.php/NauLiteraria/article/view/4839>. Acesso em: 11 maio

2023.

BRUGGER, Winfried. The Treatment of Hate Speech in German Constitutional Law (Part I). **German Law Journal**, [S.L.], v. 4, n. 1, p. 1-22, 1 jan. 2003. Cambridge University Press (CUP). Disponível em:

<https://www.cambridge.org/core/journals/german-law-journal/article/treatment-of-hate-speech-in-german-constitutional-law-part-i/FA250939FC230E7DE047B4213B1447EC>.

Acesso em: 10 maio 2023.

BULOS, Uadi L. **Curso de Direito Constitucional**. 15ª ed. São Paulo: SaraivaJur, 2022.

BUTANTAN, Portal do. **Saiba mais sobre Brian Deer, o homem que**

desmascarou a fraude que ligava vacinas ao autismo. 2021. Colaboração de

Maria Teresa Ghiuro Valentini Abdullatif e Paulo Henrique Nico Monteiro. Disponível

em: <https://butantan.gov.br/butantan-educa/saiba-mais-sobre-brian-deer-o-homem->

[que-desmascarou-a-fraude-que-ligava-vacinas-ao-autismo](#). Acesso em: 10 maio 2023.

CAVALCANTE FILHO, João Trindade. **O discurso do ódio na jurisprudência alemã, americana e brasileira: uma análise à luz da filosofia política**. 2014. 127 f. Dissertação (Mestrado) - Curso de Mestre em Constituição e Sociedade, Instituto Brasiliense de Direito Público, Escola de Direito de Brasília, Brasília, 2014. Disponível em: <https://repositorio.idp.edu.br/handle/123456789/2184>. Acesso em: 10 maio 2023.

CIDH - Comissão Interamericana de Direitos Humanos. **CONVENÇÃO AMERICANA SOBRE DIREITOS HUMANOS**. 1969. Disponível em: https://www.cidh.oas.org/basicos/portugues/c.convencao_americana.htm. Acesso em: 10 maio 2023.

COLLEONI, Elanor; ROZZA, Alessandro; ARVIDSSON, Adam. Echo Chamber or Public Sphere? Predicting Political Orientation and Measuring Political Homophily in Twitter Using Big Data. **Journal Of Communication**, [S.L.], v. 64, n. 2, p. 317-332, 19 mar. 2014. Oxford University Press (OUP). Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1111/jcom.12084>. Acesso em: 11 maio 2023.

DAHLGREN, Peter. The Internet, Public Spheres, and Political Communication: dispersion and deliberation. **Political Communication**, [S.L.], v. 22, n. 2, p. 147-162, abr. 2005. Disponível em: <https://www.tandfonline.com/doi/full/10.1080/10584600590933160>. Acesso em: 11 maio 2023.

DANTAS, Bruno; SANTOS, Caio Victor Ribeiro dos. Fake News e liberdade de expressão: contribuição para um debate necessário. *In*: ABBOUD, Georges; NERY JÚNIOR, Nelson; CAMPOS, Ricardo. **Fake News e Regulação**. 3. ed. São Paulo: Thomson Reuters Revista dos Tribunais, 2020.

DE PAULA, Lorena Tavares de; MICHALSKI, Rafael. OS BOTS DE DISSEMINAÇÃO DE INFORMAÇÃO NA CONJUNTURA DAS CAMPANHAS PRESIDENCIAIS DE 2018 NO BRASIL. **Múltiplos Olhares em Ciência da Informação**, [S. l.], v. 9, n. 1, 2019. Disponível em: <https://periodicos.ufmg.br/index.php/moci/article/view/17048>. Acesso em: 11 maio 2023.

DWORKIN, Ronald. The Right to Ridicule. **The New York Review**. 2006. Disponível em: <https://www.nybooks.com/articles/2006/03/23/the-right-to-ridicule/>. Acesso em: 10 maio 2023.

EUA – ESTADOS UNIDOS DA AMÉRICA. **Brandenburg v. Ohio**. 1969. Disponível em: <https://www.oyez.org/cases/1968/492>. Acesso em: 10 mai. 2023.

FARIAS, Edilsom Pereira de. **Liberdade de expressão e comunicação: teoria e proteção constitucional**. 2001. 290 f. Tese (Doutorado) - Curso de Doutorado em Direito, Programa de Pós-Graduação em Direito, Universidade Federal de Santa Catarina, Florianópolis, 2001. Disponível em:

<https://repositorio.ufsc.br/bitstream/handle/123456789/79426/182787.pdf?sequence=1&isAllowed=y>. Acesso em: 10 maio 2023.

FARINHO, Domingos Soares. Delimitação do espectro regulatório de redes sociais. In: ABOUD, Georges; NERY JÚNIOR, Nelson; CAMPOS, Ricardo. **Fake News e Regulação**. 3. ed. São Paulo: Thomson Reuters Revista dos Tribunais, 2020.

GARRETT, Filipe. O que é bot? Conheça os robôs que estão 'dominando' a Internet. **Techtudo**. 2018. Disponível em: <https://www.techtudo.com.br/noticias/2018/07/o-que-e-bot-conheca-os-robos-que-estao-dominando-a-internet.ghtml>. Acesso em: 11 maio 2023.

GILLESPIE, Tarleton. **Custodians of the Internet**: platforms, content moderation, and the hidden decisions that shape social media. Londres: Yale University Press, 2018.

HAN, Byung-Chul. **Infocracia**: digitalização e a crise da democracia. Petrópolis: Vozes, 2022.

HOFFMANN-RIEM, Wolfgang. **Teoria geral do direito digital**: transformação digital, desafios para o direito. 2. ed. Rio de Janeiro: Forense, 2022.

KLONICK, Kate. The New Governors: the people, rules, and processes governing online speech. **Harvard Law Review**, [s. l], v. 131, n. 11598, p. 1599-1670, mar. 2017. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2937985. Acesso em: 11 mai. 2023.

LACERDA, Marcelo. Como o PL 2630 pode piorar a sua internet. **Google Blog**. 2023. Disponível em: <https://blog.google/intl/pt-br/novidades/iniciativas/como-o-pl-2630-pode-piorar-a-sua-internet/>. Acesso em: 12 maio 2023.

LAURENTIIS, Lucas Catib de; THOMAZINI, Fernanda Alonso. Liberdade de Expressão: teorias, fundamentos e análise de casos. **Revista Direito e Práxis**, [S.L.], v. 11, n. 4, p. 2260-2301, dez. 2020. Disponível em: <https://www.e-publicacoes.uerj.br/index.php/revistaceaju/article/view/44121>. Acesso em: 10 maio 2023.

LEMBKE, Anna. **Nação dopamina**: Por que o excesso de prazer está nos deixando infelizes e o que podemos fazer para mudar. São Paulo: Vestígio, 2021.

LESSIG, Lawrence. The Limits in Open Code: regulatory standards and the future of the net. **Berkeley Technology Law Journal**, [s. l], v. 14, n. 2, p. 759-769, jun. 1999. Disponível em: <https://www.jstor.org/stable/24115665>. Acesso em: 11 maio 2023.

MACHADO, Jónatas E. M.; BRITO, Iolanda A. S. Rodrigues de. Liberdade de expressão, informações falsas e figuras públicas: o perigo de manipulação da esfera de discurso público. **Revista Populus**, Salvador, n. 8, p. 207-242, jun. 2020. Disponível em: <https://bibliotecadigital.tse.jus.br/xmlui/handle/bdtse/8174>. Acesso em: 13 mai. 2023.

MARANHÃO, Juliano; CAMPOS, Ricardo. Fake news e autorregulação regulada das redes sociais no Brasil: fundamentos constitucionais. In: ABOUD, Georges; NERY

JÚNIOR, Nelson; CAMPOS, Ricardo. **Fake News e Regulação**. 3. ed. São Paulo: Thomson Reuters Revista dos Tribunais, 2020.

MARTINS HARTMANN, Ivar Alberto; SARLET, Ingo Wolfgang. DIREITOS FUNDAMENTAIS E DIREITO PRIVADO: A PROTEÇÃO DA LIBERDADE DE EXPRESSÃO NAS MÍDIAS SOCIAIS. **Direito Público**, [S. l.], v. 16, n. 90, 2019.

Disponível em:

<https://www.portaldeperiodicos.idp.edu.br/direitopublico/article/view/3755>. Acesso em: 10 maio. 2023.

MARTINS, Flávio. **Curso de Direito Constitucional**. 6. ed. São Paulo: SaraivaJur, 2022.

MCQUAIL, Denis. **Teorias da comunicação de massas**. Trad. Carlos de Jesus. Lisboa: Fundação Calouste Gulbenkian, 2003.

MENDES, Gilmar Ferreira; BRANCO, Paulo Gustavo Gonet. **Curso de Direito Constitucional**. 8ª ed. São Paulo: Saraiva, 2013.

MILL, John Stuart. **Sobre a liberdade**. Tradução e prefácio de Alberto da Rocha Barros. Petrópolis: Vozes, 2019.

MIRANDA, Beatriz Casto; SANTOS, Carolina Xavier. A política brasileira no alcance dos dedos: os limites e as possibilidades do agir democrático no ciberespaço.

Internet & Sociedade, [S.L], v. 2, n. 1, p. 135-163, jun. 2021. Disponível em:

<https://revista.internetlab.org.br/a-politica-brasileira-no-alcance-dos-dedos-os-limites-e-as-possibilidades-do-agir-democratico-no-ciberespaço/>. Acesso em: 11 maio 2023.

OEА – Organização dos Estados Americanos. **DECLARAÇÃO CONJUNTA DE 2021 SOBRE POLÍTICOS E AUTORIDADES PÚBLICAS E LIBERDADE DE EXPRESSÃO**. 2021. Disponível em:

<https://www.oas.org/pt/cidh/expressao/showarticle.asp?artID=1214&IID=4>. Acesso em: 10 maio 2023.

ONU – ORGANIZAÇÃO DAS NAÇÕES UNIDAS. **Declaração Universal dos Direitos Humanos**. Assembleia Geral da ONU, A/Res/3/217A (10 de dezembro de 1948). Disponível em:

https://www.ohchr.org/sites/default/files/UDHR/Documents/UDHR_Translations/por.pdf. Acesso em: 10 mai. 2023.

O'REILLY, Tim. Web 2.0 Compact Definition: trying again. **O'Reilly Radar**. 2006.

Disponível em: <http://radar.oreilly.com/archives/2006/12/web-20-compact-definition-tryi.html>. Acesso em: 11 maio 2023.

PALLERO, Javier; BARATA, Joan; BETANCOURT, Valeria; PIAZZA, Andrés; MASTRINI, Guillermo; BECERRA, Martín; FREULER, Juan Ortiz. **Contribuições para uma regulação democrática das grandes plataformas que garanta a liberdade de expressão na internet**. 2019. Elaborada por Intervezes, Observacom, Desarrollo Digital, Instituto Brasileiro de Defesa do Consumidor. Disponível em:

<https://www.observacom.org/wp-content/uploads/2019/08/Contribuic%CC%A7o%CC%83es-para-uma-regulac%CC%A7a%CC%83o-democra%CC%81tica-das-grandes-plataformas-que>

[garanta-a-liberdade-de-expressa%CC%83o-na-internet.pdf](#). Acesso em: 11 maio 2023.

PIAIA, Thami Covatti; RITTER, Letícia Mousquer; SANGOI, Rafael Martins. Internet, Liberdade de Informação e o caso das Echo Chambers Ideológicas. **Revista da Faculdade de Direito do Sul de Minas**, [S. l.], v. 34, n. 2, 2018. Disponível em: <https://revista.fdsm.edu.br/index.php/revistafdsm/article/view/238>. Acesso em: 11 maio. 2023.

PIERRO, Bruno de. Epidemia de fake news. **Revista Pesquisa**. 2020. Disponível em: <https://revistapesquisa.fapesp.br/epidemia-de-fake-news/>. Acesso em: 10 maio 2023.

POLÍGRAFO. **As quatro maiores fake news colocadas a circular pelos movimentos anti-vacinas**. 2019. Disponível em: <https://poligrafo.sapo.pt/saude/artigos/as-quatro-maiores-fake-news-colocadas-a-circular-pelos-movimentos-anti-vacinas>. Acesso em: 10 maio 2023.

POSNER, Richard A. The Speech Market and the legacy of Schenck. *In*: BOLLINGER, Lee C.; STONE, Geoffrey R. **Eternally Vigilant: Free Speech in the modern era**. Chicago: University of Chicago Press, 2018.

QUATTROCIOCCHI, Walter; SCALA, Antonio; SUNSTEIN, Cass R. Echo Chambers on Facebook. **Ssrn Electronic Journal**, [S.L.], v. 1, n. 1, p. 1-15, jun. 2016. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2795110. Acesso em: 13 maio 2023.

RAMOS, Carlos Eduardo Vieira. **Direito das plataformas digitais: regulação privada da liberdade de expressão na internet**. Curitiba: Juruá, 2021.

RECUERO, Raquel; SOARES, Felipe; ZAGO, Gabriela. Polarização, Hiperpartidarismo e Câmaras de Eco: como circula a desinformação sobre covid-19 no twitter. **Revista Contracampo**, [S.L.], v. 40, n. 1, p. 1-19, 30 abr. 2021. Disponível em: <https://preprints.scielo.org/index.php/scielo/preprint/view/1154>. Acesso em: 11 maio 2023.

RIEMSDIJK, G; SOUSA, A; CRUZ, J; GONÇALVES, J; OLIVEIRA, R; LAPA ESTEVES, M; MAGALHÃES, J. O FATOR FAKE NEWS NA ATUALIDADE, NA MIRA DA PSICOLOGIA. **International Journal of Developmental and Educational Psychology**, v. 1, n. 1, 2020. Disponível em: <https://www.redalyc.org/articulo.oa?id=349863388026>. Acesso em: 10 mai. 2023.

ROBL FILHO, Ilton Norberto. **Nota Técnica sobre o Projeto de Lei nº. 2630/2020 produzida pelo Grupo de Pesquisa “Democracia Constitucional, Novos Autoritarismos e Constitucionalismo Digital”**. Brasília/DF, 2023. Disponível em: <https://www.conjur.com.br/dl/nota-tecnica-idp-fake-news.pdf>. Acesso em: 14 mai. 2023.

ROBLES-LESSA, Moyana Mariano; CABRAL, Hideliza Lacerda Tinoco Boechat; SILVESTRE, Gilberto Fachetti. Deepfake: a inteligência artificial e o algoritmo causando riscos à sociedade no ciberespaço. **Derecho y Cambio Social**, [s. l.], v. 1, n. 61, p. 475-487, jul. 2020. Disponível em:

<https://dialnet.unirioja.es/servlet/articulo?codigo=7525024>. Acesso em: 11 maio 2023.

RUSSELL, Bertrand. **História da Filosofia Ocidental**: Livro 3 – A Filosofia Moderna. Rio de Janeiro: Nova Fronteira, 2017.

SABINO, Marco Antonio da Costa. Redes Sociais são veículos de comunicação social? **Análise**. 2022. Disponível em: <https://analise.com/dna/artigos/7744>. Acesso em: 11 maio 2023.

SARLET, Ingo Wolfgang; MOLINARO, Carlos Alberto. Direito à informação e Direito de Acesso à informação como direitos fundamentais na Constituição Brasileira.

Revista da Agu, [S.L.], v. 13, n. 42, p. 9-38, 31 dez. 2014. Disponível em: <https://revistaagu.agu.gov.br/index.php/AGU/article/view/542>. Acesso em: 10 maio 2023.

SARLET, Ingo. Direitos Fundamentais em Espécie. *In*: SARLET, Ingo Wolfgang; MARINONI, Luiz Guilherme Marinoni; MITIDIERO, Daniel. **Curso de Direito Constitucional**. 11ª ed. São Paulo: Saraiva, 2022.

SARMENTO, Daniel. A liberdade de expressão e o problema do "Hate Speech". *In*: SARMENTO, Daniel. **Livres e iguais**: estudos de Direito Constitucional. Rio de Janeiro: Lumen Juris, 2006. Disponível em:

<http://professor.pucgoias.edu.br/sitedocente/admin/arquivosupload/4888/material/a-liberdade-de-expressao-e-o-problema-do-hate-speech-daniel-sarmento.pdf>. Acesso em: 10 mai. 2023.

SARMENTO, Daniel. Comentários ao artigo 5º. *In*: CONOTILHO, J. J. Gomes; MENDES, Gilmar F; SARLET, Ingo W.; STRECK, Lenio L. (Org.). **Comentários à Constituição do Brasil**. São Paulo: Saraiva/Almedina, 2013.

SCHWAB, Klaus. **A Quarta Revolução Industrial**. 1ª ed. São Paulo: Edipro, 2016.

SCHWABE, Jürgen; MARTINS, Leonardo (Org). **Cinquenta anos de jurisprudência do Tribunal Constitucional Federal alemão**. Montevideo: Konrad-Adenauer Stiftung, 2006.

SILVA, José Afonso da. **Curso de Direito Constitucional Positivo**. 44. ed., rev. e atual. São Paulo: Malheiros, 2022.

SOUZA, Carlos Affonso; LEMOS, Ronaldo. **Marco civil da internet**: construção e aplicação. Juiz de Fora: Editar Editora Associada Ltda, 2016. Disponível em: https://itsrio.org/wp-content/uploads/2017/02/marco_civil_construcao_aplicacao.pdf. Acesso em: 11 mai. 2023.

SPENCER, Michael K. Deep Fake, a mais recente ameaça distópica. **Outras Palavras**. 2019. Disponível em: <https://outraspalavras.net/tecnologiaemdisputa/deep-fake-a-ultima-distopia/>. Acesso em: 11 maio 2023.

SUSTEIN, Cass R. THE DAILY ME. **#Republic**, [S.L.], p. 1-30, 31 dez. 2020. Princeton University Press. Disponível em: <http://assets.press.princeton.edu/chapters/s8468.pdf>. Acesso em: 11 mai. 2023.

SUSTEIN, Cass R. The Future of free speech. *In*: BOLLINGER, Lee C.; STONE, Geoffrey R. **Eternally Vigilant: Free Speech in the modern era**. Chicago: University of Chicago Press, 2018.

TARUFFO, Michele. Leyendo a Ferrajoli: consideraciones sobre la jurisdicción. **Revista Cap Jurídica Central**, [S.L.], v. 3, n. 4, p. 275-296, ago. 2019. Disponível em: <https://revistadigital.uce.edu.ec/index.php/CAP/article/view/1957>. Acesso em: 10 maio 2023.

TESTA JÚNIOR, Washington Luiz. Direito fundamental à informação factual verdadeira: Proibição da mentira no estado constitucionaldemocrático de direito. *In*: **Anais do XVIII Congresso Nacional do CONPEDI**, Florianópolis: Fundação Boiteux, 2009. p.(8535)-(8550) Disponível em: http://www.publicadireito.com.br/conpedi/manaus/arquivos/anais/sao_paulo/2272.pdf. Acesso em: 15 maio 2023.

THE SANTA CLARA PRINCIPLES On Transparency and Accountability in Content Moderation. 2018. Disponível em: <https://santaclaraprinciples.org/>. Acesso em: 17 mai. 2023.

TORNADA, João. Liberdade de expressão ou “liberdade de ofender”? – o conflito entre a liberdade de expressão e de informação e o direito à honra e ao bom nome. **Book Revista O Direito**, [S.L.], v. 1, 150^a, p. 119-155, 2018. Disponível em: https://www.academia.edu/41276539/Liberdade_de_express%C3%A3o_ou_liberdade_de_ofender_o_conflito_entre_a_liberdade_de_express%C3%A3o_e_de_informa%C3%A7%C3%A3o_e_o_direito_%C3%A0_honra_e_ao_bom_nome. Acesso em: 13 mai. 2023.

UNESCO – Organização das Nações Unidas para a Educação, a Ciência e a Cultura. **Um Mundo e Muitas Vozes: comunicação e informação na nossa época**. Comissão Internacional para o Estudo dos Problemas da Comunicação. Trad. Eliane Zagury. Rio de Janeiro: Fundação Getúlio Vargas, 1983.

VANNUCHI, Camilo. O direito à comunicação e os desafios da regulação dos meios no Brasil. **Galáxia**, São Paulo, n. 38, p. 167-180, ago. 2018. Disponível em: <https://revistas.pucsp.br/index.php/galaxia/article/view/32145>. Acesso em: 10 maio 2023.

VENTURI, Thaís G. Pascoaloto. Responsabilidade civil das plataformas digitais: em busca de adequados marcos regulatórios. **Migalhas**. 2022. Disponível em: <https://www.migalhas.com.br/coluna/migalhas-de-responsabilidade-civil/362746/responsabilidade-civil-das-plataformas-digitais>. Acesso em: 12 maio 2023.

VESTING, Thomas. A mudança da esfera pública pela inteligência artificial. *In*: ABBOUD, Georges; NERY JÚNIOR, Nelson; CAMPOS, Ricardo. **Fake News e Regulação**. 3. ed. São Paulo: Thomson Reuters Revista dos Tribunais, 2020.

WARBURTON, Nigel. **Liberdade de expressão: uma breve introdução**. Belo Horizonte: editora dialética, 2020.

WARDLE, Claire. **Entender a desordem informacional**. [S.l.]: Firstdraft, 2020. Disponível em: https://firstdraftnews.org/wp-content/uploads/2020/07/Information_Disorder_Digital_AW_PTBR.pdf?x76851. Acesso em: 10 mai. 2023.

WARDLE, Claire; DERAKHSHAN, Hossein. **INFORMATION DISORDER: Toward an interdisciplinary framework for research and policy making**. [S.l.]: Council of Europe, 2017. Disponível em: <https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-researc/168076277c>. Acesso em: 10 mai. 2023.

WIMMER, Miriam; CARVALHO, Lucas Borges de. O papel e os limites do anonimato: em busca de uma interpretação constitucionalmente adequada. **Pensar - Revista de Ciências Jurídicas**, [S.L.], v. 27, n. 2, p. 16, 2 jun. 2022. Disponível em: <https://ojs.unifor.br/rpen/article/view/13041>. Acesso em: 11 maio 2023.

ZUBOFF, Shosana. **A era do capitalismo de vigilância: a luta por um futuro humano na nova fronteira do poder**. São Paulo: Intrínseca, 2021.