

O PAPEL DAS PLATAFORMAS DE REDES SOCIAIS DIANTE DO DEVER DE COMBATER O DISCURSO DE ÓDIO NO BRASIL

THE ROLE OF SOCIAL MEDIA PLATFORMS FACING THE DUTY TO COMBAT HATE SPEECH IN BRAZIL

Recebimento: 1 dez. 2021

Aceitação: 29 mar. 2022

Daury Cesar Fabriz

Doutor em Direito

Afiliação institucional: Universidade Federal do Espírito Santo – UFES – (Vitória, ES, Brasil)

Lattes iD: <http://lattes.cnpq.br/7280691457104972>

Email: daury@terra.com.br

Gabriel Heringer de Mendonça

Mestrando em Direito

Afiliação institucional: Faculdade de Direito de Vitória – FDV – (Vitória, ES, Brasil)

Lattes iD: <http://lattes.cnpq.br/6194624467048858>

Email: heringer81@gmail.com

Como citar este artigo / How to cite this article (informe a data atual de acesso / inform the current date of access):

FABRIZ, Daury Cesar; MENDONÇA, Gabriel Heringer de. O papel das plataformas de redes sociais diante do dever de combater o discurso de ódio no Brasil. **Revista da Faculdade de Direito UFPR**, Curitiba, v. 67, n. 1, p. 127-149, jan./abr. 2022. ISSN 2236-7284. Disponível em: <https://revistas.ufpr.br/direito/article/view/83904>. Acesso em: 30 abr. 2022. DOI: <http://dx.doi.org/10.5380/rfdufpr.v67i1.83904>.

RESUMO

O presente trabalho visou analisar qual o papel das plataformas de redes sociais no combate ao discurso de ódio veiculado na internet. Para tanto, utilizou-se o método dedutivo, por meio de uma revisão da bibliografia, da legislação e da jurisprudência existentes sobre o tema. Pretendeu-se refletir sobre a conceituação em torno da liberdade de expressão, a definição de discurso de ódio, a identificação da existência de um dever fundamental de combate ao discurso de ódio e a definição de redes sociais. A partir da análise realizada, argumenta-se que a liberdade de expressão não possui caráter absoluto, podendo colidir com outros direitos fundamentais, como no caso do discurso de ódio. Além disso, demonstrou-se que as redes sociais representam um importante espaço para debate e comunicação, porém servindo muitas vezes como ambiente para difusão do discurso de ódio, o que demanda a atuação do direito, no sentido de proteger as vítimas desses ataques. Ao final, concluiu-se que há um importante papel para as plataformas de redes sociais no combate ao discurso de ódio, diante da identificação de um verdadeiro dever fundamental nesse sentido. Todavia, a atuação dessas plataformas carece de aprimoramento, em prol de maior clareza, transparência e contraditório, mediante o fortalecimento da relação entre Estado e empresas de infraestrutura de internet, para assegurar que o controle não seja feito levando em consideração somente os interesses do mercado.

PALAVRAS-CHAVE

Discurso de ódio. Liberdade de expressão. Direitos fundamentais. Controle. Redes sociais.

ABSTRACT

The present work aimed to analyze the role of social media platforms in the fight against hate speech transmitted on the internet. For that, the deductive method was used, through a review of the bibliography, legislation, and jurisprudence on the subject. It was intended to reflect on the conceptualization surrounding freedom of expression, the definition of hate speech, the identification of the existence of a human duty to combat hate speech, and the definition of social networks. Based on the analysis carried out, it is argued that freedom of expression does not have an absolute character and can conflict with other human rights, as in the case of hate speech. In addition, it has been shown that social networks represent an important space for debate and communication, although often serving as an environment for the dissemination of hate speech, which demands the action of the law to protect the victims of these attacks. In the end, it was concluded that there is an important role for social media platforms in the fight against hate speech, given the identification of a true human duty in this sense. However, the performance needs improvement in its execution, in favor of greater clarity, transparency, and the contradictory right, by strengthening the relationship between the State and internet infrastructure companies, to ensure that the control is not carried out considering only the market interests.

KEYWORDS

Hate speech. Freedom of expression. Human rights. Control. Social networks.

INTRODUÇÃO

A “era digital” foi responsável pela reformulação da interação e da comunicação entre as pessoas, na medida em que as manifestações se tornaram cada vez mais corriqueiras, informais e com grande alcance. A ausência de um contato físico encoraja as pessoas, muitas vezes, a expor suas opiniões sem medo, com a falsa impressão de ausência de responsabilidade.

Nesse cenário, surgem as redes sociais como plataformas com perfis diversificados e múltiplos, servindo, com frequência, de palco para grandes conflitos. Muitos desses conflitos caracterizam verdadeiro discurso de ódio, diante do intuito de ataques baseados na intolerância e que fomentam a privação de direitos e a exclusão social de indivíduos ou grupos específicos, geralmente associados ao preconceito de raça, de gênero, de orientação sexual e de origem/nacionalidade.

No caso do discurso de ódio disseminado nas redes sociais, verifica-se uma colisão entre direitos fundamentais, quais sejam, os direitos à liberdade de expressão, à privacidade, à honra e à igualdade. Logo, o debate no meio jurídico sobre a identificação do discurso de ódio e sobre mecanismos de coibição à sua perpetuação nas redes sociais surge como um tema relevante e atual.

O método utilizado nesta pesquisa é o dedutivo e o trabalho se desenvolve mediante a revisão bibliográfica sobre o tema, incluindo as leituras de obras de autores estrangeiros e brasileiros.

Sob o enfoque dos deveres fundamentais, a proposta de problema a ser pesquisado neste artigo se refere à identificação do papel das plataformas de redes sociais diante do dever fundamental de proteção contra toda forma de discriminação, de opressão e de violência, inclusive quando materializada por meio de discurso de ódio.

A partir desse problema de pesquisa identificado, são percorridos os seguintes objetivos específicos ao longo do trabalho: a identificação dos contornos jurídicos do direito fundamental à liberdade de expressão; a caracterização e definição do discurso de ódio; a verificação da existência de um dever fundamental de combate a esse tipo de discurso; a delimitação das características das redes sociais e a análise dos mecanismos utilizados pelas plataformas de redes sociais na moderação de discurso.

A ampliação do debate sobre o tema é necessária, principalmente, diante da incipiente regulamentação existente no Brasil sobre esse importante aspecto da vida social digital dos brasileiros. Assim, mostra-se relevante identificar, no cenário jurídico do País, a atuação que compete às plataformas de redes sociais em prol do dever de contribuir para a existência de um ambiente virtual saudável de discussão, que assegure a livre manifestação do pensamento e o respeito à privacidade, à igualdade e à honra dos seus usuários.

1 LIBERDADE DE EXPRESSÃO E DISCURSO DE ÓDIO

A liberdade de expressão é um importante direito, reconhecido por diversos ordenamentos jurídicos e dispositivos internacionais e que integra o grupo dos direitos fundamentais de primeira dimensão. Em um ambiente democrático, esse direito está inserido no núcleo essencial de liberdades individuais, que decorre da própria dignidade humana.

Em sua origem, a liberdade de expressão surge como forma de defesa do cidadão diante do Estado, de modo a impedir que este cometa perseguições e censure a livre manifestação do pensamento. Nesse sentido, representa um direito negativo, um dever de abstenção do Estado.

No estudo desse importante direito, merece ser pontuada a compilação, promovida por Farias (2001, p. 57-58), das teorias que fomentam a liberdade de expressão, as quais podem ser agrupadas sob duas perspectivas: objetiva e subjetiva. O autor ressalta que a primeira perspectiva decorre de um papel supraindividual desse direito, ligado ao regime democrático e aos instrumentos necessários de participação popular. Já a segunda perspectiva decorre de uma base individualista, associada à autorrealização e à autonomia do indivíduo para pensar e se manifestar por si mesmo.

No caso do Brasil, observa-se que ambas as matrizes da liberdade de expressão se encontram positivadas na Constituição de 1988, na medida em que a dignidade da pessoa humana resta assegurada no inciso III do artigo 1º, como um fundamento do Estado democrático de direito, e, por sua vez, a sociedade livre é reconhecida como um objetivo fundamental da República Federativa do Brasil (artigo 3º, inciso I) (BRASIL, 1988). Essas disposições acabam reverberando nos demais dispositivos constitucionais que tratam da liberdade de expressão, seja em um viés individual, como no artigo 5º, incisos IV (livre manifestação do pensamento) e IX (livre expressão da atividade intelectual, artística, científica e de comunicação, independentemente de censura ou licença), ou coletivo, no caso da comunicação social disposta no artigo 220 (ausência de restrição quanto à manifestação do pensamento, à criação, à expressão e à informação, sob qualquer forma, processo ou veículo) (BRASIL, 1988).

Assentadas as bases do direito de livre expressão, resta ainda verificar se ele se reveste de um caráter absoluto, conforme análise a seguir.

1.1 CARÁTER RELATIVO DA LIBERDADE DE EXPRESSÃO

Conforme ressaltado, a liberdade de expressão encontra fundamento tanto na dignidade da pessoa humana como no pilar do regime democrático. Por isso, a importância desse direito fundamental acaba transmitindo uma aparente noção de caráter absoluto.

Assim como todo direito, o seu exercício deve observar a existência dos direitos reconhecidos aos demais indivíduos e à sociedade em geral. Nesse sentido, Berto (2012, p. 36) afirma que não se deve perder de vista a noção de que qualquer pessoa é livre para expor suas opiniões, todavia o convívio social prevê as mesmas liberdades a todos os indivíduos.

Outrossim, da mesma forma que a liberdade de expressão se encontra positivada na Constituição de 1988, outros direitos fundamentais são igualmente reconhecidos, como a inviolabilidade da intimidade, da vida privada, da honra e da imagem das pessoas (artigo 5º, inciso X) (BRASIL, 1988). Ressalta-se que, assim como a liberdade de expressão, os demais direitos citados também decorrem da dignidade da pessoa humana.

Diante de um conflito entre a liberdade de expressão e outros direitos fundamentais, não é possível sustentar a prevalência daquela, de forma prévia e sem a análise dos fatos. Isso porque não existe uma hierarquia de direitos fundamentais exposta na Constituição. Logo, diante de um caso de colisão entre a liberdade de expressão e a inviolabilidade da honra/imagem/intimidade, é necessária a análise do caso concreto.

Em vista disso, na doutrina, mostra-se pertinente a teoria defendida por Alexy (1992, p. 269, 275), que identifica uma colisão de direitos fundamentais em sentido estrito toda vez que o exercício de um direito por determinado titular traz consequências negativas sobre direitos fundamentais de outro titular. Segundo o autor, a solução do conflito estaria na ponderação dos princípios.

Tal ponderação é amplamente utilizada pelos tribunais brasileiros, conforme demonstrado no julgamento do famoso caso *Ellwanger*, que versou sobre publicações com conteúdo antissemita. Nesse julgado, o Supremo Tribunal Federal fixou o entendimento de que a liberdade de expressão não implica uma salvaguarda de condutas ilícitas, como sucede com os delitos contra a honra (BRASIL, 2004).

A limitação do direito de liberdade de expressão também encontra respaldo em documentos internacionais, tais como o marco jurídico interamericano sobre o direito à liberdade de expressão, que aposta na possibilidade de restrição do direito em casos específicos e mediante previsão legal (SILVA; BOLZAN; CIGANA, 2019, p. 233). Seguindo essa linha, a Corte Interamericana adota como critério uma análise tripartida para a restrição da liberdade de expressão, que envolve: necessidade de a restrição estar definida de forma precisa e clara por meio de lei; busca da realização de objetivos imperiosos autorizados pela Convenção Americana; e a presença, em uma sociedade democrática, de limitações proporcionais à finalidade perseguida e idôneas para alcançar o objetivo imperioso que se pretende resguardar (SILVA; BOLZAN; CIGANA, 2019, p. 233-234).

No âmbito do direito comparado, cabe destacar a posição adotada pelos Estados Unidos da América, um dos pioneiros no reconhecimento jurídico desse direito, mediante a Declaração de Virgínia de 1776 e, posteriormente, em sua Constituição, com a 1ª Emenda, em 1791. Observa-se no direito americano uma feroz defesa da liberdade de expressão, principalmente, no caso do discurso político (SARMENTO, 2006, p. 6).

No mesmo sentido, Rowland (2005, p. 59) destaca que o direito americano trabalha com níveis diferentes de proteção da liberdade de expressão, trazendo como núcleo forte o discurso político, em prol do reforço de seu caráter instrumental para a democracia. O autor ainda destaca um segundo nível de proteção, onde se encontra o discurso comercial, que recebe proteção menor do que o discurso político, e, por fim, um terceiro pavimento, no qual a proteção do discurso é excluída, como no caso de falsa informação e discurso inverídico (ROWLAND, 2005, p. 60).

Uma vez reconhecido que o direito fundamental de liberdade de manifestação não possui caráter absoluto e que a doutrina, a jurisprudência e os dispositivos internacionais trabalham com a necessidade de sua limitação e restrição, em prol de outros direitos fundamentais, cabe analisar a questão específica do discurso de ódio.

1.2 DISCURSO DE ÓDIO

O discurso de ódio surge como um ponto central de análise da relativização do direito fundamental de livre manifestação do pensamento, pois representa uma zona de conflito com outros direitos fundamentais, como igualdade e honra, diante do intuito de ataques baseados na intolerância e que fomentam a privação de direitos e a exclusão social de indivíduos ou grupos específicos, geralmente associados ao preconceito de raça, de gênero, de orientação sexual e de origem/nacionalidade.

Como base conceitual deste artigo, adota-se a caracterização do discurso de ódio promovida por Longhi (2020, p. 166), que identifica nesse tipo de discurso as menções difamatórias e degradantes à raça, à religião, à origem, ao gênero, à condição social ou à aparência física de um grupo de pessoas ou de indivíduos, bem como incitações ao ódio ou ao uso do próprio discurso como instrumento de discórdia e de promoção de ataques violentos entre grupos sociais ou símbolos nacionais.

Para a caracterização do discurso de ódio, é necessário analisar, além do seu conteúdo, a forma e o tom empregado pelo emissor, assim como sua motivação. Tais características, que permeiam o discurso, permitem identificar situações de abuso do direito de liberdade de expressão, na medida em que importem em ataques a outros direitos fundamentais previstos na Constituição da República e que deterioram o ambiente democrático.

Não se deve perder de vista que todo discurso envolve dois lados – um emissor e um receptor – e que um ambiente saudável de debate deve promover amplo acesso e oportunidades para interação. Os debatedores devem agir de forma livre, porém lastreados no respeito aos direitos dos outros e no objetivo de manutenção do diálogo democrático.

Sarmiento (2006, p. 31, 33-34) destaca que o discurso de ódio inviabiliza um ambiente de respeito mútuo entre os debatedores, pois se caracteriza como um ataque, contra o qual a vítima se lança em uma espiral em torno de dois comportamentos: revide com a mesma violência ou retirada do ambiente de discussão, diante do medo e da humilhação sofrida. No primeiro caso, a paz social resta prejudicada pelo ambiente de disputa do espaço público, e, no segundo caso, a vítima se sente desprotegida pelo Estado e abandona a esfera pública. Assim, segundo o autor, a liberdade de expressão não deve comportar o discurso de ódio, pois, além do aspecto moral, o preconceito e a intolerância veiculados em seu conteúdo não contribuem para um debate racional, gerando o comprometimento da continuidade da discussão (SARMENTO, 2006, p. 31-32).

Nessa linha de raciocínio, o discurso de ódio representa um limite da proteção ao direito de liberdade de expressão, uma vez que exorbita das próprias matrizes que sustentam o discurso livre, atentando contra a dignidade humana e deteriorando a participação social e o ambiente democrático.

Atenta aos efeitos nefastos do discurso de ódio, uma série de instrumentos internacionais estabelece a necessidade de seu combate, tais como a declaração e o plano de ação que foram redigidos na III Conferência Mundial de Combate ao Racismo, Discriminação Racial, Xenofobia e Intolerância Correlata, ocorrida em Durban, África do Sul, em 2001 (DECLARAÇÃO..., 2001).

No âmbito do direito comparado, a discussão sobre o discurso de ódio é vasta e varia de acordo com o regramento de cada país. É possível destacar a posição adotada pelos Estados Unidos da América, em prol de uma ferrenha defesa da liberdade de expressão, porém não absoluta, cedendo nos casos em que a doutrina norte-americana nomeia como *fighting words*, em que as palavras podem desencadear uma imediata reação violenta, diante da necessidade de proteger não a pessoa ofendida, mas de garantir a ordem e a paz públicas (SARMENTO, 2006, p. 9).

Em um sentido diferente, observa-se o ordenamento jurídico alemão, que aposta no combate ao discurso de ódio, principalmente, nas questões de interesse público e nas quais se justifique uma proteção mais intensa dos direitos da personalidade de minorias em situação de desvantagem. A concepção alemã reside na ponderação de interesses em conflito, lastreada em um sistema axiológico centrado no princípio da dignidade humana (SARMENTO, 2006, p. 25). Sarmiento (2006, p. 21, 33) ainda destaca que aquele país trabalha com o conceito de “democracia militante”, que demanda uma atuação do Estado contra os inimigos que tentam subverter a democracia por meio de um discurso radicalmente antidemocrático, diante do temor de que a liberdade de expressão seja utilizada para que o grupo chegue ao poder e depois consiga abolir a própria democracia.

No avanço da proposta deste trabalho, após identificar a liberdade de expressão e as restrições decorrentes do discurso de ódio, na próxima seção se discutirá a existência, no ordenamento jurídico brasileiro, de um dever de combate ao discurso de ódio.

2 DEVER FUNDAMENTAL DE COMBATE AO DISCURSO DE ÓDIO

Uma vez que o direito à liberdade de expressão não desfruta de hierarquia superior e nem serve como escudo de proteção para a perpetuação do discurso de ódio, a existência de um dever de combate desse tipo de discurso deve ser analisada à luz dos compromissos assumidos pela Constituição de 1988, de construção de uma sociedade igualitária e livre de preconceito, tal como assentado no art. 3º, incisos I (construir uma sociedade livre, justa e solidária) e IV (promover o bem

de todos, sem preconceitos de origem, raça, sexo, cor, idade e quaisquer outras formas de discriminação).

Conforme já destacado, o constituinte elencou no artigo 5º uma série de direitos fundamentais que, assim como a liberdade de manifestação do pensamento, decorrem do princípio da dignidade humana, tais como igualdade (*caput*); igualdade de gênero (inciso I); inviolabilidade da intimidade, da vida privada, da honra e da imagem (inciso X) e criminalização do racismo (inciso XLII) (BRASIL, 1988).

Esse cabedal de direitos fundamentais, para sua implementação, importa em obrigações por parte do Estado e da sociedade (ARRABAL; ENGELMANN; MELO, 2017, p. 58). Logo, a reivindicação de qualquer direito denota o cumprimento de ao menos um dever, tanto para os poderes públicos (deveres de proteção) como para as pessoas (deveres fundamentais) (LYRA et al., 2020, p. 74).

Por isso, em paralelo ao tratamento constitucional dos direitos fundamentais, os deveres fundamentais retiram seus limites no Estado de direito e na democracia, seguindo a noção de solidariedade e de forma a garantir a todos um tratamento sem discriminações injustificadas (BERNARDO SEGUNDO; PEDRA, 2011, p. 203).

Todavia, em que pese a imbricada relação mantida entre deveres e direitos fundamentais, é possível identificar uma autonomia e a ausência de uma ordem prevalente entre eles. A ótica da relação é o equilíbrio, de forma que cada cidadão exerça os seus direitos e cumpra os seus deveres, respeitando sempre os limites decorrentes da vida em sociedade e os direitos e deveres do próximo.

Assim, adota-se neste artigo o conceito de dever fundamental defendido por Lyra e outros (2020, p. 69):

[...] adotando-se conceito o instituto jurídico-constitucional [sic] com relativa autonomia em relação aos direitos, que exige juridicamente de todas as pessoas (físicas, jurídicas e coletivos despersonalizados) determinados comportamentos (fazer, não fazer, dar), umas perante as outras e todas perante a comunidade politicamente organizada a que fazem parte (Estado, inclusive), indistintamente, independentemente de capacidade (civil, penal, administrativa) e passíveis de sanção jurídica em virtude do seu descumprimento injustificado, voltadas a proporcionar as bases materiais para existência e funcionamento da sociedade e para a concretização dos direitos fundamentais de todos, decorrentes de uma ordem jurídica democrática, com posição de primazia normativa e controle de revisão (fundamentalidade formal), cujos conteúdos integram o estatuto da pessoa, formado por direitos e deveres fundamentais e orientado pela dignidade da pessoa humana (fundamentalidade material).

Outrossim, Sarmento (2006, p. 47) aponta que os direitos fundamentais reconhecidos na Constituição brasileira não se voltam exclusivamente para as relações entre o Estado e os indivíduos, diante da eficácia horizontal reconhecida, segundo a qual esses direitos também vinculam e obrigam

os particulares, ainda que de forma diferenciada, em razão da proteção constitucional conferida à autonomia privada. Assim, segundo o autor, o texto constitucional veda os atos de discriminação, de preconceito e de intolerância ao Estado e a todos os cidadãos e entidades privadas.

No caso da Carta Magna brasileira, observa-se que o constituinte condensou os direitos fundamentais até então elencados e estabeleceu no artigo 227, de forma expressa, o dever do Estado, da família e da sociedade, de proteção contra toda forma de discriminação, de violência, de crueldade e de opressão, instituindo ainda a prioridade em relação aos jovens, aos adolescentes e às crianças.

Diante desse arcabouço de normas constitucionais e considerando, principalmente, os objetivos elencados no artigo 3º do texto constitucional (BRASIL, 1988), é forçoso reconhecer que os direitos enumerados anteriormente também vinculam e obrigam os particulares, criando um verdadeiro dever de atuação contra o discurso de ódio, na medida em que este gera a segregação e a discriminação de pessoas determinadas ou grupos, em uma lógica de exclusão, de preconceito e de intolerância.

Uma vez identificada a existência desse dever, cabe continuar a análise deste artigo, por meio do estudo das plataformas de redes sociais, que despontam como importantes agentes na divulgação de conteúdos e na publicação de discursos.

3 INTERNET E REDES SOCIAIS

Perpassados os limites da liberdade de expressão, as características do discurso de ódio e o conflito entre ambos, o próximo passo é analisar a jornada desse embate em meio às redes sociais, identificando as peculiaridades do ciberespaço.

Inicialmente, cabe destacar que as redes sociais representam um fenômeno recente, fruto da revolução desencadeada pela tecnologia da informação (terceira revolução industrial). Um dos grandes pressupostos tecnológicos para sua existência se encontra no desenvolvimento da internet, cujos fundamentos técnicos tiveram origem na ARPANET, desenvolvida pelos norte-americanos com preocupações militares. A internet avançou com o desenvolvimento, pelo Núcleo de Pesquisas Nucleares da Europa (CERN), da rede nomeada de World Wide Web (WWW), o que levou à sua utilização em larga escala, a partir do início dos anos 1990 (SUNSTEIN, 2017, p. 182-183).

Conforme destacado por Longhi (2020, p. 11), a internet representa um conjunto de redes de computadores interconectados entre si, permitindo a troca de informações mediante um sistema de fragmentação e de remontagem. O autor ainda ressalta que a internet foi concebida, originalmente,

como uma estrutura aberta, o que garantiria, em tese, que ninguém seria seu dono, diante da ausência de uma entidade central ou instituição que financiasse seu desenvolvimento (LONGHI, 2020, p. 12).

A popularização da internet revolucionou a forma com que as pessoas interagem, comunicam-se e se informam, diante da criação de um novo espaço, o ciberespaço. Ela afetou desde o convívio em parques e bares, até o comércio de livros e jornais impressos e o papel do rádio e da televisão no dia a dia. Essa revolução foi incrementada com a chamada web 2.0, que transformou a internet em uma espécie de plataforma movida pelo usuário, lastreada no crescimento do mercado da gratuidade, sob a ótica da remuneração indireta, diante da aplicação de técnicas de *marketing* orientadas pelas preferências do consumidor (LONGHI, 2020, p. 20).

É importante destacar que, desde o seu surgimento, a internet ganhou aura de um lugar de liberdade. A esperança depositada no ciberespaço remonta a uma crença na neutralidade da rede, na aposta de desenvolvimento de um ambiente em que as informações possam ser trocadas entre os seus diversos usuários, que igualmente se colocam como emissores e receptores (SOUZA; SOLAGNA; LEAL, 2014, p. 144).

A internet conferiu voz a muita gente e se tornou um lugar em que as pessoas se sentem à vontade para expor suas ideias e opiniões, apoiadas, muitas vezes, em uma falsa impressão de irresponsabilidade, diante da ausência de um contato físico e da possibilidade de utilização do anonimato.

Todavia, a visão da internet como um lugar de liberdade e de amplas oportunidades não corresponde integralmente à realidade, na qual se constata a existência de grandes empresas que dominam o mercado e que desempenham grande papel na formação da opinião, entre as quais se destacam as plataformas de redes sociais.

3.1 DEFINIÇÃO DE REDES SOCIAIS

É importante elucidar aqui que as redes sociais não se confundem com a internet, cuja definição já foi apresentada anteriormente. Aquelas despontam como um tipo de serviço que se tornou muito popular neste século e que é gerenciado por grandes empresas privadas.

Uma clara definição de rede social é apresentada por Boyd e Ellison (2008, p. 211), para quem ela corresponde a um serviço ofertado na internet, no qual os indivíduos constroem seu próprio perfil (aberto ou não) e criam uma lista de outros usuários com os quais compartilham uma conexão, permitindo, assim, que eles se comuniquem entre si e que um visualize e compartilhe a lista de contatos do outro.

O grande impulso para o desenvolvimento das redes sociais deu-se a partir dos anos 2000, com a mudança da internet, capitaneada pela popularização da banda larga e dos *smartphones*, o que permitiu que as plataformas operassem com conteúdos criados pelos próprios usuários (NITRINI, 2021, p. 15).

Atualmente, as redes sociais são uma parte importante para divulgação e obtenção de conteúdo de grande parcela da população e o Brasil se encontra entre os países com maior utilização por parte de seus habitantes (POLIDO; MEYER, 2021). Nesse sentido, Nitrini (2021, p. 16) aponta que o Facebook® já bateu a marca de 2,41 bilhões de usuários ativos no mundo (correspondente a um quarto da população mundial), enquanto o Twitter® atingiu 300 milhões e o YouTube®, dois bilhões de usuários mensais ativos.

O que ressalta nesse tipo de serviço é a grande mudança em relação à disseminação de informações, uma vez que tais plataformas inovam e suplantam a tradicional distinção entre criadores e consumidores de informações (ROWLAND, 2005, p. 56). De acordo com Longhi (2020, p. 54, 163), as redes sociais são empreendimentos privados, nos quais os usuários atuam como um *webmaster* de si mesmo, enquanto o fornecedor atua como um provedor de hospedagem e se vale do conjunto de dados divulgados pelo usuário para contratar anunciantes, que oferecem produtos de acordo com as preferências que são declaradas.

Nesse contexto, as redes sociais surgem como o grande palco contemporâneo por meio do qual o discurso de ódio é disseminado. Isso porque essas plataformas contam com números expressivos de usuários, que cada vez mais se sentem à vontade para expor suas vidas e emitir suas opiniões sobre a vida alheia. Também favorece a divulgação do discurso de ódio o fato de que o universo de usuários engloba perfis e culturas diferentes e o ambiente, muitas vezes, permite o anonimato ou a criação de perfis falsos (não verificados).

Seguindo essa posição, Banks (2010, p. 233-234) assevera que o anonimato, a imediatidade e o alcance global da internet permitem sua utilização como ferramenta para a difusão do discurso de ódio, havendo, como consequência de sua popularização, o aumento de grupos de ódios que desenvolvem suas atividades *online*, inclusive por meio de redes sociais.

Assim, diante da identificação da natureza e das características das redes sociais, resta verificar quais são as regras a que estas estão sujeitas.

3.2 REGULAMENTAÇÃO DO CIBERESPAÇO

A internet e as redes sociais são uma realidade da sociedade contemporânea e inauguraram um novo espaço de convivência, o ciberespaço. Nele, assim como no mundo não virtual, conflitos de interesses aparecem, tal como no caso do discurso de ódio, e nem sempre as normas até então existentes são capazes de assegurar os interesses e valores conflitantes.

A ideia inicial da internet assentava-se em um espaço livre e de múltiplas possibilidades, do qual o governo não faria parte. Sunstein (2017, p. 178) conta que, em 1996, John Perry Barlow (um dos três fundadores da Electronic Frontier Foundation, entidade com o objetivo declarado de proteção, no contexto da era digital, dos direitos de liberdade de expressão) produziu uma declaração de independência do ciberespaço, na qual, entre outras coisas, afirmava que o governo não era bem-vindo e não desfrutaria de soberania nele. Desse modo, o surgimento do ciberespaço criou uma crença dirigida para uma espécie de extrema liberdade, que beira a anarquia. Segundo Lessig (2006, p. 19), para uma parcela da população, o ciberespaço deveria ser um lugar não sujeito à atuação dos governos, às suas leis e à política, cuja harmonia e desenvolvimento deveriam se apoiar na autorregulação e na autoconsciência dos seus usuários.

Todavia, Lessig (2006, p. 20-21) identifica que a arquitetura desenhada pelo ciberespaço caminha no sentido contrário ao da ideia inicial, pois a realidade demonstra não só a existência, mas uma constante busca pelo aperfeiçoamento do controle (códigos/algoritmos), capitaneado tanto pelos interesses dos governos quanto do mercado. Ele destaca que a luta não seria contra o governo, mas no sentido de assegurar que as liberdades essenciais sejam preservadas nesse ambiente de controle (LESSIG, 2006, p. 20-21). Ao se dirigir esforços para afastar o Estado do desenvolvimento dessa arquitetura, Lessig (2006, p. 116-117) acredita que a delimitação do que as pessoas podem ou não fazer será desenhada pelos interesses comerciais, em uma espécie de lei privada.

Lessig (2006) ainda defende que o código por trás da estrutura da internet representa valores e escolhas, não sendo uma questão apenas de engenharia. Portanto, seria um erro deixar a arquitetura do ciberespaço por conta apenas do mercado, pois a política é justamente o processo pelo qual as pessoas decidem coletivamente como devem viver (LESSIG, 2006, p. 117).

No caso específico das redes sociais, as suas plataformas surgem como grandes empresas privadas. A relação entabulada com os usuários é travada com base em uma espécie de contrato de adesão (termo do usuário). Longhi (2020, p. 37) argumenta que a relação estabelecida entre as empresas privadas de plataformas de redes sociais e seus usuários é de natureza consumerista, pois a relação não é simétrica, uma vez que usuários despontam como a parte mais vulnerável. Por

consequência, essa natureza consumerista traz reflexos na regulação da atividade, já que projeta, no Brasil, a aplicação do Código de Defesa do Consumidor.

As reflexões feitas até aqui permitem avançar neste trabalho, por meio da análise da regulamentação estatal existente e da atuação que compete, às plataformas de redes sociais, no combate ao discurso de ódio.

4 ATUAÇÃO DAS PLATAFORMAS DE REDES SOCIAIS DIANTE DO DEVER DE COMBATE AO DISCURSO DE ÓDIO

Até o presente momento, identificou-se que a liberdade de expressão não é um direito absoluto e que, constantemente, ela entra em colisão com outros direitos fundamentais, tal como na hipótese do discurso de ódio. No caso do discurso divulgado em plataformas de redes sociais, observa-se que as características desse tipo de serviço demandam, para o seu necessário combate, um olhar diferente.

A ideia inicial de que o ciberespaço seria um território sem lei não se comprovou, e a realidade indica grandes esforços existentes para aprimorar os mecanismos de controle. Destaca-se que o ambiente virtual trabalha com uma imbricada arquitetura capitaneada por códigos e algoritmos, mas eles serão sempre desenvolvidos por alguém e, conseqüentemente, servirão a um propósito.

O direito tradicional tem dificuldades de ser aplicado aos casos de discurso de ódio divulgado em redes sociais, pois a lógica empregada na relação entre autor, empresa responsável pela publicação e consumidor não se amolda completamente ao mundo virtual, no qual não resta clara uma linha editorial e nem há uma identificação da plataforma de rede social como responsável por selecionar os objetos de publicação, que são de forma quase instantânea produzidos e publicados pelos usuários.

Nesse sentido, Balkin (2018, p. 2.014-2.015) identifica um novo modelo pluralista, baseado na figura de um triângulo, no qual, em uma ponta, encontram-se os Estados-Nações e as organizações supranacionais; na outra, as empresas de infraestrutura de internet, incluindo as plataformas de redes sociais; e, na terceira ponta, têm-se organizações de mídia de massa, manifestantes, organizações da sociedade civil e *hackers*. O autor destaca que os Estados procuram controlar os falantes e as organizações de mídia de massa por meio de uma regulamentação do discurso baseada no velho modelo, que aposta em penalidades, prisão ou outras formas de punição ou de retribuição, enquanto tentam coagir e cooptar a infraestrutura da internet para o seu lado, por meio de novas regulamentações. Haveria ainda a atuação das infraestruturas da internet regulando os locutores privados e manifestantes, via técnicas de governança privada (BALKIN, 2018, p. 2.014-2.015).

É no âmbito dessa governança privada citada por Balkin que se encontra a atuação das plataformas de redes sociais no combate ao discurso de ódio, lastreada pelo dever fundamental já identificado neste artigo.

Antes de adentrar especificamente na análise da atuação das plataformas de redes sociais no combate ao discurso de ódio, cabe identificar primeiro a relação estabelecida entre o Estado brasileiro e essas empresas de infraestrutura da internet.

4.1 A RELAÇÃO ENTRE BRASIL E EMPRESAS DE INFRAESTRUTURA DE INTERNET E O DISCURSO DE ÓDIO

No Brasil, um divisor de águas na relação com as plataformas de redes sociais aconteceu com a Lei nº 12.965, de 23 de abril de 2014, conhecida como Marco Civil da Internet (BRASIL, 2014).

Antes da edição da referida lei, o Judiciário brasileiro foi instado a se manifestar em diversos casos de discursos de ódio proferidos em redes sociais. A jurisprudência caminhava para a adoção do modelo norte-americano, reconhecendo a responsabilidade da plataforma somente após a notificação feita pelo usuário (BRASIL, 2011).

Porém, o Marco Civil da Internet promoveu uma guinada, no caso do discurso de ódio, ao estabelecer que a responsabilidade da plataforma de rede social só surgiria em casos de descumprimento de uma decisão judicial. Os artigos 18 e 19 da referida lei isentam a plataforma de responsabilidade pelo discurso divulgado por terceiros, determinando a obrigação de agir somente após notificação judicial para tanto. Assim, percebe-se que o legislador brasileiro permaneceu sob orientação do modelo norte-americano, de forte privilégio da liberdade de expressão, porém incluindo mais um requisito para a exigência de atuação das plataformas de redes sociais: a notificação judicial.

Longhi (2020, p. 53) destaca que a jurisprudência brasileira e a legislação apostam em isenção de responsabilidade diante da caracterização das plataformas de redes sociais como prestadoras de serviço de fornecimento de hospedagem, e não como provedoras de conteúdo.

A mudança legal introduzida pelo Marco Civil da Internet, no seu artigo 19, é objeto de análise pelo Supremo Tribunal Federal, com repercussão geral reconhecida nos Recursos Especiais 1057258/MG e 1037396/SP. O questionamento feito nesses processos sob análise do Supremo passa pela falta de proteção gerada para a vítima, que dependeria da busca pelo Judiciário para cessar uma violação de seu direito, e pela análise da efetividade de eventual medida, diante da rapidez e da amplitude da divulgação ocorrida na internet (LONGHI, 2020, p. 81-82).

4.2 MEDIDAS ADOTADAS PELAS PLATAFORMAS DE REDES SOCIAIS PARA COMBATE AO DISCURSO DE ÓDIO

Retomando o modelo triangular apresentado por Balkin (2018), uma importante esfera de atuação no combate ao discurso de ódio nas redes sociais recai sobre as grandes empresas que disponibilizam o serviço, já que elas dominam a infraestrutura que mantém o ambiente virtual funcionando.

As medidas adotadas por essas plataformas no combate ao discurso de ódio estão lastreadas nos termos celebrados com seus usuários, que determinam quais condutas serão aceitas. Alguns fatores acabam norteando a atuação das plataformas no controle do discurso, tais como: crença nas normas de liberdade de expressão; senso de responsabilidade corporativa; e necessidade de inserir nas normas dos usuários regras que garantam uma viabilidade econômica (KLONICK, 2018, p. 1.618).

Um ponto de partida para entender a atuação dessas empresas no controle de discurso é de que redes sociais não são ambientes neutros, diante da expressiva atuação de filtros e de derrubadas de conteúdos que denotam um ambiente regido pela lógica do permitido/proibido e do visível/invisível (NITRINI, 2021, p. 79). Não se pode esquecer que códigos e algoritmos são desenvolvidos com a participação de trabalho humano e, portanto, representam escolhas e servem a um propósito. E aqui entram fatores que pressionam tais escolhas, como mercado, normas sociais, limitações da tecnologia e regras jurídicas de variados países (NITRINI, 2021, p. 129).

Em seu caminhar na atuação no combate ao discurso de ódio, as plataformas de redes sociais, inicialmente, adotaram o modelo capitaneado pelos Estados Unidos, apostando em um controle posterior, mediante sistema de denúncias por parte dos usuários. Esse tipo de mecanismo é conhecido como *flagging* e, geralmente, desencadeia um processo de revisão feito por moderadores humanos (NITRINI, 2021, p. 64).

Como exemplo, pode ser identificado o modelo do Facebook®, que segue uma lógica de atuação de moderadores humanos em forma de pirâmide, na qual o primeiro nível conta com revisores terceirizados que aplicam aos casos analisados as regras mais claras. O segundo nível abarca moderadores mais experientes, que supervisionam a atuação do primeiro nível e se encarregam de temas definidos como prioritários. Por fim, o último nível é composto por moderadores ainda mais experientes (KLONICK, 2018, p. 1.640-1.643).

Todavia, os incentivos da legislação norte-americana para o combate à pornografia infantil e para a defesa de direitos autorais demandou uma atuação das empresas do ramo no desenvolvimento de mecanismos de controle prévio (antes da publicação) – modelo que depois foi expandido para outras áreas de controle de discurso, como no caso do discurso de ódio (NITRINI, 2021, p. 53).

O controle prévio ocorre no intervalo de tempo entre o envio do conteúdo pelo usuário e sua efetiva publicação e, em regra, acontece de forma automatizada, mediante uma filtragem algorítmica sem revisão humana (KLONICK, 2018, p. 1.636). Nesse sentido, Nitrini (2021, p. 57) destaca que, em 2018, metade do conteúdo referente a discursos de ódio que foi removido proativamente pelo Facebook® ocorreu de forma automatizada, por meio de ferramentas para análise de linguagem e monitoramento de imagens.

O bloqueio geográfico desponta como outra forma de controle da publicação, em função da localização dos usuários, normalmente associada a seu país de origem e a peculiaridades culturais e legais locais (NITRINI, 2021, p. 60).

Em 2018, o Facebook® inovou ao criar um órgão independente, formado por pessoas externas à empresa, com a atribuição de tomar decisões finais no campo de moderação de conteúdo, chamado de comitê supervisor, que atua como uma espécie de corte de revisão (NITRINI, 2021, p. 143, 145).

Merecem ainda ser mencionadas algumas medidas mais graves adotadas para o controle de conteúdo, como a suspensão e o banimento de usuários, geralmente em casos de incitação de distúrbios. Exemplo recente aconteceu com o ex-presidente norte-americano Donald Trump, que teve suas contas suspensas pelo Facebook®, com decisão confirmada pelo comitê supervisor, conforme noticiado pela rede de notícias BBC (FACEBOOK..., 2021).

Esses mecanismos aqui indicados trabalham com a lógica do permitido/proibido; todavia, a engenharia por trás das redes sociais também controla o discurso por meio de outra lógica, a do visível/invisível, realizando direcionamento e determinando a maior exposição de dado conteúdo, e, assim, estimulando ou não a divulgação de informações sobre certo tema. O algoritmo trabalha para identificar preferências dos usuários e, dessa forma, estimular sua navegação.

Outro aspecto importante nesta etapa do trabalho é a identificação de como as plataformas de redes sociais definem o que deve ser bloqueado ou removido por caracterizar discurso de ódio. Os critérios são usados pelos moderadores e pelos processos automatizados aplicados por cada empresa. Assim, Klonick (2018, p. 1.631-1.632) identifica uma mudança de paradigma de governança – de, inicialmente, orientado por *standard* normativo, aberto e vago, para um novo e cada vez mais complexo sistema baseado em regras, com delimitações mais precisas.

4.3 O PAPEL QUE DEVE SER DESEMPENHADO PELAS PLATAFORMAS DE REDES SOCIAIS NO COMBATE AO DISCURSO DE ÓDIO

As redes sociais são uma realidade da vida cotidiana contemporânea e, por isso, o controle do discurso de ódio veiculado em suas plataformas é um tema atual e relevante.

À luz do tratamento constitucional brasileiro, a liberdade de expressão não possui caráter absoluto e pode ceder em casos concretos de colisão com outros direitos fundamentais, como no caso do discurso de ódio. Mas além da relatividade da livre manifestação do pensamento, a Carta de 1988 reconhece todo um sistema protetivo que coíbe o preconceito e a discriminação, em prol de uma sociedade justa e igualitária.

O texto do artigo 227 (BRASIL, 1988) é cristalino ao estabelecer que existe um dever de atuação não só do Estado, mas também da família e da sociedade, contra toda e qualquer forma de discriminação, violência, crueldade e opressão, que são justamente as notas definidoras de um discurso de ódio. Nessa ótica, surge o reconhecimento do dever de atuação das plataformas de redes sociais no combate ao discurso de ódio.

Nitrini (2021, p. 132), ao discorrer sobre o tema, trabalha com a noção de um constitucionalismo digital, como forma de atualização, para o ambiente digital, da noção de eficácia horizontal de direitos fundamentais entre particulares. O autor destaca a incidência de direitos fundamentais nas relações entre particulares – no caso, a liberdade de uma empresa de definir as regras de discursos em seus ambientes *versus* a liberdade de expressão dos usuários (NITRINI, 2021, p. 168).

Além do dever fundamental de atuação contra o preconceito e a discriminação, o modelo de controle de discurso de ódio implementado pelas plataformas de redes sociais segue a ótica de negócio, como uma exigência do mercado para angariar cada vez mais usuários, por meio da necessidade de manutenção de um ambiente seguro e atraente para seus consumidores (BALKIN, 2018, p. 2.022).

A atuação de plataformas de redes sociais no controle da divulgação do discurso de ódio é decorrência de uma nova realidade, em que técnicas tradicionais de regulação do discurso baseada em penalidades criminais e responsabilidade civil não são suficientes e demandam a colaboração de infraestrutura criada nessas plataformas (BALKIN, 2014, p. 2.298).

Por mais que seja uma realidade a atuação das plataformas de redes sociais no controle do discurso de ódio, as medidas adotadas são objetos de críticas e geram preocupações. Um dos

principais questionamentos diz respeito ao viés incorporado no efetivo controle do discurso por essas empresas, diante do risco de politização. Todavia, a neutralidade não é algo que se deve esperar dos algoritmos, pois exprimem ponderações feitas previamente pelos programadores (LONGHI, 2020, p. 90).

Klonick (2018, p. 1.661) destaca que tal preocupação com a neutralidade na moderação de conteúdo surge do fato de que, apesar de serem serviços prestados por empresas privadas, adquirem uma natureza essencial diante do destaque de algumas plataformas que detêm uma quota de mercado cada vez mais poderosa.

Entretanto, conforme identificado por Nitrini (2021, p. 168), se as redes sociais não tiverem um certo grau de liberdade de criação de regras para seus ambientes, haverá uma desconsideração do que elas são, o que fazem e qual o papel que possuem ou devem possuir na esfera da liberdade de expressão e do livre fluxo de informações na sociedade.

Outra crítica relevante diz respeito a uma falta de transparência dos critérios utilizados na moderação de conteúdo e ao fato de não haver um devido processo, no qual o usuário bloqueado ou removido possa argumentar e tentar reverter a ação da plataforma. Esse problema surge como motivação para o trabalho desenvolvido por entidades da sociedade civil e acadêmicos que editaram, em 2018, os Princípios de Santa Clara (THE SANTA..., 2021), os quais propõem a adoção das seguintes medidas por parte das plataformas de redes sociais: divulgação e disponibilização, ao público, de números para remoção de postagens; notificação dos usuários que tiverem suas postagens derrubadas ou contas suspensas, identificando a cláusula de moderação de conteúdo infringida e como se chegou à decisão (automático, controle humano, reclamação-denúncia); e previsão de apelação, mediante oportunidade de recorrer da decisão de derrubada ou suspensão do perfil (NITRINI, 2021, p. 141).

Apesar das críticas identificadas, a oposição à moderação do discurso, seja ela governamental ou não, apresenta-se como incoerente, diante da impossibilidade de ausência de uma regulação perante os ditames constitucionais. Portanto, o que deve ser discutido é o tipo de regulação que se quer, qual o processo e os valores a serem observados.

Nesse sentido, o importante não é buscar anular a atuação das plataformas de redes sociais na moderação do discurso, mas sim investir no fortalecimento da relação entre Estado e empresas de infraestrutura de internet. A atuação conjunta das duas esferas é essencial para o sucesso no combate ao discurso de ódio e para a manutenção do nível necessário de liberdade de expressão.

Um exemplo apontado por Nitrini (2021, p. 63) demonstra essa necessidade, no caso do vídeo conhecido como “inocência dos muçulmanos”, veiculado no YouTube®. O vídeo representava

uma sátira que causou revolta nos países muçulmanos e sofreu pressão do governo norte-americano para que fosse retirado do ar. A plataforma, inicialmente, negou e manteve o vídeo, porém, após os ataques sofridos por representações diplomáticas norte-americanas no Egito e na Líbia, a empresa bloqueou o acesso em ambos os países.

O exemplo demonstra que a moderação de conteúdo feita pelas plataformas precisa estar em consonância com a cultura, a religião e as disposições legais locais, pois o que está em jogo não é somente a manutenção, a perda e o crescimento de usuários, diante da possibilidade de desencadear ondas de violência como as observadas no caso citado.

Por outro lado, a retirada da autonomia das plataformas na moderação do discurso não parece ser uma estratégia eficiente. No caso do Brasil, a exigência legal de uma atuação das empresas somente após uma notificação judicial se mostra ineficiente. Isso decorre das características das redes sociais, que possuem um alcance e uma velocidade de divulgação muito grande e não compatível com o tempo necessário para obtenção de uma resposta judicial. As consequências para a vítima podem ser enormes e irreversíveis diante da demora do Judiciário para apreciar os conflitos a ele levados e para a eles responder.

O estágio atual de desenvolvimento do ciberespaço demonstra que o modelo de controle tradicional dos Estados não é totalmente aplicável nesses casos e que as empresas de infraestrutura da internet precisam ser estimuladas a cooperar com os Estados no combate eficiente ao discurso de ódio, pois somente elas desfrutam de meios rápidos e de custos menores para agir nas situações necessárias.

Conforme Balkin (2018, p. 2.019) ressalta, a cooperação entre a esfera pública e a privada representa uma das características da nova escola de moderação de conteúdo. Afinal, proteger a liberdade de expressão em uma era digital muitas vezes envolve soluções técnicas, regulatórias e administrativas que se aplicam em contextos em que a legislação tradicional não alcança (BALKIN, 2018, p. 2.032). O autor ainda destaca que os objetivos dessa cooperação entre público e privado devem ser orientados pela prevenção ou melhora da censura colateral e de novas formas de contenção digital prévia, assim como pela proteção das pessoas perante novos métodos de vigilância e de manipulação digital (BALKIN, 2018, p. 2.032-2.033).

Nesse ponto, convém retomar a posição exarada por Lessig (2006), de que a atuação governamental na regulação do ciberespaço deve ocorrer, pois, caso contrário, a lógica aplicada pelas empresas de infraestrutura de internet seguirá somente as exigências de mercado. A esfera governamental precisa atuar em parceria com as plataformas, pois nos governos ocorre o debate

político, o qual permite de forma democrática a definição dos valores e dos objetivos que devem nortear a vida em sociedade, inclusive no ciberespaço.

A atuação das plataformas de redes sociais na moderação do discurso é necessária, porém carece de aprimoramento, em prol de maior transparência, delimitação clara de critérios e procedimentos a serem adotados, e incorporação de mecanismos de revisão de decisões, tal como sugerido pelos Princípios de Santa Clara.

5 CONSIDERAÇÕES FINAIS

O objetivo deste artigo foi identificar o papel das plataformas de redes sociais no combate ao discurso de ódio. A discussão ganha relevo diante do grande espaço ocupado pelo meio digital no mundo atual e da crescente utilização das redes sociais como forma de publicação e de consumo de conteúdo.

Assim, é inegável que, ao permitir que os usuários criem e divulguem seus próprios conteúdos, as redes sociais se tornam o grande palco contemporâneo para a divulgação de discurso de ódio e, com isso, atraem para dentro de suas estruturas a discussão sobre o exercício do direito de liberdade de expressão e da necessidade de atuação na moderação do discurso.

Um dos pilares estabelecidos neste trabalho foi a identificação da existência, no texto constitucional brasileiro, de um dever de combater o discurso de ódio. Esse dever fundamental resultaria dos compromissos assumidos, pela Constituição de 1988, de construção de uma sociedade igualitária e livre de preconceito, e restaria consolidado na redação do artigo 227, que traz de forma expressa o dever, da sociedade, de proteção contra toda forma de discriminação, de violência, de crueldade e de opressão, estabelecendo, ainda, a prioridade em relação aos jovens, aos adolescentes e às crianças.

Outro pilar deste trabalho foi a verificação de que o direito de liberdade de expressão não possui caráter absoluto e nem posição superior em relação aos demais direitos fundamentais previstos na Constituição de 1988. Regulamentos internacionais, doutrina e jurisprudência brasileiras apontam a necessidade de a liberdade de expressão sofrer limites em casos de abusos, em prol da manutenção do debate democrático e harmônico. O discurso de ódio surge como um ponto central de análise dessa relativização, pois representa um ataque não só às pessoas, mas à democracia, diante do intuito de intolerância e do fomento da privação de direitos, exclusão social de indivíduos ou grupos específicos.

Também foi debatido ao longo do artigo a ideia de internet como um espaço sem lei, que serviu de atrativo para que as pessoas publiquem e exponham suas vidas e emitam opiniões sobre as

vidas alheias. Todavia, a realidade comprovou que o ambiente virtual também sofre controles, servindo a interesses tanto do mercado como dos governos.

Não se deve esquecer que as plataformas de redes sociais são empresas privadas e que celebram, com seus usuários, contratos que estabelecem regras de utilização do serviço. A lógica empregada na pactuação é de mercado, pois, em que pese a gratuidade do acesso, essas empresas lucram indiretamente com a disponibilização de anúncios pagos. Assim, diante da necessidade de manter um ambiente atrativo para cada vez mais pessoas, tais empresas acabam desenvolvendo mecanismos de controle e de moderação de conteúdo, tanto de forma posterior, por meio de denúncias dos próprios usuários, como de forma prévia, por meio de filtros automatizados.

Essas regras de comportamento no ambiente virtual desenvolvidas pelas plataformas de redes sociais geram um modelo de governança privada, o qual, além de fatores culturais, sociais e legais, também envolve aspectos tecnológicos, representados pela engenharia do código/ algoritmo.

Apesar da existência de desconfiança em relação à moderação de discurso realizada pelas plataformas de redes sociais, principalmente, diante da falta de transparência e definição clara quanto aos procedimentos e critérios usados, assim como pelo temor de um viés político, a realidade demonstra que essa moderação veio para ficar.

De fato, diante do dever, reconhecido pela Constituição de 1988, de combater toda e qualquer forma de discriminação, de opressão e de crueldade, não há como questionar se a atuação das plataformas de redes sociais encontra, na mediação do discurso, respaldo constitucional, diante da eficácia horizontal e da previsão expressa, no artigo 227, de tal dever para toda a sociedade.

Logo, a preocupação com a preservação da liberdade de expressão não deve ser canalizada contra a existência do controle, mas sim à delimitação da forma pela qual o controle deve ser exercido, buscando assegurar o seu exercício de forma transparente, com procedimentos e critérios claros e previamente divulgados, com a previsão de disponibilização de estatísticas e de mecanismos de revisão de decisões de moderação de conteúdo.

Por fim, foi identificado neste trabalho que, para se atingir o aprimoramento da forma de moderação de discurso e garantir que ela não seja efetivada somente sob a influência dos interesses do mercado, é importante o fortalecimento da relação entre Estado e empresas de infraestrutura de internet, sendo esse aprimoramento relevante porque a esfera pública representa o espaço para o debate político, no qual, de forma democrática, é possível delimitar os valores e objetivos a serem observados também no ciberespaço.

REFERÊNCIAS

- ALEXANDER, Robert. Colisão de Direitos Fundamentais e Realização de Direitos Fundamentais no Estado de Direito Democrático. Tradução Luís Afonso Heck. **Revista da Faculdade de Direito da UFRGS**, Porto Alegre, v. 17, 1992.
- ARRABAL, A. K.; ENGELMANN, W.; MELO, M. P. Liberdade e anonimato no contexto da cibercultura. **Revista de Direitos e Garantias Fundamentais**, Vitória (ES), v. 18, n. 2, p. 55-76, maio/ago. 2017.
- BALKIN, Jack M. Free Speech is a triangle. **Columbia Law Review**, New York, v. 118, n. 7, p. 2.011-2.055, 2018.
- BALKIN, Jack M. Old-School/New-School Speech Regulation. **Harvard Law Review**, [s. l.], v. 127, n. 8, p. 2.296-2.342, 2014.
- BANKS, James. Regulating Hate Speech Online. **International Review of Law, Computers & Technology**, [s. l.], v. 24, n. 3, p. 233 -239, 2010.
- BERNARDO SEGUNDO, Ronaldo Louzada; PEDRA, Adriano Sant'Ana. Limites ao dever de tolerância. In: BUSSINGER, Elda Coelho de Azevedo (org.). **Direitos fundamentais: pesquisa**. Curitiba: CRV, 2011. p. 201-208.
- BERTO, Matheus. O *cyberbullying* e a liberdade de expressão: uma proposta de análise dos limites que garantem a manutenção do convívio social. **Revista Tecer**, Belo Horizonte, v. 5, n. 8, p. 29-39, maio 2012.
- BOYD, Danah M.; ELLISON, Nicole B. Social network sites: Definition, history, and scholarship. **Journal of Computer-Mediated Communication**, [s. l.], v. 13, n. 1, article 11, p. 210-230, 2008.
- BRASIL. Constituição da República Federativa do Brasil de 1988. **Diário Oficial da União**, Brasília, DF, 5 out. 1988. Disponível em: <https://bit.ly/3KZO2BG>. Acesso em: 21 dez. 2020.
- BRASIL. Lei nº 12.965, de 23 de abril de 2014. Estabelece princípios, garantias, direitos e deveres para o uso da Internet no Brasil. **Diário Oficial da União**, Brasília, DF, 24 abr. 2014. Disponível em: <https://bit.ly/3OhWw9D>. Acesso em: 5 jul. 2021.
- BRASIL. Superior Tribunal de Justiça. Recurso Especial 1193764/SP. Relatora: Ministra Nancy Andrighi. Terceira Turma, Brasília, 14 dez. 2010. **DJe**, 08/08/2011. Disponível em: <https://bit.ly/3EzOuo4>. Acesso em: 5 jul. 2021.
- BRASIL. Supremo Tribunal Federal. Habeas Corpus 82424. Relator: Ministro Moreira Alves. Relator p/ Acórdão: Ministro Maurício Corrêa. Tribunal Pleno, 17 set. 2003. **DJ** 19-03-2004, PP-00024, EMENT VOL-02144-03, PP-00524. Disponível em: <https://bit.ly/3OmG7Az>. Acesso em: 26 jun. 2021.
- DECLARAÇÃO e o Plano de Ação emitidos na III Conferência Mundial de Combate ao Racismo, Discriminação Racial, Xenofobia e Intolerância Correlata. Durban, 2001. Disponível em: <https://bit.ly/3JWuRYr>. Acesso em: 27 jun. 2021.

FACEBOOK suspends Trump accounts for two Years. **BBC**, United States and Canada, 4 jun. 2021. Disponível em: <https://bbc.in/3L3EGVS>. Acesso em: 14 jun. 2021.

FARIAS, Edilsom Pereira de. **Liberdade de expressão e comunicação**: teoria e proteção constitucional. 2001. 287 f. Tese (Doutorado em Direito) – Programa de Pós-Graduação em Direito, Universidade Federal de Santa Catarina, Florianópolis, 2001.

KLONICK, Kate. The New Governors: The People, Rules, and Processes Governing Online Speech. **Harvard Law Review**, [s. l.], v. 131, p. 1.598-1.670, 2018.

LESSIG, Lawrence. **Code**: And Other Laws of Cyberspace, Version 2.0. New York: Basic Books, 2006. *E-book*.

LONGHI, João Victor Rozatti. **Responsabilidade civil e redes sociais** – retirada de conteúdo, perfis falsos, discurso de ódio e *fake news*. 1. ed. São Paulo: Foco, 2020.

LYRA, J. F. D. C. *et al.* A era dos deveres: a necessidade de um estatuto da pessoa humana para a eficácia social dos direitos fundamentais. **Revista Mexicana de Derecho Constitucional**, Ciudad de México, n. 43, jul.-dic. 2020.

NITRINI, Rodrigo Vidal. **Liberdade de expressão nas redes sociais**: o problema jurídico da remoção de conteúdo pelas plataformas. Belo Horizonte: Dialética, 2021.

POLIDO, Fabrício Bertini Pasquot; MEYER, Emilio Peluso Neder. Disinformation, Digital Platforms and COVID-19: Making State Agents Accountable in Brazil. **Int'l J. Const. L. Blog**, [s. l.], Mar. 26, 2021. Disponível em: <https://bit.ly/3MjwOQf>. Acesso em: 16 mar. 2022.

ROWLAND, Diane. Free Expression and Defamation. *In*: KLANG, Mathias; MURRAY, Andrew. **Human Rights in the Digital Age**. London: Glasshouse Press, 2005. p. 55-71.

SARMENTO, Daniel. **A liberdade de expressão e o problema do “hate speech”**. Rio de Janeiro, 2006. Disponível em: <https://bit.ly/388mDPK>. Acesso em: 2 abr. 2021.

SILVA, R. L. da; BOLZAN, B. E. T.; CIGANA, P. F. A liberdade de expressão e seus limites na Internet: uma análise a partir da perspectiva da Organização dos Estados Americanos. **Revista de Direitos e Garantias Fundamentais**, [s. l.], v. 20, n. 1, p. 219-250, 2019.

SOUZA, Rebeca H. V. de; SOLAGNA, Fabrício; LEAL, Ondina F. As políticas globais de governança e regulamentação da privacidade na internet. **Horizontes Antropológicos**, Porto Alegre, ano 20, n. 41, p. 141-172, jan./jun. 2014.

SUNSTEIN, Cass R. **#republic**. Divided Democracy in the Age of Social Media. Princeton, NJ: Princeton University Press, 2017.

THE SANTA Clara Principles: On Transparency and Accountability in Content Moderation. 2021. Disponível em: <https://bit.ly/3EpCo0y>. Acesso em: 23 maio 2021.