

REVISTA

DIREITO, INOVAÇÃO E REGULACÕES

I. DOCTRINA NACIONAL

2

A REGULACÃO CONSTITUCIONAL DO DISCURSO DE
ÓDIO NO FACEBOOK

THE CONSTITUTIONAL REGULATION OF HATE
SPEECH ON FACEBOOK

*Fausto Santos de Moraes¹
Álerton Poletto²*

1 Fausto Santos de Moraes. Doutor e Mestre em Direito (UNISINOS). Professor do Curso de Mestrado em Direito do PPGD/IMED, coordenador do Grupo de Pesquisas IAJUS – Inteligência Artificial e Direito. Sócio fundador da Associação Ibero-americana de Direito e Inteligência Artificial. Editor Chefe da Revista Brasileira de Direito e da Revista Brasileira de Direito e Inteligência Artificial. faustomorais@gmail.com

2 Álerton Emanuel Poletto. Mestrando em Direito do PPGD da Faculdade Meridional - IMED, bolsista CAPES/PROSUP, membro do Grupo de Pesquisa IAJUS - Inteligência Artificial e Direito. alertonep@gmail.com.

POLETTO, Álerton; MORAIS, Fausto Santos de. **A regulação constitucional do discurso de ódio no Facebook**. Revista Direito, Inovação e Regulações - Centro Universitário de Cascavel (UNIVEL). Cascavel. Jan. 2022; V. 1 (1): 15-30. DOI 10.5281/zenodo.14060625

RESUMO

Malgrado o Facebook ser um espaço virtual de livres manifestações, determinadas manifestações são vedadas na rede social, conforme a ordem constitucional brasileira. O discurso de ódio é reconhecido pela jurisprudência pátria como um limite ao pleno exercício da liberdade de expressão e de comunicação. Por discurso de ódio compreende-se toda a manifestação inflamatória e discriminatória com base nas características de raça, etnia, nacionalidade, religião, orientação sexual, sexo, gênero, identidade de gênero e deficiências. Desse modo, o objetivo deste trabalho é analisar se as regras de moderação de discurso de ódio, representadas pelos padrões da comunidade do Facebook, estão de acordo com os limites constitucionais fixados pelo Supremo Tribunal Federal (STF). Pela análise jurisprudencial do discurso de ódio, a proteção do Facebook está constitucionalmente alinhada ao que preceitua o STF, dada a pressão jurídica do constitucionalismo digital. O método de abordagem refere-se ao fenomenológico-hermenêutico e o procedimental é do tipo exploratório, monográfico e jurisprudencial.

Palavras-chave: constitucionalismo digital; discurso de ódio; Facebook; jurisprudência; Padrões da Comunidade.

ABSTRACT

Despite Facebook being a virtual space for free manifestations, certain manifestations are prohibited on the social network, according to the Brazilian constitutional order. Hate speech is recognized by national jurisprudence as a limit to the full exercise of freedom of expression and communication. Hate speech is understood to mean any inflammatory and discriminatory manifestation based on the characteristics of race, ethnicity, nationality, religion, sexual orientation, sex, gender, gender identity and disabilities. Thus, the objective of this work is to analyze whether the rules of hate speech moderation, represented by the Facebook community standards, are in accordance with the constitutional limits set by the Federal Supreme Court (STF). According to the jurisprudential analysis of hate speech, the protection of Facebook is constitutionally aligned with the provisions of the STF, given the legal pressure of digital constitutionalism. The method of approach refers to the phenomenological-hermeneutic and the procedural is exploratory, monographic and jurisprudential.

Keywords: digital constitutionalism; hate speech; Facebook; jurisprudence; Community Standards.

INTRODUÇÃO

As redes sociais, especialmente o Facebook, são ambientes virtuais de livres manifestações dos usuários. Contudo, alguns conteúdos publicados são limitados tanto pela ordem jurídica quanto pelas regras da própria plataforma de aplicação. Neste último caso, o Facebook constituiu os Padrões da Comunidade como regras e orientações aos usuários do que pode ou não circular na sua rede social.

Tendo em vista que o Facebook é uma das principais plataformas de divulgação de conteúdo, permitindo que os usuários conectados globalmente interajam, busca-se, portanto, debruçar o olhar às regras de proteção constitucional contra os discursos inflamatórios e discriminatórios, baseados nas características de raça, etnia, nacionalidade, religião, orientação sexual, sexo, gênero, identidade de gênero e deficiências.

Assim, o presente estudo tem como objetivo analisar a moderação de conteúdo pelo Facebook do conteúdo marcado como discurso de ódio, examinando as regras de moderação do Facebook em contraste com a jurisprudência do Supremo Tribunal Federal, de modo a perquirir se as regras do Facebook para a moderação de conteúdo marcado como discurso de ódio estão de acordo com os limites constitucionais fixados pelo Supremo.

Do ponto de vista procedimental, a pesquisa é do tipo exploratória, monográfica e jurisprudencial, partindo da exploração de obras de vários autores que debatem a temática abordada, configurando-se o *corpus* da pesquisa realizada. No que tange à linha de abordagem, para a elaboração da temática utilizou-se do método fenomenológico-hermenêutico, pela compreensão de como se dá a moderação do conteúdo e do tratamento jurídico dado ao problema do discurso de ódio nas redes sociais, de acordo com a jurisprudência pátria.

1 OS PADRÕES DA COMUNIDADE DO FACEBOOK

Inicialmente, cumpre apresentar os Padrões da Comunidade definidos pelo Facebook, a fim de ter uma compreensão de como a rede social delinea o que pode ou não ser publicado, relacionado à atividade de moderação de conteúdo que decorre da identificação de conteúdo violador na rede social.

As plataformas digitais se tornaram um meio de debate público e, de certa forma, até mesmo um instrumento de exercício da democracia. Contudo, há uma restrição nesse exercício, uma vez que as plataformas realizam a moderação do conteúdo que é publicado e definem o que pode ou não pode estar na rede.

A moderação de conteúdo nas plataformas digitais, aqui tratando-se especificamente do Facebook, por estar restringindo a liberdade de expressão virtual (compreendida como direito fundamental), deve estar em conformidade com a ordem constitucional brasileira, no que tange ao processo de relativização desses direitos.

Todo o conteúdo que é publicado no Facebook, seja por texto ou por imagens, submete-se à atividade de moderação pela plataforma. A moderação do conteúdo é compreendida como um conjunto de mecanismos de governança, utilizados por provedores das plataformas digitais, para classificar conteúdos publicados por usuários, a fim de prevenir abusos. A moderação opera como um filtro do que pode ou não estar disponível nas plataformas, podendo, assim, remover conteúdos, tanto de forma manual quanto utilizando de ferramentas automatizadas.

Para realizar tal atividade, o Facebook criou regras que orientam tanto os moderadores de conteúdo, quanto os usuários, do que é permitido publicar na rede social. O processo de moderação de conteúdo pelo Facebook tem como estratégia gerencial “remover, reduzir e informar”. Isto é, a plataforma tem como preceito remover o conteúdo prejudicial, contrário às políticas da rede social, reduzir a distribuição de conteúdo problemático que não viola as políticas do Facebook e informar as pessoas com contexto adicional para que possa decidir o que clicar, ler ou compartilhar¹

Ainda segundo os Padrões da Comunidade do Facebook, a identificação de conteúdo violador na rede social importa no esforço tecnológico de inteligência artificial da plataforma, associado às equipes de revisão. Essa tecnologia detecta e remove proativamente a grande maioria dos conteúdos violadores antes que os usuários os denunciem.

Bem assim, as orientações para a moderação de conteúdo pelo Facebook pautam-se nos Padrões da Comunidade desenvolvidos pela plataforma, com o intuito de explicitar as diretrizes de conteúdos de podem circular na plataforma. Desse modo, a moderação de conteúdo recaíra sobre conteúdos contrários aos padrões, baseado nas áreas de tecnologia, segurança pública e direitos humanos. (FACEBOOK, 2021)

Nessa esteira, o objetivo da criação dos Padrões da Comunidade pelo Facebook justifica-se da seguinte forma:

¹ Compreensão exposta nos Padrões da Comunidade do Facebook, na aba Discurso de Ódio, disponível em: <<https://transparency.fb.com/pt-br/policies/community-standards/hate-speech/>>.

O objetivo de nossos Padrões da Comunidade sempre foi criar um espaço para expressão e dar voz às pessoas. Isso não mudou e não mudará. Criar uma comunidade e unir pessoas depende da capacidade de compartilhar diversos pontos de vista, experiências, ideias e informações. Queremos que as pessoas possam falar abertamente sobre os assuntos importantes para elas, ainda que sejam temas que geram controvérsias e objeções. Em alguns casos, permitimos conteúdo que poderia ir contra os nossos Padrões da Comunidade para fins de conscientização pública, caso ele seja interessante e de utilidade pública. Para fazer esses julgamentos, consideramos o valor do interesse público e o risco de dano. Também observamos os padrões internacionais relativos a direitos humanos. Dessa forma, avaliamos a importância do conteúdo publicado por alguém, incluindo organizações de notícias e usuários individuais. Por exemplo, permitimos conteúdo que retrate graficamente uma guerra ou as consequências dela quando isso é importante para o discurso público. (FACEBOOK, 2021)

Apesar de a plataforma objetivar a promoção da comunicação online, as limitações de conteúdo possuem um interesse comercial. Esse interesse “é uma razão central para que empresas privadas de mídia social proíbam manifestação que não é ilegal e, portanto, limitem a liberdade de expressão muito além do exigido pela lei” (HARTMANN; SARLET, 2019, p. 96).

Nessa esteira, os Padrões da Comunidade classificam os conteúdos publicados em 25 tópicos dentro de seis diretrizes, representadas na Figura 1:

Figura 1 - Padrões da Comunidade do Facebook.



Fonte: Elaborado pelo autor, 2021.

A primeira diretriz, chamada “Violência e comportamento criminoso”, engloba todas as publicações na plataforma que contenham violência e comportamento criminosos de modo a evitar a circulação desse tipo de conteúdo, a fim de evitar danos reais e potenciais aos usuários. (FACEBOOK, 2021). Essa diretriz subdivide-se em: a) violência e incitação; b) organização e indivíduos perigosos; c) coordenação de danos e divulgação de crimes; d) produtos controlados; e) fraude e dolo.

Na diretriz de “Segurança” o Facebook afirma o seu compromisso em fazer da plataforma um lugar seguro, onde sejam moderados e excluídos conteúdos com ameaças, com intimidações e com exposição dos usuários a determinados conteúdos que possam causar riscos à própria segurança física ou mental, tais como: a) automutilação e suicídio; b) exploração sexual, abuso ou nudez infantil; c) exploração sexual de adultos; d) *bullying* e assédio; e) exploração humana; f) violações de privacidade e direitos de privacidade de imagem.

Já na diretriz dos “Conteúdos questionáveis” encontram-se publicações com conteúdos controversos que requerem uma análise profunda das nuances e dos contextos da publicação, a fim de realizar as distinções entre o debate lícito e a violação dos limites de assuntos, sendo as subdivisões: a) discurso de ódio; b) violência e conteúdo explícito; c) abordagem sexual; d) nudez adulta e atividades sexuais.

Outra importante diretriz diz respeito à “Integridade e autenticidade”, que compreende a atuação da plataforma na garantia da segurança cibernética dos usuários e das

contas contra perfis falsos, *spam* e manipulações de conteúdo. Essa diretriz subdivide-se em: a) integridade da conta e identidade autêntica; b) *spam*; c) segurança cibernética; d) comportamento não autêntico; e) notícias falsas; f) mídia manipulada; g) perfil memorial.

Em atenção aos direitos autorais, essa diretriz disciplina as publicações sobre o uso de marcas comerciais e outros direitos legais alheios, bem como os problemas relativos à propriedade intelectual que possam decorrer do uso da rede social. Por fim, o último ponto dos Padrões da Comunidade refere-se às “Solicitações e decisões relativas a conteúdo”, como requerimentos ao Facebook para a revisão das decisões de moderação de conteúdo.

Cada diretriz apresentada firma orientações para os usuários do conteúdo violador que se enquadra em determinado aspecto da política. Portanto, se determinado usuário publicar algo que se enquadre como conteúdo proibido pelos Padrões da Comunidade do Facebook, terá a sua publicação moderada e restringida.

Dessa forma, os Padrões da Comunidade do Facebook representam as regras de conteúdo permitido na rede social para os usuários, bem como a matriz orientativa para a atividade de moderação de conteúdo, que consiste no filtro e no controle de todo o conteúdo que circula na plataforma, implicando na marcação e na remoção do conteúdo violador.

2 O DISCURSO PROIBIDO NO FACEBOOK

Em recente atualização das diretrizes apresentadas no título anterior, o Facebook buscou explicitar cada tópico das políticas. Assim, o objetivo desta seção é apresentar as regras de conteúdo marcados como discurso de ódio pelo Facebook, conforme a política do “Conteúdo Questionável”.

A política de moderação de conteúdo classificado como discurso de ódio inicia pela visão do Facebook de que permitir discursos de ódio na plataforma cria um ambiente de intimidação e exclusão, podendo ocorrer casos de violência no mundo físico (FACEBOOK, 2021).

Assim, a definição de como o Facebook compreende esse fenômeno, conforme o fundamento da política, traduz-se do seguinte modo:

Definimos discurso de ódio como um ataque direto a pessoas, e não a conceitos e instituições, baseado no que chamamos de características protegidas: raça, etnia, nacionalidade, religião, orientação sexual, casta, sexo, gênero, identidade de gênero e doença grave ou deficiência. Definimos ataques como discursos violentos ou desumanizantes, estereótipos

prejudiciais, declarações de inferioridade, expressões de desprezo, repugnância ou rejeição, xingamentos e apelos à exclusão ou segregação. Também proibimos o uso de estereótipos prejudiciais, que definimos como comparações desumanizantes que têm sido historicamente usadas para atacar, intimidar ou excluir grupos específicos, e que muitas vezes estão ligadas à violência no meio físico. (FACEBOOK, 2021)

Percebe-se que o Facebook estabeleceu os limites da classificação com base na proteção das características individuais dos usuários, considerando também o contexto regional e linguístico da publicação. A classificação do que se enquadra como discurso de ódio parte de três níveis de violação, dado que as pessoas possuem diferentes níveis de tolerância para falas sobre características (FACEBOOK, 2021).

Desta feita, o Facebook (2021) recomenda aos usuários que não publiquem, sob pena de remoção da publicação, em primeiro nível, conteúdos “visando um indivíduo ou grupo de pessoas (incluindo todos os subconjuntos, salvo os que descrevem o cometimento de crimes violentos ou ofensas sexuais)”, como:

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: Insetos, Animais culturalmente percebidos como inferiores física ou intelectualmente, Sujeira, bactérias, doenças e excrementos, Predadores sexuais, Subumanidade, Criminosos sexuais e violentos, Outros criminosos (incluindo, entre outros, “ladrões” e “assaltantes de banco” ou afirmando que “todo [característica protegida ou semiprotégida] é ‘criminoso’”), Declarações negando a existência; **Deboche do conceito, de eventos ou de vítimas de crimes de ódio, mesmo que nenhuma pessoa real apareça na imagem; Comparações ou generalizações desumanizantes designadas, ou afirmações comportamentais (por escrito ou visuais) que incluam:** Pessoas negras e macacos ou seres semelhantes a macacos; Pessoas negras e equipamentos agrícolas; Caricaturas de pessoas negras com prática de *blackface*. Judeus e ratos, Judeus comandando o mundo ou controlando grandes instituições, como redes de mídia, a economia ou o governo, Negação ou distorção de informações sobre o Holocausto, Muçulmanos e porcos, Muçulmanos e relações sexuais com cabras ou porcos, Mexicanos e seres semelhantes a vermes, Mulheres como objetos domésticos ou referência à mulher como propriedade ou “objeto”, Pessoas transgêneras ou não binárias sendo chamadas de “isso”,

Dalits, pessoas de casta registrada ou de "casta inferior" como trabalhadores braçais. (FACEBOOK, 2021, grifo nosso)

Em um segundo nível de tolerância, o Facebook orienta que não seja publicado conteúdo visando um usuário ou um grupo, com base em características protegidas, que contenha:

Generalizações afirmando inferioridade (por escrito ou visuais) sob as seguintes formas: **Deficiências físicas são definidas em termos de:** Higiene, incluindo, entre outros, sujo, imundo, fedorento; Aparência física, incluindo, entre outros, feio, medonho. **Deficiências mentais são definidas em termos de:** Capacidade intelectual, incluindo, entre outros, burro, imbecil, idiota; Educação, incluindo, entre outros, analfabeto, atrasado; Saúde mental, incluindo, entre outros, doente mental, retardado, louco, maluco. **Deficiências morais são definidas em termos de:** Traços de personalidade culturalmente tidos como negativos, incluindo, entre outros, covarde, mentiroso, arrogante e ignorante; Termos pejorativos relacionados a atividades sexuais, incluindo, entre outros, vagabunda, vadia, pervertido. **Outras declarações de inferioridade, definidas em termos de:** Expressões sobre a falta de adequação, incluindo, entre outras, inútil, incapaz; Expressões sobre ser melhor/pior do que outra característica protegida, incluindo, entre outras: “Eu acredito que os homens são superiores às mulheres”; Expressões sobre desvio das normas, incluindo, entre outras, esquisito, anormal. **Expressões de desprezo (de forma visual ou escrita), definidas como:** Autoadmissão de intolerância com base em características protegidas, incluindo, entre outras, homofóbico, islamofóbico, racista; Expressões indicando que uma característica protegida não deveria existir; Expressões de ódio, incluindo, entre outras, desprezo, repulsa; Expressões de desaprovação, incluindo, entre outras, não respeito, não gosto, não me importo; Expressões de repulsa (de forma visual ou escrita), definidas como; Expressões que sugiram que o alvo causa náusea, incluindo, entre outras, vômito, regurgitação; Expressões de repulsa ou nojo, incluindo, entre outras, nojento, abjeto, eca. **Xingamentos, definidos como:** [...] Termos profanos ou frases com a intenção de insultar. (FACEBOOK, 2021, grifo nosso)

Já no terceiro nível da política de discurso de ódio, a rede social proíbe que seja publicado conteúdo, sob pena de remoção, conteúdo que “descreva ou prejudique pessoas com calúnias, definidas como palavras intrinsecamente ofensivas e usadas como frases ofensivas”, com qualquer dos seguintes termos (FACEBOOK, 2021):

Segregação sob a forma de conclamações, declarações de intenção, declarações intencionais ou condicionais, ou declarações que defendem ou apoiem a segregação. Exclusão por meio de conclamações, declarações de intenção, declarações intencionais ou condicionais, ou declarações que defendam ou apoiem a exclusão, definidas como: Exclusão explícita, como expulsar certos grupos ou dizer que eles não são permitidos; Exclusão política, como negar o direito à participação política; Exclusão econômica, no sentido de negar o acesso aos direitos econômicos e limitar a participação no mercado de trabalho; Exclusão social, como negar o acesso a espaços (físicos e online) e serviços sociais. (FACEBOOK, 2021).

Por fim, o Facebook exige informações e/ou contexto adicional dos usuários para a aplicação os Padrões da Comunidade, a fim de promover a moderação de determinados conteúdos que, de forma explícita, forneça ou ofereça produtos ou serviços com o objetivo de mudar a orientação sexual ou a identidade de gênero das pessoas ou, ainda, que ataca conceitos, instituições, ideias, práticas ou crenças associadas a características protegidas, que provavelmente contribuem para danos corporais iminentes, intimidação ou discriminação contra as pessoas associadas a essa característica protegida. (Facebook, 2021)

Para além disso, é importante destacar que:

O Facebook considera vários sinais para determinar se há uma ameaça de dano no conteúdo. Isso inclui, entre outros, conteúdo que poderia incitar violência ou intimidação iminente; se existe um período de alta tensão, como uma eleição ou conflito em andamento; e se existe um histórico recente de violência contra o grupo protegido. Em alguns casos, podemos também considerar se o orador é uma figura pública ou ocupa uma posição de autoridade. (Facebook, 2021)

O discurso de ódio na rede social ora analisada é tratado com complexidade e amplitude conceitual, delineando todo o âmbito de proteção dos usuários perante os diferentes níveis de tolerância, possibilitando, ainda, a eventual contextualização da publicação quando identificada como discurso de ódio, com o intuito de proteger, também, a liberdade de expressão. Essas regras devem, portanto, estar de acordo com o ordenamento jurídico.

3 O DISCURSO DE ÓDIO NA INTERNET SEGUNDO A JURISPRUDÊNCIA BRASILEIRA

Apresentada a visão do Facebook sobre o discurso de ódio, delineadas por meio das suas políticas próprias, cumpre examinar os limites do discurso permitido ou considerado de ódio no ordenamento jurídico pátrio, bem como a jurisprudência do Supremo Tribunal Federal acerca do tema. Esses tópicos serão abordados no presente capítulo.

Oportuno destacar que a temática reflete diretamente no direito à liberdade de expressão, uma vez que o discurso de ódio se manifesta como um abuso à liberdade de expressão exercida nas redes sociais. No Estado democrático, o direito à liberdade importa na liberdade de comunicação, a qual desdobra-se na liberdade de expressão ou liberdade de pensamento e na liberdade de informação (SILVA, 2014, p. 245).

A liberdade de informação pode ser dividida em liberdade geral de informação e liberdade de imprensa e de informação. No que tange à liberdade de comunicação, insta compreender que representa um “conjunto de direitos, formas, processos e veículos, que possibilitam a coordenação desembaraçada da criação, expressão e difusão do pensamento e da informação”. (SILVA, 2014, p. 245).

A liberdade de informação funda-se na procura, no acesso, no recebimento e na difusão de informações ou de ideias, independente do meio e sem submissão à censura. Em decorrência da liberdade de pensamento, a liberdade de informação abrange a liberdade de informar e de ser informado, que persegue o interesse da coletividade de que todos estejam informados no exercício consciente das liberdades públicas. (SILVA, 2014, p. 247-248)

Ainda segundo Silva (2014), não obstante, à liberdade de informação jornalística, enquanto meio de efetivação da liberdade de informar e ser informado, impõem-se o dever de divulgar informações corretas e parciais sem faltar com a verdade ou adulterar o sentido original, caso contrário, não será informação, mas deformação.

A liberdade de expressão representa um direito essencial à Democracia, que só existe pela consagração do pluralismo, da tolerância e do diálogo, sendo que a liberdade de expressão compreende as informações consideradas inofensivas, indiferentes ou favoráveis, como também aquelas que causam transtorno, resistência e inquietação. (MORAES, 2006, p. 207)

Nesse sentir, Barroso (2004, p. 18) leciona que a liberdade de expressão “destina-se a tutelar o direito de externar ideias, opiniões, juízos de valor, em suma, qualquer manifestação do pensamento humano”.

Percebe-se que a liberdade de expressão possui posição preferencial ante a sua limitação na arquitetura constitucional brasileira (SARLET; SIQUEIRA, 2020), dado que a própria Constituição, desde a sua raiz, fixa os impeditivos legislativos materiais e formais,

conforme os §§ 1º e 3º do artigo 220 da Constituição Federal, e proíbe veemente a censura (§ 2º, art. 220 da Constituição Federal).

Por sua vez, a Declaração Universal dos Direitos Humanos de 1948 preceitua, no artigo XIX, que “todo ser humano tem direito à liberdade de opinião e expressão; este 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”.

Nesta senda, Balkin (2009) alerta acerca da preocupação da inserção dos valores da liberdade de expressão em um conjunto de preocupações do século vinte e um, que ele denomina de política de informação, bem como da necessidade de uma teoria da liberdade de expressão adequada às mudanças tecnológicas, uma vez que, atualmente, é possível transmitir e publicar conteúdos para um grande número de pessoas em todo o mundo. Assim, nos tornamos produtores ativos de conteúdo e não apenas destinatários ou consumidores.

Feitas breves considerações sobre a liberdade de expressão, convém indicar que a Constituição Federal de 1988 firmou a garantia da liberdade de expressão e a proibição ao preconceito e à discriminação. Contudo, quanto aos princípios que envolvem tais premissas, é necessário analisar as manifestações do Supremo Tribunal Federal (STF).

Segundo um conceito doutrinário sobre o discurso de ódio, este pode ser lido como a “manifestação ou expressão, motivada por preconceito ou intolerância, através da qual uma pessoa ou um grupo é discriminado, com base em suas características identitárias”. (ANDRADE, 2020, p. 151)

Pacificamente, a doutrina afirma que há incompatibilidade entre a liberdade de expressão e os discursos de ódio, vez que “não restam dúvidas de que esse tipo de discurso não está protegido pela liberdade de expressão, sendo, inclusive, constitucionalmente reprovado com o adicional de imprescritibilidade”. (TASSINARI; JACOB DE MENEZES NETO, 2014, p. 32)

Dito isso, cumpre explicar o entendimento do Supremo Tribunal Federal acerca do discurso de ódio. Em uma pesquisa jurisprudencial na plataforma digital do STF, utilizando os termos “discurso de ódio” e “*hate speech*”, obteve-se o resultado de 7 acórdãos relevantes ao assunto.

A primeira decisão abordando o discurso de ódio foi proferida em 2003, no julgamento do caso Ellwanger (HC 82.424/RS). Em breve síntese fática, Siegfried Ellwanger fora condenado por ter publicado, na qualidade de autor e editor, obras apontadas como racistas e discriminatórias contra o povo judeu.

A tese fixada pelo STF no julgamento supracitado impôs limites ao exercício da liberdade de expressão em não abrigar, em sua abrangência, “manifestações de conteúdo imoral que implicam ilicitude penal”, dado que o preceito fundamental de liberdade de expressão não consagra o direito à incitação ao racismo. Desse modo, a primeira manifestação do STF estabeleceu que o discurso de ódio consiste na manifestação que decorre de ilicitude penal.

O julgamento de outro *habeas corpus*, de 2013, decorreu da tipificação de crime contra a honra contra manifestações injuriosas. Na decisão, a Primeira Turma entendeu que os limites da liberdade de expressão são de “coibir qualquer manifestação preconceituosa e discriminatória que atinja valores da sociedade brasileira, como o da harmonia inter-racial, com repúdio ao discurso de ódio” (Habeas Corpus 109.676, STF).

Na sequência, no julgamento do RHC 134.682 (2016), o STF preceituou que a liberdade religiosa e a de expressão não alcançam o discurso discriminatório criminoso, e fixou requisitos de materialização: caráter cognitivo, em que atestada a desigualdade entre grupos e/ou indivíduos; viés valorativo, no qual se assenta suposta relação de superioridade entre eles e, suposição de legítima dominação, exploração, escravização, eliminação, supressão ou redução de direitos fundamentais, encontrando-se desse modo, em situação desfavorável.

Na mesma esteira, com a decisão do RHC 146.303 (2018), a Segunda Turma firmou o entendimento de que o discurso que ataca a crença religiosa ou seus seguidores, no sentido de atingi-la, rebaixá-la ou desmerecê-la, não encontra proteção no direito à liberdade de crença religiosa, por ser essencialmente oposto.

De forma paradigmática, no julgamento da ADO 26 (2019), o STF estabeleceu uma nova implicação jurídica ao discurso de ódio. O Supremo construiu uma compatibilidade entre o crime de racismo e a repressão penal de homotransfobia, de modo a afastar o livre exercício da liberdade, seja de expressão ou religiosa, quando as manifestações configurarem 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” (ADO 26, 2019, STF).

Na análise entre a liberdade religiosa e a liberdade artística, o STF firmou UMA distinção entre a intolerância e a crítica, dado que a mera crítica, desde que não incite violência, é legítima e a “proibição de divulgação de determinado conteúdo deve-se dar apenas em casos excepcionalíssimos, como na hipótese de configurar ocorrência de prática ilícita, de incitação à violência ou à discriminação, bem como de propagação de discurso de ódio” (Reclamação Constitucional 38.782, 2020, STF).

Por fim, em recente julgado, o Pleno da Corte consolidou que a liberdade de comunicação social e de imprensa dever ser exercidas em harmonia com os preceitos constitucionais e dentro dos limites da discriminação e do discurso de ódio (ADI 5.418, 2021, STF).

Percebe-se que o Supremo Tribunal Federal, na arquitetura jurisprudencial, firma uma posição preferencial da liberdade de expressão, em suas diversas faces, ante a sua restrição ou limitação. Entretanto, tem-se claro que o exercício da liberdade de manifestação não encontra proteção nos discursos de ódio e incitações à violência.

Em uma análise com os Padrões da Comunidade do Facebook há pontos de convergência entre a análise de tolerância das publicações na rede social e a jurisprudência pátria. De fato, ao fazer a apreciação caso a caso, o Supremo Tribunal Federal estabelece preceitos mínimos de proibição de manifestações preconceituosas, que configurem ilícitos, independente do contexto discriminatório e de ódio.

4 A CONSTITUCIONALIZAÇÃO DAS REGULAGÕES DO FACEBOOK

A regulação do Facebook e da atividade de moderação de conteúdo, através dos Padrões da Comunidade, preceitua regras permissivas e proibitivas do que pode ou não ser publicado na rede social. Tais regras possuem um “tom constitucional” por fazerem uso do tradicional estilo de textos constitucionais e a adoção da configuração peculiar que articula conteúdos em termos de direitos, princípios e deveres (CELESTE, 2018, p. 3). Desse modo, esta seção buscará compreender a constitucionalização das regulações do Facebook, sob o olhar do constitucionalismo digital.

Fica evidente o “tom constitucional” dos Padrões da Comunidade, apresentados anteriormente, diante de uma análise formal e material. Do ponto de vista formal, o Facebook adotou uma redação tradicional de textos jurídicos, ao passo que do ponto de vista material, foram adaptados princípios constitucionais tradicionais, como a liberdade de expressão e a proteção contra o discurso de ódio no contexto da rede social, indicando a ideia de espaço democrático. (CELESTE, 2018, p. 4)

Isso pode ser explicado em razão das transformações no Direito que decorrem das implicações tecnológicas e a necessidade de rediscutir novos meios de proteger os direitos fundamentais dos indivíduos. A tecnologia gera alterações no ecossistema constitucional, o que “dificulta a possibilidade de lograr consensos sociais amplos voltados para a ordenação global do conjunto da sociedade por meio de instrumentos constitucionais”. (CALEJON, 2020, p. 589)

A proteção constitucional aos direitos exercidos no Facebook surge de “uma corrente teórica do Direito Constitucional contemporâneo que se organiza a partir de prescrições normativas comuns de reconhecimento, afirmação e proteção de direitos fundamentais no ciberespaço”. (MENDES; FERNANDES, 2020, p. 5)

Dentre as definições dessa corrente, tem-se que o Constitucionalismo digital corresponde, desse modo, a um termo guarda-chuva que corresponde ao esforço do movimento de afirmação de direitos fundamentais na internet, através de uma ideologia constitucional que se estrutura em um quadro normativo de proteção dos direitos fundamentais e de reequilíbrio de poderes na governança do ambiente digital, por meio de princípios e de valores que guiam o processo de produção normativa para a constitucionalização do ambiente digital (REDEKER; et al, 2018).

O constitucionalismo digital é “capaz de fornecer diretrizes normativas suficientes para guiar a sua aplicação” afastando a necessidade de judicialização do espaço digital como “régua da ação estatal na internet” (MENDES; FERNANDES, 2020, p. 5). Os termos e os Padrões do Facebook, sob a ótica do constitucionalismo digital, “não podem mais ser qualificadas como relações contratuais individuais, por aperfeiçoar características da legislação geral”, ante a necessidade de observar os princípios de direito constitucional (TEUBNER, 2017, p. 205).

O constitucionalismo digital está traduzindo e adaptando os valores e os princípios constitucionais existentes às peculiaridades da sociedade digital contemporânea (CELESTE, 2018). A institucionalização dessa corrente pode ser percebida através das regulações do Facebook, ao ponto de convergirem preceitos constitucionais nos Padrões da comunidade da rede social.

Resta claro que os Padrões do Facebook possuem um tom constitucional e perseguem os princípios constitucionais democráticos, que decorre da pressão jurídica do movimento de constitucionalização do ambiente digital para a proteção dos direitos dos indivíduos frente às regulações das redes sociais.

CONCLUSÃO

A rede social do Facebook firmou-se como um ambiente de compartilhamento de opiniões e de debates sobre os assuntos mais diversos, ainda que alguns gerem controvérsias e objeções. Todo o conteúdo que é publicado no Facebook submete-se à atividade de moderação, classificando os conteúdos e identificando eventuais violações.

A moderação de conteúdo é guiada pelas regras do Facebook do que pode ou não estar na rede social. Tais regras são identificadas como os Padrões da Comunidade que orientam os usuários e a atividade moderadora, seja ela por Inteligência Artificial ou humana. Nesses Padrões, dentro da diretriz de “Conteúdo questionável”, o Facebook aborda a limitação de publicações de discurso de ódio.

Na ordem constitucional brasileira, o discurso de ódio é tratado como uma limitação constitucionalmente adequada e legítima ao exercício da liberdade de expressão e manifestação, de modo a proteger os indivíduos contra a intolerância e discriminação de qualquer gênero. Pode-se afirmar que o discurso de ódio, segundo o Facebook e a jurisprudência pátria, portanto, é compreendido como toda a manifestação que incite a violência e ódio contra raça, etnia, nacionalidade, religião, orientação sexual, sexo, gênero, identidade de gênero e deficiências.

Assim, tem-se evidente que a definição dada pelo Facebook está constitucionalmente alinhada ao que preceitua o Supremo Tribunal Federal. Não obstante, o Facebook vai além e lista os diversos conteúdos que poderão ensejar a moderação de conteúdo, a fim de tornar claras e transparentes as suas regras e evitar a discricionariedade. Desse modo, as regras de remoção de conteúdo marcado como discurso de ódio pelo Facebook preceituam o mesmo objeto de proteção constitucional reconhecido pelo Supremo Tribunal Federal.

REFERÊNCIAS BIBLIOGRÁFICAS

BALKIN, Jack M. The Future of Free Expression in a Digital Age. *Pepperdine Law Review*, [s.l.], v. 36, n. 2, 2009. Disponível em: <http://digitalcommons.pepperdine.edu/plr/vol36/iss2/9>. Acesso em: 06 dez. 2021.

BARROSO, Luís Roberto. *Curso de direito constitucional contemporâneo: os conceitos fundamentais e a construção do novo modelo*. São Paulo: Saraiva, 2009.

BASSINI, Marco. Fundamental rights and private enforcement in the digital age. *Eur Law J.*, v. 25, 182–197, 2019.

BRASIL. Supremo Tribunal Federal. *Ação Direta de Inconstitucionalidade 5.418*. Relator: DIAS TOFFOLI, julgado em 11 de março de 2021. Disponível em: <https://jurisprudencia.stf.jus.br/pages/search/sjur446943/false>>. Acesso em 06 dez 2021.

BRASIL. Supremo Tribunal Federal. *Ação Direta de Inconstitucionalidade por Omissão 26*. Relator: CELSO DE MELLO, julgado em 13 de junho de 2019. Disponível em: <<https://jurisprudencia.stf.jus.br/pages/search/sjur433180/false>>. Acesso em 06 dez 2021.

BRASIL. Supremo Tribunal Federal. *Habeas Corpus 8.242*. Relator: MOREIRA ALVES, Relator p/ Acórdão: MAURÍCIO CORRÊA, julgado em 17 de setembro de 2003. Disponível em: <<https://jurisprudencia.stf.jus.br/pages/search/sjur96610/false>>. Acesso em 06 dez 2021.

BRASIL. Supremo Tribunal Federal. *Habeas Corpus 109.676*. Relator: LUIZ FUX, julgado em 11 de junho de 2013. Disponível em: <<https://jurisprudencia.stf.jus.br/pages/search/sjur238436/false>>. Acesso em 06 dez 2021.

BRASIL. Supremo Tribunal Federal. *Reclamação 38.782*. Relator: GILMAR MENDES, julgado em 03 de novembro de 2020. Disponível em: <<https://jurisprudencia.stf.jus.br/pages/search/sjur440937/false>>. Acesso em 06 dez 2021.

BRASIL. Supremo Tribunal Federal. *Recurso Ordinário em Habeas Corpus 134.682*. Relator: EDSON FACHIN, julgado em 29 de novembro de 2016. Disponível em: <<https://jurisprudencia.stf.jus.br/pages/search/sjur372168/false>>. Acesso em 06 dez 2021.

BRASIL. Supremo Tribunal Federal. *Recurso Ordinário em Habeas Corpus 146.303*, Relator: EDSON FACHIN, Relator p/ Acórdão: DIAS TOFFOLI, julgado em 06 de março de 2018. Disponível em: <<https://jurisprudencia.stf.jus.br/pages/search/sjur388361/false>>. Acesso em 06 dez 2021.

CALLEJÓN, Francisco Balaguer. REDES SOCIAIS, COMPANHIAS TECNOLÓGICAS E DEMOCRACIA. *REI - REVISTA ESTUDOS INSTITUCIONAIS*, [S.l.], v. 6, n. 2, p. 579-599, set. 2020.

CELESTE, Edoardo. Digital Constitutionalism: Mapping the Constitutional Response to Digital Technology's Challenges. *HIIG Discussion Paper Series*. v. 2018-02, 2018.

FACEBOOK. *Transparency Center - Padrões da Comunidade*. [S.l.] 2021. Disponível em: <<https://transparency.fb.com/pt-br/policies/community-standards/>>. Acesso em: 06 dez. 2021

FACEBOOK. *Padrões da Comunidade*. [S.l.] 2021. Disponível em:
<<https://www.facebook.com/communitystandards/introduction>>. Acesso em: 05 fev. 2022

HARTMANN, Ivar Alberto M.; 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, dez. 2019. Disponível em:
<<https://www.portaldeperiodicos.idp.edu.br/direitopublico/article/view/3755>>. Acesso em: 07 dez. 2021.

MENDES, Gilmar Ferreira; FERNANDES, Victor Oliveira. Constitucionalismo digital e jurisdição constitucional: uma agenda de pesquisa para o caso brasileiro. *Revista Brasileira de Direito*, Passo Fundo, v. 16, n. 1, p. 1-33, out. 2020. Disponível em:
<<https://seer.imed.edu.br/index.php/revistadedireito/article/view/4103>>. Acesso em: 01 dez. 2021.

MORAES, Alexandre de. *Direito constitucional*. São Paulo: Atlas, 2006.

NAÇÕES UNIDAS. *Declaração Universal dos Direitos Humanos de 1948*. Disponível em:
<<https://www.ohchr.org/EN/UDHR/Pages/Language.aspx?LangID=por>>. Acesso em 06 dez 2021.

REDEKER, Dennis; GILL, Lex; GASSER, Urs. Towards digital constitutionalism? Mapping attempts to craft an Internet Bill of Rights. *International Communication Gazette*, v. 80, n. 4, jun. 2018.

SARLET, Ingo Wolfgang; SIQUEIRA, Andressa De Bittencourt. LIBERDADE DE EXPRESSÃO E SEUS LIMITES NUMA DEMOCRACIA: O CASO DAS ASSIM CHAMADAS “FAKE NEWS” NAS REDES SOCIAIS EM PERÍODO ELEITORAL NO BRASIL. *Revista Estudos Institucionais*, v. 6, n. 2, p. 534-578, maio/ago. 2020.

SILVA, José Afonso da. *Curso de Direito Constitucional Positivo*. São Paulo: Malheiros, 2014.

TASSINARI, Clarissa; JACOB DE MENEZES NETO, Elias. Liberdade de expressão e Hate Speeches: as influências da jurisprudência dos valores e as consequências da ponderação de princípios no julgamento do caso Ellwanger. *Revista Brasileira de Direito*, Passo Fundo, v. 9, n. 2, p. 7-37, jan. 2014. ISSN 2238-0604. Disponível em: <<https://seer.imed.edu.br/index.php/revistadedireito/article/view/461/435>>. Acesso em: 03 dez. 2021.

TEUBNER, Gunter. Horizontal Effects of Constitutional Rights in the Internet: A Legal Case on the Digital Constitution. *THE ITALIAN LAW JOURNAL*, v. 3, n. 1, jul. 2017. Disponível em: <<https://www.theitalianlawjournal.it/teubner-1/>>. Acesso em: 18 out. 2021

Artigo recebido em: 30/09/2021
Artigo aceito em: 22/11/2021