

REVISTA  
EUROLATINOAMERICANA DE  
DERECHO ADMINISTRATIVO

VOL. 10 | N.1 | ENERO/JUNIO 2023 | ISSN 2362-583X



RED DOCENTE  
EUROLATINOAMERICANA  
DE DERECHO ADMINISTRATIVO



# O enfrentamento da desinformação e do discurso de ódio: um novo modelo regulatório para as redes sociais

## *Facing fake news and hate speech: a new regulatory model for social media*

VIVIAN CRISTINA LIMA LOPEZ VALLE I, \*

I Pontificia Universidade Católica do Paraná (Curitiba-PR, Brasil)

vivian.lima@pucpr.br

<https://orcid.org/0000-0002-5793-2912>

ANALÍA ANTIK II, \*\*

II Universidad Nacional de Rosario (Rosario, Argentina)

aantik@fder.unr.edu.ar

<https://orcid.org/0009-0005-1625-5715>

EDUARDO MAGNO CASSITAS CAVALCANTE DE LIMA I, \*\*

I Pontificia Universidade Católica do Paraná (Curitiba-PR, Brasil)

eduardomagno.cavalcante@gmail.com

<https://orcid.org/0000-0002-9430-2047>

**Recibido el/Received:** 17.01.2023 / 17 January 2023

**Aprobado el/Approved:** 18.05.2023 / 18 May 2023

### RESUMO

Através de uma análise bibliográfica e jurisprudencial, estudaram-se os modelos implementados nos Estados Unidos da América, na Alemanha e no Brasil para a regulação de redes sociais. Verificaram-se insuficiências em cada regime que obstam o combate efetivo da desinformação e

### ABSTRACT

Through bibliographical and jurisprudential analysis, the methods of regulating social media in the United States of America, Germany and Brazil were studied. All the regimes demonstrated flaws that impede the effective defence against Fake News and Hate Speech. Therefore, the

Como citar este artículo | *How to cite this article:* VALLE, Vivian Cristina Lima López; ANTIK, Analía; LIMA, Eduardo Magno Cassitas Cavalcante de. O enfrentamento da desinformação e do discurso de ódio: um novo modelo regulatório para as redes sociais. **Revista Eurolatinoamericana de Derecho Administrativo**, Santa Fe, vol. 10, n. 1, e239, ene./jun. 2023. DOI 10.14409/redoeda.v10i1.12372

\* Professora Titular de Direito Administrativo da Pontificia Universidade Católica do Paraná (Curitiba-PR, Brasil). Doutora e Mestre em Direito do Estado pela Universidade Federal do Paraná. Especialista em Contratação Pública pela Universidade de Coimbra. Especialista em Direito Administrativo pelo IBEJ. Coordenadora Adjunta do Curso de Direito da PUCPR. Membro do Instituto Paranaense de Direito Administrativo e do Instituto de Direito Romeu Felipe Bacellar. Advogada.

\*\* Profesora Asociada Ordinaria de la Cátedra de Derecho Administrativo de la Universidad Nacional de Rosario (Rosario, Argentina). Doctora en Derecho de la Universidad Nacional de Rosario. Directora de la Maestría en Derecho Público de la UNR. Docente estable de la mencionada Maestría y de las Carreras de Especialización en Derecho Inmobiliario, Urbanístico y de la Construcción, en Derecho Administrativo y en Derecho Sucesorio de la UNR.

\*\* Bacharel em Direito pela Pontificia Universidade Católica do Paraná (Curitiba-PR, Brasil).



do discurso de ódio. Assim sendo, sugere-se a adoção de um regime regulatório misto, através do qual o Estado e a plataforma de rede que hospeda o debate teriam competência concorrente para a regulação *ex-post* de conteúdo, de maneira que a liberdade de expressão possa ser otimizada, sem prejuízo ao respeito da dignidade da pessoa humana. O modelo proposto permite que não haja a exclusão da capacidade regulatória estatal meramente por se tratar de uma plataforma privada, ao mesmo tempo em que permite a ação tipicamente mais célere e eficiente da regulação pela plataforma em si.

**Palavras-chave:** regulação; redes sociais; fóruns cibernéticos; desinformação; discurso de ódio.

*adoption of a mixed regulatory model is suggested, through which the State and the platform hosting the debate would have the shared attribution to regulate content ex-post. This tactic would permit the maximization of freedom of speech, without disregarding respect for human dignity. Additionally, the proposed model would allow the State's regulatory capacity not to be excluded on the grounds of the host platform being a private entity, while still permitting the platform's self-regulatory practices (which are typically quicker and more efficient than governmental measures).*

**Keywords:** regulation; social media; cybernetic forums; fake news; hate speech.

## SUMARIO

**1.** Introdução. **2.** A regulação nos EUA. **3.** A regulação na Alemanha. **4.** A regulação no Brasil. **5.** O modelo de autoregulação. **6.** O modelo de regulação estatal (judicial). **7.** Um regime jurídico diferenciado: o modelo proposto para a regulação dos fóruns cibernéticos. **8.** Considerações finais. Referências.

## 1. INTRODUÇÃO

As transformações e evoluções sociais se dão de maneira muito mais veloz que avanços jurídicos, em razão dos trâmites legislativos. A revolução 4.0 só agravou esse cenário, impulsionando a velocidade de comunicação e de transformação social a níveis sem precedentes.<sup>1</sup> Essa aceleração traz muitos benefícios, mas também exacerbou problemas pré-existentes e traz consigo novos desafios. No ciberespaço, onde essa aceleração é ainda mais marcada, as capacidades regulatórias do Estado estão sob questionamento,<sup>2</sup> especialmente no que diz respeito às ferramentas de comunicação em geral e às redes sociais em particular. Percebe-se a necessidade urgente de resubstancializar a relação jurídica administrativa de controle, uma vez que a regulação eficiente das redes sociais, particularmente enquanto fóruns cibernéticos,<sup>3</sup> é absolutamente essencial.

<sup>1</sup> ROCHA, Leonel Severo, Paradoxos da auto-observação: percursos da teoria jurídica contemporânea. *apud* TACCA, Adriano; ROCHA, Leonel Severo, Inteligência Artificial: reflexos no sistema de Direito. **Revista do Programa de Pós-Graduação em Direito da UFC**, [s.l.], vol. 38, n. 2, p. 53-68, jul/dez. 2018.

<sup>2</sup> VALLE, Vivian Cristina Lima López; GALLO, William Ivan. Inteligência artificial e capacidades regulatórias do Estado no ambiente da administração pública digital. **A&C – Revista de Direito Administrativo & Constitucional**, Belo Horizonte, n. 82, p. 67-86, out./ dez. 2020.

<sup>3</sup> Fóruns cibernéticos é um termo utilizado neste trabalho para se referir coletivamente a todos os espaços virtuais cuja função primária é a difusão e discussão de informações e ideias. Abrange tanto fóruns de discussão online quanto algumas das funcionalidades de redes sociais. Este termo será abordado em maior profundidade na seção “2. OBJETIVOS”.



Os fóruns cibernéticos são amplamente utilizados para o relacionamento de cidadãos entre si, bem como entre cidadãos e a Administração Pública,<sup>4</sup> o que por si só torna essencial a sua regulação eficiente. Ademais, o uso inapropriado dessas tecnologias representa um grande risco às bases da democracia, uma vez que o direito ao voto, pilar da democracia representativa, está intrinsecamente ligado ao direito de acesso à informação.<sup>5</sup>

Essa ligação levou a Suprema Corte Suíça a anular em 2019 um referendo público de 2016, pela primeira vez na história moderna do país. A corte afirmou que, em razão dos resultados extremamente próximos e do oferecimento de informações incorretas e incompletas pelo Governo Federal, houve a violação do direito de voto dos eleitores, gerando a nulidade do resultado do referendo.<sup>6</sup> Assim sendo, a difusão *deliberada* de desinformação, visando alterar a percepção pública e impedir o voto informado, indubitavelmente constitui grave atentado à democracia.

O discurso de ódio, por sua vez, viola a moral e os sentimentos das vítimas, além de colocar em risco sua segurança, haja vista os amplos estudos demonstrando a correlação entre o discurso de ódio e a violência física contra seus alvos.<sup>7</sup> O genocídio da população muçulmana Rohingya em Myanmar em 2017/2018,<sup>8</sup> atribuído em parte a publicações de discurso de ódio por meio do Facebook, conforme admite a própria empresa,<sup>9</sup> serve para ilustrar a gravidade dessa ameaça.

<sup>4</sup> VALLE, Vivian Cristina Lima López; FELISBERTO, Jéssica Heinzen. Administração Pública digital: limites e possibilidades em atenção à desigualdade social e ao custo dos direitos. **Revista Eurolatinoamericana de Derecho Administrativo**, Santa Fe, vol. 9, n. 1, p. 151-179, ene./jun. 2022.

<sup>5</sup> ONO, Taynara Tiemi, A regulação da desinformação na internet em período de campanha eleitoral: leis, normas sociais, arquitetura de rede e atuação da justiça eleitoral como mecanismos regulatórios adotados no contexto brasileiro. **Revista dos Tribunais**, [s.l.], vol. 1008, p. 93-109. 2019.

<sup>6</sup> Julgamento final pela Suprema Corte Suíça da causa *Béglé contre Conseil d'État du Canton de Vaud*, de 10/04/2019, que tratava da taxação mais elevada de casais em comparação aos demais indivíduos, gerando a iniciativa popular *Pour le couple et la famille - Non à la pénalisation du mariage* ("Pelo casal e a família – não à penalização do casamento", em tradução livre). Inicialmente, o Governo afirmou que isso afetava aproximadamente 80.000 casais; posteriormente, verificou-se que o número era mais próximo de 454.000 afetados. O referendo foi rejeitado por uma margem de 0,8% dos eleitores. Bundesgericht, **Arrêt 1C\_338/2018**, vol. 1/2019 (145) – 207. Disponível em: < [https://www.bger.ch/ext/eurospider/live/fr/php/aza/http/index.php?highlight\\_docid=aza%3A%2F%2F10-04-2019-1C\\_338-2018&lang=fr&type=show\\_document&zoom=NO&](https://www.bger.ch/ext/eurospider/live/fr/php/aza/http/index.php?highlight_docid=aza%3A%2F%2F10-04-2019-1C_338-2018&lang=fr&type=show_document&zoom=NO&)>. Último acesso em 29 jan. 2021.

<sup>7</sup> MÜLLER, Karsten; SCHWARZ, Carlo. From Hashtag to Hate Crime: Twitter and Anti-Minority Sentiment (Nov. 2, 2019), disponível em <[https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3149103](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3149103)>. Último acesso em 13 de fevereiro de 2021; MÜLLER, Karsten; SCHWARZ, Carlo. Fanning the Flames of Hate: Social Media and Hate Crime (Nov. 3, 2019), disponível em <<https://ssrn.com/abstract=3082972>>. Último acesso em 13 de fevereiro de 2021; EDWARDS, Griffin; RUSHIN, Stephen. The Effect of President Trump's Election on Hate Crimes (Jan. 14, 2018), disponível em <[https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3102652](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3102652)>. Último acesso em 13 de fevereiro de 2021.

<sup>8</sup> WILSON, Richard A.; LAND, Molly K. Hate speech on Social Media: Towards a Context-Specific Content Moderation Policy. **Connecticut Law Review**, Connecticut, vol. 52. 2020. p. 15-17.

<sup>9</sup> "[...] prior to this year, we weren't doing enough to help prevent our platform from being used to foment division and incite offline violence. We agree that we can and should do more." WAROFKA, Alex. An Independent



Parte da dificuldade da regulação efetiva desses espaços advém da ampla utilização do modelo tradicional de Direito Administrativo, que faz distinção entre espaços públicos e aos privados, definindo dessa forma a competência (e responsabilidade) pela sua regulação. Em se tratando de espaços privados, sua regulação caberia aos agentes particulares, dentro dos limites da lei, e vice-versa. Os fóruns cibernéticos, todavia, não se enquadram nitidamente em nenhuma dessas duas categorias, pelo que essa dicotomia é de utilidade questionável na sua regulação.

Por um lado, são hospedados por empresas particulares, que tecnicamente detém seu domínio; por outro, possuem finalidades públicas, como proporcionar o debate político ou permitir a comunicação direta entre a Administração e os administrados. Apresentam características de ambas as categorias, além de particularidades próprias. Ademais, já foi demonstrado que tais ferramentas são suscetíveis de abuso tanto por entes privados, como no caso de redes de desinformação para manipular eleições ao redor do mundo,<sup>10</sup> como pelos públicos, no fenômeno conhecido como autoritarismo digital.<sup>11</sup>

Da mesma forma que a regulação eficaz de empresas públicas e sociedades de economia mista pressupõe a criação de um modelo regulatório distinto do regime público tradicional, a criação de um regime jurídico diferenciado é imprescindível para a regulação eficaz dos fóruns cibernéticos. O regime administrativo tradicional é apropriado para o exercício de funções estritamente públicas, e não para os raros casos em que o Estado explora diretamente a atividade econômica – da mesma maneira, os regimes de controle aplicáveis a espaços públicos ou privados não são apropriados para a regulação de espaços que não se enquadram em nenhuma dessas categorias. A criação de um regime jurídico diferenciado, deixando de lado as categorizações de espaços públicos e privados, é necessária para possibilitar a regulação efetiva dos fóruns cibernéticos e o combate da desinformação e do discurso de ódio nessas plataformas – por todos os agentes.

Para coibir ao máximo a possibilidade de abusos no uso dessas ferramentas sem restringir demasiadamente a liberdade de expressão, sustenta-se que será necessário

---

Assessment of the Human Rights Impact of Facebook in Myanmar. **About Facebook**, November 5, 2018, disponível em: <<https://about.fb.com/news/2018/11/myanmar-hria/>>. Último acesso em 30 jan. 2021.

<sup>10</sup> Acerca do uso de redes de desinformação para conquistar votos para Trump: ALCOTT, Hunt; GENTZKOW, Mathew, Social Media and Fake News in the 2016 Election. **Journal of Economic Perspectives**, [s.l.], vol. 31, n. 2, p. 211-236. 2017; acerca da manipulação inerente aos debates em plataformas como o Facebook e de “political micro-targeting”: BENNET, Colin J.; LYON, David. Data driven elections: implications and challenges for democratic societies. **Internet Policy Review**, vol. 8, n. 4, p. 1-16. 2019. DOI: 10.14763/2019.4.1433. No mesmo sentido, vide o documentário “O Dilema das Redes” (O dilema das redes. Direção de Jeff Orlowski. Netflix: 2020. 94 min.). Sobre o tema nas eleições presidenciais brasileiras de 2018: SANTOS, Gustavo Ferreira. Social media, disinformation, and regulation of the electoral process: a study based on 2018 Brazilian election experience. **Revista de Investigações Constitucionais**, Curitiba, vol. 7, n. 2, p. 429-449, maio/ago. 2020.

<sup>11</sup> WILSON, Richard A.; LAND, Molly K. Hate speech on Social Media: Towards a Context-Specific Content Moderation Policy. **Connecticut Law Review**, Connecticut, vol. 52. 2020. p. 29.



implementar um regime em que tanto as plataformas quanto o Poder Público detenhams poderes significativos de regulação *ex post*, podendo exercê-los concorrentemente. Dessa forma, haveria um sistema de freios e contrapesos, como no modelo de organização estatal democrático, minimizando a possibilidade de abusos por qualquer um dos lados. Ainda, em um modelo que priorize o controle *ex-post*, seja por meio da remoção de postagens, seja através da indicação de que o conteúdo de determinada publicação é falso ou duvidoso, minimizar-se-ão as restrições ao debate e à liberdade de expressão. Finalmente, outro aspecto que impede o combate efetivo à desinformação e o discurso de ódio é a sua indefinição; parâmetros mais objetivos precisam ser estabelecidos, ou a aplicação das normas regulatórias jamais será eficaz.

A fim de verificar a viabilidade e a efetividade de um modelo regulatório nos moldes sugeridos, o presente estudo analisará os métodos de regulação de redes sociais aplicados nos EUA, na Alemanha, e no Brasil. Pretende-se verificar os benefícios e malefícios inerentes a cada modelo, e se a alternativa proposta seria justificável em sua implementação.

Reconhecendo a amplitude de funcionalidades e aplicações de redes sociais, o estudo se limitará à análise de apenas um aspecto de determinadas redes, doravante denominado *fórum cibernético*: espaços virtuais cuja função *primária* é a *difusão e discussão de informações e ideias*. Para os fins desse trabalho, redes sociais e fóruns cibernéticos serão tratados como sinônimos, utilizando essa definição. O referido termo abrange tanto fóruns de discussão online, como algumas das funcionalidades de redes sociais. Com isso, excluem-se ferramentas de troca de mensagens instantâneas, como o Facebook Messenger ou chats individuais do WhatsApp,<sup>12</sup> debruçando-se exclusivamente sobre ferramentas como o Mural do Facebook e chats em grupo cuja finalidade se alinhe à da definição. Redes sociais com finalidades especializadas diversas, como o Tinder, também serão excluídas.

O termo é reconhecidamente imperfeito: é relativamente vago e a expressão “finalidade primária” carece de objetividade, o que dificultaria a eventual produção de provas.

Contudo, faz-se necessário adotá-lo a fim de delimitar adequadamente o escopo do estudo, excluindo ferramentas muito distintas, mas preservando a amplitude do estudo. Uma delimitação maior ainda, abordando somente uma ou duas redes sociais, não refletiria adequadamente os desafios de uma regulação efetiva genérica, privando o presente trabalho de qualquer valor prático.

---

<sup>12</sup> Embora o WhatsApp, como um todo, seja uma ferramenta muito mais próxima de serviços de mensagens instantâneas, é importante reconhecer a dimensão de fóruns cibernéticos de chats em grupo com essa finalidade. A regulação dessa ferramenta é absolutamente necessária, haja vista seu papel pronunciado na disseminação de desinformação, particularmente durante as eleições. Acerca do tema: EVANGELISTA, Rafael; BRUNO, Fernanda. WhatsApp and political instability in Brazil: targeted messages and political radicalisation. *Internet Policy Review*, [s.l.], vol. 8, n. 4, p. 1-23. 2019. DOI: 10.14763/2019.4.1434.



## 2. A REGULAÇÃO NOS EUA

O estudo iniciou por meio da análise do cenário estado-unidense, onde ficou claro que uma das maiores preocupações ali discutidas é a restrição da liberdade de expressão no combate à desinformação e ao discurso de ódio. A liberdade de expressão é o valor mais caro à democracia americana,<sup>13</sup> pelo que prevalece a tese de que o combate à desinformação e o discurso de ódio deve se dar através do livre debate. Há uma forte corrente doutrinária que defende que não deve haver censura de discurso de ódio, pois normas regulatórias de fala tendem a desfavorecer pontos de vista minoritários,<sup>14</sup> pelo que a liberdade de expressão deveria prevalecer. Tal afirmação se justifica pela história do país, onde essa forma de controle foi utilizada como censura; seja no que diz respeito a discursos abolicionistas, que “tinham o potencial de incitar violência e rebelião”, seja no que se refere a discursos socialistas.<sup>15</sup> Como exemplo recente, tem-se o movimento *Black Lives Matter*, contra a violência policial e a favor da valorização de vidas negras, que foi incluído pelo Comitê Republicano Nacional e por alguns órgãos estaduais dentro das resoluções de discurso de ódio. Em razão disso, nas palavras de Nadine Strosser, essa corrente acredita que “a cura é pior que a doença” no que diz respeito à regulação do discurso de ódio.<sup>16</sup>

Disso resulta a ausência de definição legal para discurso de ódio nos EUA. A Suprema Corte Americana reiteradamente rechaçou tentativas de estabelecer uma definição constitucional para discurso de ódio e excluir tais discursos da proteção do *First Amendment*, como nos julgamentos *Matal v. Tam (2011)* e *Snyder v. Phelps (2011)*.<sup>17</sup>

Outro empecilho para a regulação estatal da desinformação e do discurso de ódio nos EUA, paradoxalmente, é o próprio sistema legal. As únicas restrições de conteúdo vigentes naquele território são aquelas impostas pelas próprias plataformas, haja vista que o *First Amendment* americano expressamente preclui a “supressão de discurso com base em seu conteúdo ou ponto de vista”. A aplicação rigorosa dessa norma não permite implementar leis que combatam efetivamente a desinformação ou o discurso de ódio nas redes, onde o conteúdo em si dos discursos, aliado ao contexto da discussão, é um parâmetro indispensável. A regulação cabe exclusivamente às plataformas.

<sup>13</sup> EBERLE, Edward J. *Dignity and Liberty: Constitutional Visions in Germany and the United States*. Praeger Publishers, Westport, 2002. p. 190-191.

<sup>14</sup> WILSON, Richard A.; LAND, Molly K. Hate speech on Social Media: Towards a Context-Specific Content Moderation Policy. *Connecticut Law Review*, Connecticut, vol. 52. 2020. p. 11.

<sup>15</sup> WILSON, Richard A.; LAND, Molly K. Hate speech on Social Media: Towards a Context-Specific Content Moderation Policy. *Connecticut Law Review*, Connecticut, vol. 52. 2020. p. 11.

<sup>16</sup> STROSSEN, Nadine. *Hate: Why We Should Resist It With Free Speech, Not Censorship*. Oxford University Press, New York. 2018. p. 17.

<sup>17</sup> WILSON, Richard A.; LAND, Molly K. Hate speech on Social Media: Towards a Context-Specific Content Moderation Policy. *Connecticut Law Review*, Connecticut, vol. 52. 2020. p. 10



Todavia, não há qualquer incentivo legal para que as plataformas regulem o conteúdo ali postado, em razão da proteção oferecida pelo *Communications Decency Act*, de 1996. Esse dispositivo legal adveio de decisões conflitantes em cortes nova iorquinas acerca da responsabilidade de plataformas pelo conteúdo ali postado (*Cubby, Inc. v. CompuServe, Inc.* e *Stratton Oakmont, Inc. v. Prodigy Services Co.*)<sup>18</sup>. No primeiro caso, o raciocínio tradicional empregado em casos de difamação foi utilizado para escusar a *CompuServe*: a empresa foi equiparada à uma biblioteca, sem conhecimento ou responsabilidade do conteúdo específico de seus livros. No segundo, contudo, a suprema corte de Nova Iorque afirmou que, haja vista o grau de controle sobre o conteúdo de seu site, *Stratton* se aproximaria mais à uma editora que uma biblioteca, pelo que não seria possível excluir sua responsabilidade. O parâmetro para responsabilização passou a ser o grau de controle editorial exercido<sup>19</sup>, o que poderia resultar na ausência de autorregulação das plataformas a fim de afastar sua responsabilidade. Essa preocupação levou à ampliação do Ato, com a seção 230(c), que preclui a responsabilização de “provedores de serviços informáticos interativos” por gozarem de funções editoriais e de controle<sup>20</sup>, levando à criação de uma *imunidade federal generalizada* para essa classe de prestadores de serviço. A aplicação desse dispositivo resulta na ausência de responsabilidade das plataformas pelos conteúdos nela hospedados, independentemente do grau de controle exercido sobre eles.

A priori, a ausência de responsabilidade das empresas serviria como desincentivo ao combate ao discurso de ódio e à desinformação pelas redes sociais. Contudo, em razão de pressão social, a maioria dessas plataformas vêm proibido e removido discurso de ódio.<sup>21</sup> Através de regulamentos extensos, as plataformas implementaram “*platform law*”,<sup>22</sup> fazendo uso de algoritmos sofisticados e sistemas de moderação humana para remover conteúdos potencialmente abusivos. Estas “leis”, por não serem promulgadas pelo Estado, não estão sujeitas às limitações do *First Amendment* – estão completamente à discrição das plataformas, independentemente dos impactos de tais “leis” sobre a liberdade de expressão.<sup>23</sup>

<sup>18</sup> CDA 230: Legislative History. Electronic Frontier Foundation, disponível em: <<https://www.eff.org/issues/cda230/legislative-history>>. Último acesso em 31 de janeiro de 2021.

<sup>19</sup> FLICK, Dallas. Combatting Fake News: Alternatives To Limiting Social Media Disinformation And Rehabilitating Quality Journalism. **Science and Technology Law Review**, [s.l.], vol. 20, n. 2, p. 375-405. 2017. p. 7.

<sup>20</sup> Section 230(c): “No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider.” Disponível em: <<https://www.law.cornell.edu/uscode/text/47/230>>. Último acesso em 31 de janeiro de 2021.

<sup>21</sup> WILSON, Richard A.; LAND, Molly K. Hate speech on Social Media: Towards a Context-Specific Content Moderation Policy. **Connecticut Law Review**, Connecticut, vol. 52. 2020. p. 26.

<sup>22</sup> Termo utilizado de demonstrar o grau de controle exercido e a estrutura das normas em comento, conforme David Kaye, relator especial da ONU. KAYE, David. *Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression*, **United Nations Human Rights Council**, 38<sup>th</sup> Session, 6 April 2018. p. 3.

<sup>23</sup> Nesse sentido, vide *Manhattan Cmty. Access Corp. v. Halleck* (2019), onde a Suprema Corte afirmou cabalmente que hospedar diálogo não é uma função exclusivamente pública, e não faz com que entidades





Tal entendimento, a prima facie, vai de encontro com decisões judiciais que entenderam que partes dessas plataformas podem ser consideradas fóruns públicos,<sup>24</sup> determinando a aplicação da 1ª Emenda Constitucional à plataforma em comento. Interpretados conjuntamente, percebe-se que o único parâmetro para sua aplicabilidade passa a ser a qualidade pública ou não do espaço, definida com base na titularidade da página, *independentemente dos efeitos do ato atentatório ou regulatório sobre o bem jurídico protegido* (qual seja, a liberdade política e de expressão).

Assim sendo, um presidente não pode bloquear usuários de sua página em redes sociais, para não impedir o exercício de seus direitos políticos, uma vez que sua página constituiria um fórum público – mas a mesma rede social pode bloquear um presidente e forçá-lo a apagar publicações por violar as normas de conteúdo da plataforma,<sup>25</sup> independentemente das repercussões políticas, ou até mesmo deletar sua página. Em todos os casos haveria o impedimento do diálogo direto entre os cidadãos e um dos mais influentes oficiais do governo, que seria o fundamento para a não permissão de bloqueio pelo presidente; contudo, ao ente particular é permitida essa conduta, e àquele tido como público, não.

Percebe-se que nos EUA prevalece o modelo autorregulatório no que diz respeito ao discurso de ódio e à desinformação, tanto nas redes como fora. Não há definição legal desses conceitos (embora haja resoluções que propõe definições), não há obrigações legais para as plataformas de monitorar ou moderar os conteúdos hospedados e não há possibilidade de controle estatal desses conteúdos – nem mesmo indiretamente, através de legislação ou remédios equivalentes. Tampouco há perspectiva de mudança dessa situação, no estado atual da legislação e da jurisprudência.

---

privadas sejam transformadas em “atores estatais” sujeitos às limitações do *First Amendment*. UNITED STATES SUPREME COURT. **Manhattan Community Access Corp. v. Halleck** [syllabus]. 2019. Disponível em: <<https://www.law.cornell.edu/supremecourt/text/17-1702>>. Último acesso em 12 fev. 2021. Vide também Prager Univ. v. Google LLC (2020), onde o 9º circuito estabeleceu que o Youtube é um fórum privado, e não público, e que, portanto, seu regulamento não está sujeito ao *First Amendment*. UNITED STATES COURT OF APPEALS, Ninth Circuit. **Prager University v. Google LLC, FKA Google Inc; Youtube, LLC** [opinion]. 2019. Disponível em: <<https://cdn.ca9.uscourts.gov/datastore/opinions/2020/02/26/18-15712.pdf>>.

<sup>24</sup> Vide **Knight First Amendment Institute at Columbia University v. Donald J. Trump et al.** (2018), que reconheceu a página de Twitter de Donald Trump como fórum público, cujo acesso não poderia ser restringido através do bloqueio de usuários. UNITED STATES SUPREME COURT. **Knight First Amendment Institute at Columbia University v. Donald J. Trump et. Al.** [syllabus]. 2018. Disponível em: <<https://www.law.cornell.edu/rio/citation/928F.3d226>>. Último acesso em 12 fev. 2021. Vide também Brian Davison v. Phyllis Randall et al (2019), no mesmo sentido quanto à página oficial de um oficial do governo. UNITED STATES COURT OF APPEALS, Fourth Circuit. **Brian Davison v. Phyllis Randallet. Al.** [syllabus]. 2019. Disponível em: <<https://www.ca4.uscourts.gov/opinions/172002.P.pdf>>. Último acesso em 12 fev. 2021.

<sup>25</sup> PAUL, Kari. Twitter and Facebook lock Donald Trump's accounts after video address. The Guardian. 7 Jan. 2021. Disponível em: <https://www.theguardian.com/us-news/2021/jan/06/facebook-twitter-youtube-trump-video-supporters-capitol>. Último acesso em 12 fev. 2021.



### 3. A REGULAÇÃO NA ALEMANHA

Na Alemanha, por sua vez, o Estado intervém de maneira mais pronunciada na regulação dos conteúdos das redes. Enquanto nos EUA só há legislação vedando a difamação, na Alemanha há também dispositivos legais acerca do discurso de ódio e da disseminação deliberada de desinformação,<sup>26</sup> que também são expressamente combatidos. Deve-se manter em mente que não há a preponderância da liberdade de expressão sobre os demais direitos fundamentais na Alemanha; de fato, pela leitura da constituição alemã, pode-se concluir que a dignidade da pessoa humana está em posição superior, ou ao menos equivalente, a essa liberdade.<sup>27</sup> Ademais, o mesmo artigo que trata liberdade de expressão afirma que toda pessoa poderá se informar sem obstáculos de fontes publicamente acessíveis, expressamente consignando uma preocupação com o direito ao acesso à informação (que embasa o combate à desinformação).

Assim sendo, uma vez que os mesmos direitos protegidos pelo Estado em ambientes físicos podem ser ameaçados nos virtuais, o governo Alemão exerce (e exige que as plataformas em questão exerçam) um certo grau de controle sobre os conteúdos postados em redes sociais.

No que diz respeito ao discurso de ódio (*Volkshverhetzung*), há uma definição legal positivada<sup>28</sup> que pode se sobrepor às definições estabelecidas por cada plataforma, embora as definições das plataformas por vezes sejam ainda mais amplas.<sup>29</sup> Antes da promulgação da *Netzwerkdurchsetzungsgesetz* (“Lei de Fiscalização de Redes” em tradução livre, doravante *NetzDG*), provedores de serviços de telecomunicação não seriam responsáveis pelo conteúdo publicado por terceiros em suas plataformas contanto que não tivessem ciência de seu conteúdo; uma vez notificados a respeito, todavia, teriam

<sup>26</sup> HARFF, Graziela; DUQUE, Marcelo Schenk. Discurso de ódio nos contextos alemão e brasileiro. **A&C – Revista de Direito Administrativo & Constitucional**, Belo Horizonte, ano 21, n. 84, p. 199-225, abr./jun. 2021.

<sup>27</sup> A constituição alemã trata da dignidade humana em seu primeiro artigo, afirmando ser inviolável e que respeitá-la e protegê-la será o dever de toda autoridade estatal. A liberdade de expressão, por sua vez, é tratada no quinto artigo, onde se afirma que toda pessoa terá o direito de se expressar livremente e disseminar suas opiniões oralmente e por escrito, e que não haverá censura. Todavia, o segundo inciso do referido artigo informa que o direito à liberdade de expressão será limitado pelas provisões de leis gerais, de dispositivos de proteção de jovens e no direito à honra pessoal. (Conforme a tradução em inglês da constituição, disponível em: [https://www.gesetze-im-internet.de/englisch\\_gg/englisch\\_gg.html#p0019](https://www.gesetze-im-internet.de/englisch_gg/englisch_gg.html#p0019), consultado em 24/06/2021). A despeito disso, reconhece-se a importância da liberdade de expressão, com a Corte Constitucional Alemã chegando a afirmar que essa liberdade “é de certa maneira a base de qualquer tipo de liberdade, a matriz, a indispensável condição de quase toda outra forma de liberdade”, no julgamento “7 BVerfGE, 198, 208 (1958)”. SARMENTO, Daniel. A liberdade de expressão e o problema do “hate speech”. **Revista de Direito do Estado**, Rio de Janeiro, ano 1, n. 4, p. 1-58, out./dez. 2006. p. 21.

<sup>28</sup> Vide seção 130 do Código Penal Alemão, particularmente os itens 1.1 e 1.2. Código disponível em inglês neste link: [https://www.gesetze-im-internet.de/englisch\\_stgb/englisch\\_stgb.html](https://www.gesetze-im-internet.de/englisch_stgb/englisch_stgb.html). Último acesso em 31/01/2021.

<sup>29</sup> Como no caso do Youtube, que prevê a remoção exclusivamente com base no conteúdo, enquanto a legislação alemã dispõe o requisito adicional que a mensagem de discurso de ódio seja veiculada “de maneira adequada a causar um distúrbio da paz pública” (Seção 130.1 do Código Penal Alemão), ou seja, que tenha o potencial de causar dano real.



de removê-lo imediatamente para preservar sua imunidade.<sup>30</sup> Contudo, o governo alemão julgou que o cumprimento da norma era de difícil fiscalização e que seus resultados eram insuficientes,<sup>31</sup> pelo que a NetzDG foi promulgada em 2017, impulsionada pela influência de campanhas de desinformação durante as eleições estado-unidenses de 2016.

Essa lei se aplica exclusivamente a provedores de serviço de redes sociais<sup>32</sup> com 2 milhões ou mais de usuários registrados na Alemanha, e não a plataformas que publicam conteúdo jornalístico original, nem a serviços de e-mail ou de envio de mensagens.<sup>33</sup> Aos serviços abrangidos, impõe-se a obrigação de remover conteúdo manifestamente ilegal dentro de 24 horas do recebimento de uma reclamação de qualquer usuário. Ausente a manifesta ilegalidade, a plataforma tem o prazo de 7 dias para investigar e deletar o conteúdo, com a possibilidade de extensão do período caso mais fatos sejam necessários à apuração da veracidade do conteúdo publicado, ou no caso de contratação de um terceiro para a realização desse procedimento. As referidas plataformas devem obrigatoriamente oferecer um mecanismo fácil e transparente de reclamação, e a decisão final da empresa deve ser fundamentada e comunicada prontamente ao usuário denunciante.

Ademais, plataformas que recebem mais de 100 denúncias anuais de conteúdo ilegal são obrigadas a publicar relatórios bianuais em alemão acerca de como lidam com as denúncias. Os relatórios devem estar disponíveis na página de acesso do serviço e no Diário Federal (*Bundesgesetzblatt*) no máximo um mês após o término do período de meio ano, e deverão ser de livre e imediato acesso, assim como permanentemente disponíveis. O descumprimento intencional ou culposos de qualquer das disposições

<sup>30</sup> GESLEY, Jenny. Initiatives to Counter Fake News in Selected Countries. **Law Library of Congress**, [s.l.], p. 33-40, abr. 2019. p. 2.

<sup>31</sup> Wissenschaftliche Dienste [Parliamentary Research Services], Der Umgang mit Fake-News. Rechtslage und Reformansätze [Dealing with Fake News. Current Legal Situation and Reform Proposals], Report No. WD 10 - 3000 - 067/16 (Dec. 20, 2016), <https://www.bundestag.de/blob/494418/4321d229204080dce488ebd0356b5db2/wd-10-067-16-pdf-data.pdf>. Archived at: <<http://perma.cc/ZD2V-SSHP>>. *apud* GESLEY, Jenny. Initiatives to Counter Fake News in Selected Countries. **Law Library of Congress**, [s.l.], p. 33-40, abr. 2019. p. 2.

<sup>32</sup> Definidos como “provedores de serviços telemáticos que operam plataformas online com o intuito de gerar lucro, e nas quais usuários podem compartilhar conteúdo com outros usuários ou tornar esse conteúdo publicamente disponível,” em tradução livre do item 1.1 da NetzDG.

<sup>33</sup> Tal definição é distinta daquela de fóruns cibernéticos defendida neste trabalho, principalmente por distinguir entre plataformas em sua integralidade, e não explicitamente uma de suas funcionalidades; chats de grupo do WhatsApp, por exemplo, estariam excluídos por meio dessa definição, apesar de serem um veículo notável de desinformação e assemelham-se a murais do Facebook em funcionalidade. Ainda, tal definição incluiria redes sociais com um enfoque em conteúdo puramente artístico (e apolítico), e não necessariamente de discussão ou compartilhamento de informações e opiniões, que estão mais ligados ao exercício de direitos políticos, merecendo uma proteção diferenciada. Em suma, a definição pela NetzDG possui um escopo diferente, não considera a multiplicidade de espaços em uma mesma plataforma que devem ter regimes jurídicos distintos, e utiliza como critério a possibilidade de compartilhar conteúdo publicamente, ao invés de considerar a finalidade da ferramenta em comento. Por todos esses motivos, entende-se que o termo fórum cibernético é mais apropriado, ao menos para os fins deste estudo.



dessa lei estará sujeito a duras sanções patrimoniais, de até 50 milhões de euros, condicionadas a uma decisão judicial.

O Estado alemão, portanto, também optou por deixar as obrigações de regulação dos conteúdos das plataformas às próprias plataformas, em conjunto com seus usuários, em uma espécie de “autorregulação regulada”<sup>34</sup>. Há mais normas regulamentares acerca de como se dará esse controle, diferentemente do que ocorre nos EUA, com a definição de um procedimento a ser seguido e definições legais positivadas de conteúdos que devem obrigatoriamente ser removidos, mas o dever de operacionalizar o referido controle continua nas mãos das plataformas. É um meio termo entre as chamadas *Old-School Speech Regulation* e *New-School Speech Regulation*,<sup>35</sup> em que o controle é exercido pelo Estado ou por entes particulares, respectivamente; uma espécie de regime misto, em que as plataformas são obrigadas a servir o Estado no cumprimento da lei.

Prevalece a ideia de uma democracia militante, definida por Daniel Sarmiento como “a noção de que o Estado deve defender a democracia dos seus ‘inimigos’, que não aceitam as regras do jogo democrático e pretendem subvertê-las.”<sup>36</sup> Da mesma forma que o passado americano influenciou o culto à liberdade de expressão irredutível, não há dúvidas que a construção da democracia alemã, particularmente após a 2ª Guerra Mundial, influencia a sua tradição jurídica, resultando no combate ativo às manifestações de ódio. A constituição alemã é expressa no sentido de que o desrespeito aos princípios democráticos ou o comprometimento da democracia podem justificar medidas extremas, como a declaração de inconstitucionalidade de partidos políticos (art. 21, II) ou na privação dos direitos fundamentais utilizados abusivamente (art. 18), como já ocorreu no passado.<sup>37</sup>

#### 4. A REGULAÇÃO NO BRASIL

No Brasil, como na Alemanha, o Estado intervém na esfera privada para a garantia dos direitos fundamentais, dentre os quais a dignidade da pessoa humana, a integridade psíquica e a igualdade. Embora tanto o discurso de ódio como a difusão de

<sup>34</sup> CUEVA, Ricardo Villas Bôas. Alternativas para a remoção de fake news das redes sociais. **Revista de Direito e as Novas Tecnologias**, [s.l.], vol. 1/2018, p. 1-10, out./dez. 2018. p. 1.

<sup>35</sup> BALKIN, Jack M. Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation. **UC Davis Law Review**, vol. 51/2018, p. 1149-1210. 2018. Disponível em: [https://digitalcommons.law.yale.edu/fss\\_papers/5160/](https://digitalcommons.law.yale.edu/fss_papers/5160/). Último acesso em 31/01/2021.

<sup>36</sup> SARMENTO, Daniel. A liberdade de expressão e o problema do “hate speech”. **Revista de Direito do Estado**, Rio de Janeiro, ano 1, n. 4, p. 1-58, out./dez. 2006. p. 21.

<sup>37</sup> Vide os casos dos partidos SRP (Partido Socialista do Reich), por suas tendências neonazistas, e a do KDP (Partido Comunista Alemão), na Guerra Fria, em 1952 e 1956, respectivamente. Julgamentos 2 BVerfGE 1 (1952) e 5 BVerfGE 85 (1956). SARMENTO, Daniel. A liberdade de expressão e o problema do “hate speech”. **Revista de Direito do Estado**, Rio de Janeiro, ano 1, n. 4, p. 1-58, out./dez. 2006.



desinformação sejam combatidas no Brasil, não são objetos de leis específicas, pelo que a subsunção da conduta à norma é um obstáculo em si.

O discurso de ódio é tutelado sob os crimes contra a honra tipificados no Código Penal ou aqueles previstos na Lei para Crimes de Racismo (Lei 7.716/89)<sup>38</sup>, recordando-se que atos discriminatórios que não tenham como base a raça da vítima ainda podem ser enquadrados (como no caso de homofobia). Tal tutela é justificável não somente pelo direito à dignidade, à integridade (física e psíquica) e à igualdade das vítimas, como pelo objetivo expresso da Constituição Federal de combater a discriminação e a desigualdade.<sup>39</sup>

Há casos emblemáticos que trataram da colisão principiológica da dignidade humana e a liberdade de expressão em matéria de discurso de ódio, como no célebre Caso Ellwanger (o HC 82.424, julgado em 2004). O caso tratava da condenação de um escritor e editor antisemita, julgado pelo crime de racismo em razão da publicação, venda e distribuição de material difusor dessa “ideologia”. A condenação foi mantida, restando expresso na ementa do acórdão que a liberdade de expressão é “garantia constitucional que não se tem como absoluta. Limites morais e jurídicos. [...] As liberdades públicas não são incondicionais, por isso devem ser exercidas de maneira harmônica, observados os limites definidos na própria Constituição Federal [...]. Prevalência dos princípios da dignidade da pessoa humana e da igualdade jurídica.” A Corte expressamente afirmou que não há proteção constitucional do discurso de ódio, prevalecendo os direitos à igualdade e à dignidade.<sup>40</sup>

A tutela da desinformação, por sua vez, se dá através da aplicação de múltiplas normas, eleitorais ou não, dentre as quais enumera o Ministro Toffoli: a Lei 12.891/2013 (criminalização de contratação de grupos para emitir mensagens ou comentários na internet para denigrir ou ofender candidato, partido ou coligação); a Lei 13.488/2017 (penalização de usuários que se cadastram em aplicação de internet com a intenção de falsar identidade [perfis falsos], permitindo a responsabilização do beneficiário e do responsável pela veiculação); a Resolução n. 23.551/2017 (acerca da divulgação de fatos sabidamente inverídicos); a Lei 12.965/14, denominada o Marco Civil da Internet (notavelmente em seu art. 19º, que permite, mediante ordem judicial, a indisponibilização

<sup>38</sup> SANTOS, Elaine Gomes dos; RIBEIRO, Raisa Duarte da Silva. Restrições à liberdade de expressão e crimes cibernéticos: a tutela penal do discurso de ódio nas redes sociais. **Revista dos Tribunais**, [s.l.], vol. 997/2018, p. 515-514, nov. 2018. p. 5.

<sup>39</sup> “Art. 3º Constituem objetivos fundamentais da República Federativa do Brasil: I - construir uma sociedade livre, justa e solidária; [...] III - erradicar a pobreza e a marginalização e reduzir as desigualdades sociais e regionais; IV - promover o bem de todos, sem preconceitos de origem, raça, sexo, cor, idade e quaisquer outras formas de discriminação.” Sem grifos no original.

<sup>40</sup> COSTA, Rodrigues de Souza; RIBEIRO, Raisa Duarte da Silva. Discurso de ódio no Brasil: uma análise do HC 82.424/RS e seus reflexos. **Direitos e Garantias Fundamentais II - XXV Congresso Nacional do Conpedi**, 2015, Florianópolis, p. 138-158. Disponível em: <<http://site.conpedi.org.br/publicacoes/02q8agmu/j8667276/r177M5DQXgWii38p.pdf>>. Último acesso em 25/06/2021.



pelo provedor de aplicações de internet do conteúdo danoso gerado por terceiro, permitindo-se a tutela antecipada).<sup>41</sup>

Assim sendo, a regra geral aplicada no Brasil é a ausência de responsabilidade das plataformas quanto aos conteúdos ali postados, até que haja determinação judicial expressa para a remoção. Contudo, vale ressaltar que em 2017 o STJ ampliou esse entendimento no REsp 1.342.640/SP. Afirmou-se que embora as plataformas não possuam o dever de exercer um controle prévio (ao jurisdicional) do conteúdo postado em seus sites pelos usuários, devem removê-los imediatamente assim que tomarem conhecimento inequívoco das publicações. Ademais, para permitir a responsabilização de usuários que postem conteúdos ilegais, a plataforma deve manter um sistema de identificação de usuários.<sup>42</sup>

Como se vê, não há legislação dedicada à tutela do discurso de ódio ou da desinformação, embora haja projetos de lei em votação. A situação é particularmente preocupante no que diz respeito à regulação da desinformação,<sup>43</sup> uma vez que o discurso de ódio pode muitas vezes ser enquadrado nas hipóteses dos crimes por raça ou dos crimes contra a honra, enquanto condutas relacionadas à criação e disseminação (deliberada) de desinformação não se amoldam com tanta facilidade aos tipos legalmente previstos. A própria subsunção da conduta à norma representa um obstáculo significativo no combate a essa tática – o que é particularmente problemático diante da necessidade de controle judicial para a remoção de publicações.

## 5. O MODELO DE AUTORREGULAÇÃO

A autorregulação, tanto na modalidade mais pura presente nos EUA como a “autorregulação regulada” da Alemanha, apresenta benefícios e malefícios com relação ao modelo da regulação estatal vigente no Brasil.

Em um primeiro momento, cabe ressaltar que a desinformação e o discurso de ódio (assim como discursos inflamatórios e polarizantes em geral) não prosperam nos fóruns por mera coincidência. Conforme diversos estudos, esse tipo de publicação

---

<sup>41</sup> TOFFOLI, José Antonio Dias. Fake News, Desinformação e Liberdade de Expressão. **Interesse Nacional**, [s.l.], ano 12, n. 46, p. 1-15, jul./set. 2019. Disponível em: <<http://interessenacional.com.br/2019/07/11/fake-news-desinformacao-e-liberdade-de-expressao/>>. Último acesso em 25/06/2021.

<sup>42</sup> Confira-se breve trecho da ementa do acórdão: “Esta Corte fixou entendimento de que “(i) não respondem objetivamente pela inserção no site, por terceiros, de informações ilegais; (ii) não podem ser obrigados a exercer um controle prévio do conteúdo das informações postadas no site por seus usuários; (iii) devem, assim que tiverem conhecimento inequívoco da existência de dados ilegais no site, removê-los imediatamente, sob pena de responderem pelos danos respectivos; (iv) devem manter um sistema minimamente eficaz de identificação de seus usuários, cuja efetividade será avaliada caso a caso”. Precedentes.” REsp 1.342.640/SP, julgado em 14/02/07. Relatoria da Ministra Nancy Andrighi.

<sup>43</sup> FORNASIER, Mateus de Oliveira; BORGES, Gustavo Silveira. The current influence of social media on democratic debate, political parties and electioneering. **Revista de Investigações Constitucionais**, Curitiba, vol. 9, n. 1, p. 73-102, jan./abr. 2022.



consistentemente resulta em um engajamento muito maior pelos usuários do que outros tipos de publicações, gerando comentários e compartilhamentos.<sup>44</sup> Ora, uma vez que as redes sociais monetizam o engajamento,<sup>45</sup> o combate à desinformação e o discurso de ódio vão de encontro com o seu modelo de negócios, gerando nítido conflito de interesses.

Outro problema é a questão da legitimidade para regular.<sup>46</sup> O Estado possui legitimidade para agir enquanto guardião da liberdade de expressão, da dignidade da pessoa humana e da saúde coletiva, uma vez que deve zelar pelos interesses coletivos dos cidadãos. Mas que legitimidade possui uma empresa de tecnologia para agir dessa mesma maneira – e mais, para excluir essa regulação da competência estatal?

A aplicação do regime americano significa que o Facebook poderia remover publicações contendo informações falsas sobre a vacina contra a Covid-19, criadas e compartilhadas para semear pânico e desincentivar a vacinação (embora não tenha qualquer obrigação de fazê-lo, haja vista a exclusão de sua responsabilidade pelo CDA). Contudo, o Estado não poderia fazer nada, não lhe sendo legítimo intervir na esfera privada ou censurar a liberdade de expressão dos semeadores de pânico e desinformação. Uma plataforma poderia censurar discursos contra a vacinação que prejudicariam a saúde pública injustificadamente – caracterizando nítida ofensa a um bem coletivo – mas o próprio Estado, guardião da saúde pública, não poderia fazê-lo. A totalidade do poder regulatório e fiscalizatório é concentrado nas mãos de entes particulares com

---

<sup>44</sup> Segundo o estudo *"The science of fake news"*, de março de 2018, "o poder de proliferação das fake News é três vezes maior que a de outros conteúdos", mesmo controlando a presença de robôs (bots). LAZER, David et al. *The science of fake news*. **Science**, [s.l.], vol. 359, n. 6380, p. 1094-1096. 2018. *apud* CARNEIRO, Gustavo Ferraz Sales. Autorregulação de fake news no Facebook: incentivos e freios à proliferação de desinformação. **Revista de Direito e as Novas Tecnologias**, [s.l.], vol. 3/2019, p. 1-11, abr./jun. 2019. p. 3. O estudo *"The spread of true and false news online"*, por sua vez, vai além, afirmando que falsidades possuem um alcance, velocidade e profundidade de difusão muito maior do que o apresentado por notícias verdadeiras. Em termos de velocidade, o estudo aponta que a verdade demorava 6 vezes mais tempo para alcançar 1.500 pessoas que notícias falsas, e 20 vezes mais para alcançar a profundidade de 10. As notícias falsas de cunho político possuem potencial de difusão ainda maior, alcançando 20.000 pessoas quase 3 vezes mais rapidamente que outros tipos de falsidades levaram para alcançar 10.000 pessoas. VOSOUGH, Sorous; ROY, Deb; ARAL, Sinan. *The spread of true and false news online*. **Science**, [s.l.], vol. 359, n. 6380, p. 1146-1151. 2018.

<sup>45</sup> Este é o fenômeno conhecido como "economia de atenção", no qual cada segundo de atenção do usuário é um recurso econômico a ser explorado. Vide o documentário "O Dilema das Redes", bem como o artigo "Autorregulação de fake news no Facebook: incentivos e freios à proliferação de desinformação". O dilema das redes. Direção de Jeff Orlowski. Netflix: 2020. 94 min. CARNEIRO, Gustavo Ferraz Sales. Autorregulação de fake news no Facebook: incentivos e freios à proliferação de desinformação. **Revista de Direito e as Novas Tecnologias**, vol. 3/2019, 2019.

<sup>46</sup> Ricardo Villas Bôas Cueva afirma que uma das críticas dirigidas à NetzDG seria justamente essa da legitimidade. Afirma-se que "a nova lei seria uma espécie de terceirização ou delegação inconstitucional de atividade típica do Estado a empresas. O Estado, que detém o monopólio da repressão e punição de comportamentos ilícitos, não poderia atribuir tais atividades ao setor privado." CUEVA, Ricardo Villas Bôas. Alternativas para a remoção de fake news das redes sociais. **Revista de Direito e as Novas Tecnologias**, [s.l.], vol. 1/2018, p. 1-10, out./dez. 2018. p. 6.



finalidades econômicas específicas, que muitas vezes vão de encontro aos interesses coletivos. Seria mesmo um cenário desejável?

Esse controle por particulares, ademais, não estaria sujeito a princípios fundamentais para a atuação a Administração Pública, como os princípios da transparência, da publicidade e da fundamentação da decisão pública. Isso é exacerbado pelo método de controle algorítmico empregado, que, apesar de célere e eficiente, apresenta problemas em termos de fiscalização e transparência.<sup>47</sup> A regulação por meio de algoritmos deve ser criteriosa e muito bem regulada,<sup>48</sup> sob pena da inteligência artificial aparentemente neutra ser utilizada para reforçar preconceitos e vícios sociais pré-existentes.<sup>49</sup>

Ainda, a autorregulação das redes sociais se apoia na participação popular, através de denúncias por usuários, o que pode ser a origem de mais complicações. Uma das críticas dirigidas ao modelo é que “a notificação de supostos ilícitos por leigos suscita inevitáveis dificuldades de enquadramento jurídico, o que pode levar a uma aparente ineficácia do sistema de autorregulação, em decorrência de uma sobrecarga de reclamações que podem ser arquivadas, à míngua de concretude.”<sup>50</sup>

O modelo de “autorregulação regulada” alemão combate algumas dessas questões, impondo um grau maior de transparência na tomada de decisões através das disposições da NetzDG, principalmente à luz dos relatórios obrigatórios. Ademais, redes sociais como o Twitter hoje contam com uma rede de usuários e organizações considerados “parceiros confiáveis”<sup>51</sup>, o que é um diferencial e tanto – se houver o respeito criterioso a organizações especializadas confiáveis, como a OMS em matéria de saúde, atenuam-se as preocupações de legitimidade, expertise e conflitos de interesses.

---

<sup>47</sup> CUEVA, Ricardo Villas Bôas. Alternativas para a remoção de fake news das redes sociais. **Revista de Direito e as Novas Tecnologias**, [s.l.], vol. 1/2018, p. 1-10, out./dez. 2018. p. 6. Juarez Freitas também menciona esse problema, pelo que afirma que “é instada a ensejar, ao máximo do tecnicamente viável, o acesso à sequência dos passos lógicos da decisão algorítmica” para a concretização dos princípios da transparência e do direito à explicação, positivado no art. 20 da Lei Brasileira de Proteção de Dados (Lei 13.709/2020). FREITAS, Juarez. Direito Administrativo e inteligência artificial. **Interesse Público** – IP, Belo Horizonte, ano 21, n. 114, p. 15-29, mar./abr. 2019. p.11.

<sup>48</sup> PUSCHEL, André Felipe Silva; RODRIGUES, Roberto Tessis; VALLE, Vivian Cristina Lima López. O dilema ético da decisão algorítmica na administração pública. **A&C – Revista de Direito Administrativo & Constitucional**, Belo Horizonte, ano 22, n. 90, p. 207-226, out./dez. 2022; ARAÚJO, Valter Shuenquener de; ZULLO, Bruno Almeida; TORRES, Maurílio. Big Data, algoritmos e inteligência artificial na Administração Pública: reflexões para a sua utilização em um ambiente democrático. **A&C – Revista de Direito Administrativo & Constitucional**, Belo Horizonte, ano 20, n. 80, p. 241-261, abr./jun. 2020.

<sup>49</sup> A esse respeito, vide: VINUESA, Ricardo et al. The role of artificial intelligence in achieving the Sustainable Development Goals. **Nature Communications**, [s.l.], vol 11, n. 1, artigo 233, p. 1-10. 2020. Disponível em: <<http://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-168566>>. Último acesso em 27/06/2021.

<sup>50</sup> PEIFER, Karl-Nikolaus. Netzwerkdurchsetzungsgesetz: Selbstbehauptung des Rechts oder erster Schritt in die selbstregulierte Vorzensur? – **Zivilrechtliche Aspekte**. *apud* CUEVA, Ricardo Villas Bôas. Alternativas para a remoção de fake news das redes sociais. **Revista de Direito e as Novas Tecnologias**, [s.l.], vol. 1/2018, p. 1-10, out./dez. 2018. p. 8-9.

<sup>51</sup> WILSON, Richard A.; LAND, Molly K. Hate speech on Social Media: Towards a Context-Specific Content Moderation Policy. **Connecticut Law Review**, Connecticut, vol. 52. 2020. p. 42.





Contudo, mesmo com esses mecanismos, os pontos problemáticos permanecem presentes, pelo que regimes de regulação alternativos permanecem atrativos.

Dito isso, é inegável que também há vantagens inerentes a esse modelo. A autorregulação possui maior agilidade na identificação, bloqueio ou supressão de conteúdos ilícitos, particularmente em contraste com o modelo de controle judicial.<sup>52</sup> No mundo virtual, a remoção tardia de uma publicação resulta no potencial de compartilhamento e armazenamento por inúmeros usuários, frustrando o resultado útil da medida. Ademais, com a delegação dessa atividade às empresas, reduz-se a carga (econômica, técnica e humana) do Estado, sem mencionar a possível geração de empregos.

Ademais, uma vez que essas empresas estão inseridas na sociedade, as normas estabelecidas pelas plataformas estão sujeitas a valores sociais e pressão popular. Por um lado, isso pode resultar em medidas discriminatórias, como em regimes populistas e fascistas – por outro, há a possibilidade de que resulte em normas consensualmente formadas, conferindo-lhes legitimidade e atendendo os anseios da população. A elaboração e discussão dessas normas pode conduzir à evolução social, como qualquer outra instância de exercício pleno da democracia.

É interessante notar que, em razão da ampla autonomia “legislativa” das plataformas e de valores sociais amplamente compartilhados, os regimes de “platform law” aplicáveis ao redor do mundo aproximam-se ao modelo regulatório implementado na Europa. Isso ocorre tão marcadamente que David Kaye afirma que é possível que, com o passar do tempo, não haja muita diferença entre a regulação material Europeia e a Estado-Unidense,<sup>53</sup> apesar de tais normas serem de origem completamente particular nos EUA. É possível que, com o tempo, houvesse a criação orgânica de uma espécie de Direito Internacional das Redes, avançando a integração global da humanidade.

## 6. O MODELO DE REGULAÇÃO ESTATAL (JUDICIAL)

A regulação através da técnica da reserva de jurisdição, por sua vez, apresenta vantagens e desvantagens distintas. É o modelo atualmente utilizado no Brasil para combater a desinformação e o discurso de ódio nas redes sociais.

Ao invés de ter como base a teoria do Estado absentista – amplamente aceita nos EUA – segundo a qual o Estado deve limitar ao máximo sua intervenção na esfera particular a fim de respeitar a liberdade dos cidadãos, permitindo que soluções justas emergam do livre debate, o modelo brasileiro tem como base a teoria do Estado garantidor. Segundo essa concepção, não basta a igualdade formal dos participantes de um debate, por exemplo, para que a verdade ou a solução mais justa advenham do “mercado

<sup>52</sup> CUEVA, Ricardo Villas Bôas. Alternativas para a remoção de fake news das redes sociais. **Revista de Direito e as Novas Tecnologias**, [s.l.], vol. 1/2018, p. 1-10, out./dez. 2018. p. 8.

<sup>53</sup> Conforme o discurso de David Kaye no Conselho de Carnegie para Ética em Assuntos Internacionais, 6 de Julho de 2019, disponível em: <https://www.youtube.com/watch?v=W6PDZ-o5Khg> (39:34).



livre de ideias”<sup>54</sup>; o Estado deve agir para garantir os direitos dos cidadãos, ao invés de se omitir. A despeito da importância do livre debate e da liberdade de expressão e opinião, a busca pela verdade através do mercado livre de ideias só terá êxito em um ambiente de livre competição entre as ideias – ou seja, um ambiente em que haja *igualdade de material* entre os participantes e seus discursos.<sup>55</sup> Essa teoria justifica a intervenção estatal nos discursos entre particulares, em especial no que diz respeito ao discurso de ódio (que não é uma real tentativa de engajar em debate, e sim uma prática social que visa promover a violência e intolerância contra determinados indivíduos ou grupos).<sup>56</sup>

A mais significativa desvantagem desse modelo relativo ao anterior é a celeridade. Conforme aduzido acima, a remoção tardia de um conteúdo pode resultar na falta de eficácia da medida, uma vez que o conteúdo já terá sido armazenado e compartilhado repetidamente. Esse trâmite mais longo também pode contribuir para a sobrecarga do Judiciário Brasileiro, o que representa custos adicionais com recursos (humanos, financeiros e tecnológicos) que serão pagos pela população em geral. A viabilidade desse controle em larga escala também é limitada, haja vista o tempo e a expertise necessárias.

Contudo, o mesmo procedimento que diminui a celeridade possui benefícios importantes. Dentre estes, estão o exercício do amplo contraditório, bem como a estabilidade da decisão judicial.<sup>57</sup> Ademais, uma vez que a decisão judicial é pautada nos princípios de atuação da Administração Pública, a transparência e a publicidade, como regra, são plenamente aplicáveis.

Embora uma preocupação comum com a regulação estatal seja o surgimento de uma regulação preventiva excessiva por parte das redes sociais em razão da possibilidade de responsabilização (o chamado “chilling effect”),<sup>58</sup> o modelo utilizado no Brasil

<sup>54</sup> MILL, John Stuart. *On Liberty*. **Ticknor and Fields**, Boston, 1863. *apud* SANTOS, Elaine Gomes dos; RIBEIRO; Raísa Duarte da Silva. Restrições à liberdade de expressão e crimes cibernéticos: a tutela penal do discurso de ódio nas redes sociais. **Revista dos Tribunais**, [s.l.], vol. 997/2018, p. 515-514, nov. 2018. p. 4.

<sup>55</sup> SANTOS, Elaine Gomes dos; RIBEIRO; Raísa Duarte da Silva. Restrições à liberdade de expressão e crimes cibernéticos: a tutela penal do discurso de ódio nas redes sociais. **Revista dos Tribunais**, [s.l.], vol. 997/2018, p. 515-514, nov. 2018. No mesmo sentido, SARMENTO, Daniel. A liberdade de expressão e o problema do “hate speech”. **Revista de Direito do Estado**, Rio de Janeiro, ano 1, n. 4, p. 1-58, out./dez. 2006. p. 31.

<sup>56</sup> MOURA, Marco Aurélio. **O discurso do ódio em redes sociais**. São Paulo: Lura Editorial, 2016. p. 44. *apud* SANTOS, Elaine Gomes dos; RIBEIRO; Raísa Duarte da Silva. Restrições à liberdade de expressão e crimes cibernéticos: a tutela penal do discurso de ódio nas redes sociais. **Revista dos Tribunais**, [s.l.], vol. 997/2018, p. 515-514, nov. 2018. p. 9.

<sup>57</sup> CUEVA, Ricardo Villas Bôas. Alternativas para a remoção de fake news das redes sociais. **Revista de Direito e as Novas Tecnologias**, [s.l.], vol. 1/2018, p. 1-10, out./dez. 2018. p. 8.

<sup>58</sup> Foi a principal preocupação que levou à criação da excludente geral de responsabilidade aos provedores de serviços virtuais no Communications Decency Act nos EUA. Em igual sentido, na Alemanha, acerca das penalidades e obrigações criadas pela NetzDG: MÜLLER-FRANKEN, Sebastian. *Netzwerkdurchsetzungsgesetz: Selbstbehauptung des Rechts oder erster Schritt in die selbstregulierte Vorzensur? – Verfassungsrechtliche Fragen*, 2018. *apud* CUEVA, Ricardo Villas Bôas. Alternativas para a remoção de fake news das redes sociais. **Revista de Direito e as Novas Tecnologias**, [s.l.], vol. 1/2018, p. 1-10, out./dez. 2018.



evita esse problema. Segundo o Marco Civil da Internet, a remoção por parte das redes sociais só precisa ocorrer *após a determinação judicial*, momento em que a plataforma passa a ser passível de responsabilização. Assim sendo, de certa forma, isso garante um grau maior de liberdade de expressão do que a imposição de obrigações sancionáveis às redes, (como ocorre na Alemanha,) o que poderia resultar na censura prévia pela plataforma, particularmente quando houver dúvida quanto a permissibilidade do conteúdo.

## 7. UM REGIME JURÍDICO DIFERENCIADO: O MODELO PROPOSTO PARA A REGULAÇÃO DOS FÓRUMS CIBERNÉTICOS

Tanto a autorregulação como a regulação pela reserva de jurisdição apresentam vantagens e desvantagens próprias. A fim de conservar as principais vantagens de cada regime e combater seus problemas inerentes, sugere-se um modelo distinto, no qual a Administração Pública e as plataformas de redes sociais atuam conjuntamente.

As principais vantagens do regime de autorregulação (celeridade e efetividade) se dão pela possibilidade de ação independente pelas plataformas de redes sociais. As suas principais desvantagens, por sua vez, (falta de legitimidade, transparência e o conflito de interesses supramencionado) tem como origem a exclusão da atuação estatal.

As principais vantagens do regime de regulação judicial (a transparência, o amplo contraditório, a possibilidade recursal, a estabilidade da decisão e a falta de “censura preventiva”) advém dos princípios aplicáveis à atuação da Administração Pública, bem como da organização do Judiciário como um todo. Suas principais desvantagens, por outro lado, (a falta de celeridade, e, conseqüentemente, a possível inefetividade da tutela) decorrem da exclusão da atuação independente pelas empresas. Dessa forma, percebe-se que *a regulação efetiva desses espaços depende da cooperação entre o Estado e as empresas*. Afinal, não são espaços propriamente públicos ou privados, apresentando traços típicos de ambas as categorias, e há benefícios inerentes a atuação de ambas as classes de agente.

Tanto a Administração como as empresas devem ter a possibilidade de agir por iniciativa própria, removendo conteúdo diretamente, alertando a população quanto a conteúdos desinformativos e até mesmo suspendendo usuários.

A verificação da veracidade de determinadas informações é um assunto delicado, pois não é desejável tornar o Estado ou qualquer outra entidade em árbitro absoluto da verdade. Contudo, esse controle, em alguma medida, se demonstra necessário. Através da conferência do conteúdo de determinada publicação com o que consta nos sites de organizações internacionais especializadas (como a OMS, em matéria de saúde em geral, ou de vacinas em particular), seria possível realizar um certo controle, limitando



a margem de discricionariedade. Para evitar o abuso estatal dessa posição, todavia, a fim de evitar a concretização de uma distopia Orwelliana, sugere-se que o controle da veracidade de informações se dê *exclusivamente* através de um alerta na publicação quanto ao conteúdo inverídico ou duvidoso, e não através da remoção direta do conteúdo. Quanto aos usuários que já visualizaram ou compartilharam a informação falsa antes da aposição desse indicativo, seria interessante que lhes fosse encaminhada uma notificação, alertando-os da inveracidade.

*O discurso de ódio, por sua vez, deve ser removido diretamente das redes, ao invés de ter um indicativo informando usuários de sua natureza. Por ser uma espécie de expressão que atenta contra a dignidade da pessoa humana, fere a construção de uma sociedade justa, igualitária e solidária sem contribuir efetivamente ao debate. Não deve ser tolerado. Quando o conteúdo de uma publicação for manifestamente discurso de ódio, a rede social deve ter a obrigação de removê-lo diretamente, sem a necessidade de provocação estatal ou por usuários. A rede social poderia ser responsabilizada pela não remoção do conteúdo dentro de determinado prazo, conforme a NetzDG alemã.*

Quando a natureza do discurso não for manifestamente aquela de discurso de ódio, contudo, o tratamento deve ser diferente. *Nos casos em que a natureza do discurso não for manifestamente de ódio, as empresas não devem ser responsabilizáveis*, sob pena de haver uma censura prévia de qualquer publicação que dê vazão à dúvida, ocasionando grande prejuízo à liberdade de expressão. Nesses casos, seria necessária a atuação Estatal, seja através do Judiciário – quando provocado – seja através de uma agência reguladora.

Reconhece-se, contudo, que a própria definição de “discurso de ódio manifesto” já acarreta uma discussão complexa. A definição dos critérios exatos caberia à Administração Pública, através da evolução legislativa e jurisprudencial. Como critério tentativo, a fim de minimizar a restrição da liberdade de expressão, mas proteger a dignidade das minorias alvejadas por tais discursos, sugere-se: “será considerado discurso de ódio manifesto qualquer fala que explicitamente defenda a violência ou a inferioridade de qualquer pessoa ou grupo, em razão de sua cor, gênero, raça, orientação sexual, etnia, entre outros” (onde, sem dúvida, é necessário um estudo maior para a realização de uma listagem exhaustiva).

Com a criação de uma espécie de “competência concorrente” entre o Estado e as empresas de redes sociais para a regulação do conteúdo postado, seria possível efetuar um combate mais efetivo contra a desinformação e o discurso de ódio. Em casos em que houver o enquadramento manifesto de uma publicação em uma dessas categorias, a empresa seria responsável pela remoção (embora o Estado pudesse igualmente determiná-la, caso necessário); nos casos em que houvesse dúvida, o Estado poderia agir diretamente ou ser provocado pela empresa. A definição exata do que constituiria a “manifesta” desinformação ou discurso de ódio dependeria da elaboração de critérios legislativos próprios, ou da evolução jurisprudencial.



Seria possível recorrer ao Judiciário no caso de discordância pelo usuário quanto às decisões de remoção pela empresa ou pelo Estado. Finalmente, quanto à responsabilização penal dos usuários, não deve haver a delegação desse dever às empresas particulares, por falta de legitimidade, controle e transparência. A empresa deveria encaminhar uma denúncia diretamente à entidade estatal competente, em protocolo simplificado a ser estabelecido.

Com isso, pretende-se conservar os benefícios de celeridade e efetividade da autorregulação, bem como as garantias de segurança oferecidas pelo controle judicial. Ao mesmo tempo, a falta de legitimidade e a preocupação com o “chilling effect” são eliminadas, ou ao menos atenuadas.

Por fim, reconhece-se que o modelo proposto não elimina o conflito de interesses inerente à autorregulação das redes sociais, no que diz respeito à desinformação e o discurso de ódio. Enquanto esse tipo de publicação gerar um engajamento maior do que outros tipos de conteúdo, e o modelo de negócios das empresas em comento estiverem centralizados na monetização do engajamento, o problema irá subsistir. Contudo, não se vislumbra um modelo de regulação que permita a ação célere e eficaz, sem prejuízos significativos à liberdade de expressão, que não pressuponha a iniciativa regulatória pelas empresas, e o exercício pleno dessa iniciativa.

Assim sendo, apesar de ser uma medida drástica, sugere-se que o modelo de negócios da chamada “economia de atenção” seja o objeto de estudos próprios, para verificar a possibilidade de compatibilização com o combate à desinformação e ao discurso de ódio. Caso não seja possível essa compatibilização, o combate efetivo a essas ameaças pressupõe a proibição desse modelo de negócios, que trata os usuários como nódulos de produção a serem explorados, sem considerar os efeitos que essa exploração terá sobre o indivíduo ou a sociedade em geral.

## 8. CONSIDERAÇÕES FINAIS

A desinformação e o discurso de ódio são problemas complexos, cuja solução não depende exclusivamente do Governo, das plataformas de redes sociais, ou mesmo da sociedade civil em geral. O combate efetivo dessas ameaças exige a cooperação de todos os setores da sociedade. E não há dúvida que são ameaças significativas, haja vista os efeitos dessas táticas na manipulação de eleitores no Brasil e nos EUA, bem como no cometimento de genocídios em Myanmar. Diversos estudos, alguns dos quais citados nesse trabalho, demonstram a forte correlação entre a disseminação de discurso de ódio e a agressão física contra os alvos desses discursos. A desinformação pode ser utilizada para exacerbar preconceitos pré-existentes ou criar novos, um fenômeno particularmente preocupante quando se leva em conta que sua velocidade de difusão é muito superior à da verdade. Quando essas táticas são utilizadas em redes sociais, a



própria estrutura da plataforma facilita sua difusão e polariza cada vez mais os ânimos dos usuários, através das chamadas “câmaras de eco”.

Para a manutenção da democracia, a regulação efetiva desses espaços é imprescindível. Todavia, os regimes atualmente vigentes nos EUA, na Alemanha e no Brasil apresentam falhas preocupantes e faz-se necessário um modelo regulatório alternativo. Assim sendo, propôs-se um modelo regulatório misto que visa aproveitar os benefícios inerentes à Autorregulação e a regulação judicial e mitigar as falhas inerentes a cada modelo, a fim de combater efetivamente essas ameaças.

Contudo, talvez não seja necessário ir tão longe. Reconhece-se, que o regime de “autorregulação regulada”, *após sucessivos aperfeiçoamentos*, talvez bastasse para a regulação eficaz dos fóruns cibernéticos (apesar da legitimidade questionável das plataformas para regular matérias como a saúde pública e a liberdade de expressão).

Quanto ao grau de utilidade da dicotomia tradicional, os resultados são menos claros. A princípio, verifica-se a necessidade de cooperação entre a Administração Pública e as empresas particulares, ao ponto em que seria interessante que todas pudessem ter a iniciativa regulatória e fiscalizatória, em uma espécie de “competência concorrente”. Por outro lado, mesmo com essa concorrência, segundo o modelo proposto, os papéis da Administração e dos administrados não seriam os mesmos. Ambos poderiam remover uma publicação ou determinar sua remoção, por exemplo, mas recursos dessa decisão só poderiam ser propostos perante o Estado, e a definição dos casos mais controvertidos caberia igualmente ao Estado com exclusividade. Em alguns momentos do controle, os entes públicos se confundem com os particulares, mas em outros a diferença é nítida.

Diante disso, não se afirma que a dicotomia é inútil na criação do regime diferenciado necessário à regulação efetiva dos fóruns cibernéticos – somente que deve haver muita cautela no seu emprego, com a flexibilização dos respectivos conceitos conforme a necessidade.

## REFERÊNCIAS

ALCOTT, Hunt; GENTZKOW, Mathew, Social Media and Fake News in the 2016 Election. **Journal of Economic Perspectives**, [s.l.], vol. 31, n. 2, p. 211-236. 2017.

ARAÚJO, Valter Shuenquener de; ZULLO, Bruno Almeida; TORRES, Maurílio. Big Data, algoritmos e inteligência artificial na Administração Pública: reflexões para a sua utilização em um ambiente democrático. **A&C – Revista de Direito Administrativo & Constitucional**, Belo Horizonte, ano 20, n. 80, p. 241-261, abr./jun. 2020.

BALKIN, Jack M. Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation. **UC Davis Law Review**, vol. 51/2018, p. 1149-1210. 2018. Disponível em: [https://digitalcommons.law.yale.edu/fss\\_papers/5160/](https://digitalcommons.law.yale.edu/fss_papers/5160/). Último acesso em 31/01/2021.



BAPTISTA, Patrícia; KELLER, Clara Iglesias. Por que, quando e como regular as novas tecnologias? Os desafios trazidos pelas inovações disruptivas. **RDA – Revista de Direito Administrativo**, Rio de Janeiro, vol. 273, p. 123-163. 2016.

BENNET, Colin J.; LYON, David. Data driven elections: implications and challenges for democratic societies. **Internet Policy Review**, vol. 8, n. 4, p. 1-16. 2019. DOI: 10.14763/2019.4.1433.

BUNDESGERICHT, **Arrêt 1C\_338/2018**, vol. I/2019 (145) – 207. Disponível em: <[https://www.bger.ch/ext/eurospider/live/fr/php/aza/http/index.php?highlight\\_docid=aza%3A%2F%-2F10-04-2019-1C\\_338-2018&lang=fr&type=show\\_document&zoom=NO](https://www.bger.ch/ext/eurospider/live/fr/php/aza/http/index.php?highlight_docid=aza%3A%2F%-2F10-04-2019-1C_338-2018&lang=fr&type=show_document&zoom=NO)>. Último acesso em 29 jan. 2021.

CARNEIRO, Gustavo Ferraz Sales. Autorregulação de fake news no Facebook: incentivos e freios à proliferação de desinformação. **Revista de Direito e as Novas Tecnologias**, [s.l.], vol. 3/2019, p. 1-11, abr./jun. 2019.

CDA 230: Legislative History. **Electronic Frontier Foundation**, disponível em: <https://www.eff.org/issues/cda230/legislative-history>. Último acesso em 31 de janeiro de 2021.

COSTA, Rodrigues de Souza; RIBEIRO, Raisa Duarte da Silva. Discurso de ódio no Brasil: uma análise do HC 82.424/RS e seus reflexos. **Direitos e Garantias Fundamentais II - XXV Congresso Nacional do Conpedi**, 2015, Florianópolis, p. 138-158. Disponível em: <<http://site.conpedi.org.br/publicacoes/02q8agmu/j8667276/r177M5DQXgWii38p.pdf>>. Último acesso em 25/06/2021.

CUEVA, Ricardo Villas Bôas. Alternativas para a remoção de fake news das redes sociais. **Revista de Direito e as Novas Tecnologias**, [s.l.], vol. 1/2018, p. 1-10, out./dez. 2018.

DAVID KAYE: Speech Police: The Global Struggle to Govern the Internet. 2019. 1 video (1:04:18). Publicado pelo canal Carnegie Council for Ethics in International Affairs. Disponível em: <<https://www.youtube.com/watch?v=W6PDZ-o5Khg>>. Último acesso em 31 jan. 2021.

EBERLE, Edward J. Dignity and Liberty: Constitutional Visions in Germany and the United States. **Praeger Publishers**, Westport, 2002.

EDWARDS, Griffin; RUSHIN, Stephen. The Effect of President Trump's Election on Hate Crimes. 2018. SSRN. Disponível em: <[https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3102652](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3102652)>. Último acesso em 13 de fevereiro de 2021.

EVANGELISTA, Rafael; BRUNO, Fernanda. WhatsApp and political Instability in Brazil: targeted messages and political radicalisation. **Internet Policy Review**, [s.l.], vol. 8, n. 4, p. 1-23. 2019. DOI: 10.14763/2019.4.1434.

FLICK, Dallas. Combatting Fake News: Alternatives To Limiting Social Media Disinformation And Rehabilitating Quality Journalism. **Science and Technology Law Review**, [s.l.], vol. 20, n. 2, p. 375-405. 2017.

FORNASIER, Mateus de Oliveira; BORGES, Gustavo Silveira. The current influence of social media on democratic debate, political parties and electioneering. **Revista de Investigações Constitucionais**, Curitiba, vol. 9, n. 1, p. 73-102, jan./abr. 2022.



FREITAS, Juarez. Direito Administrativo e inteligência artificial. **Interesse Público** – IP, Belo Horizonte, ano 21, n. 114, p. 15-29, mar./abr. 2019.

GESLEY, Jenny. Initiatives to Counter Fake News in Selected Countries. **Law Library of Congress**, [s.l.], p. 33-40, abr. 2019.

HARFF, Graziela; DUQUE, Marcelo Schenk. Discurso de ódio nos contextos alemão e brasileiro. **A&C – Revista de Direito Administrativo & Constitucional**, Belo Horizonte, ano 21, n. 84, p. 199-225, abr./jun. 2021.

KAYE, David. *Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression*, **United Nations Human Rights Council**, 38<sup>th</sup> Session, 6 April 2018.

LAZER, David et al. The science of fake news. **Science**, [s.l.], vol. 359, n. 6380, p. 1094-1096. 2018.

MILL, John Stuart. On Liberty. **Ticknor and Fields**, Boston, 1863.

MOURA, Marco Aurélio. O discurso do ódio em redes sociais. **Lura Editorial**, São Paulo, 2016.

MÜLLER, Karsten; SCHWARZ, Carlo. Fanning the Flames of Hate: Social Media and Hate Crime. 2019. SSRN. Disponível em: <<https://ssrn.com/abstract=3082972>>. Último acesso em 13 de fevereiro de 2021.

MÜLLER, Karsten; SCHWARZ, Carlo. From Hashtag to Hate Crime: Twitter and Anti-Minority Sentiment. 2019. Disponível em: <[https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3149103](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3149103)>. Último acesso em 13 de fevereiro de 2021.

ONO, Taynara Tiemi, A regulação da desinformação na internet em período de campanha eleitoral: leis, normas sociais, arquitetura de rede e atuação da justiça eleitoral como mecanismos regulatórios adotados no contexto brasileiro. **Revista dos Tribunais**, [s.l.], vol. 1008, p. 93-109. 2019.

PAUL, Kari. Twitter and Facebook lock Donald Trump's accounts after video address. *The Guardian*. 7 Jan. 2021. Disponível em: <<https://www.theguardian.com/us-news/2021/jan/06/facebook-twitter-youtube-trump-video-supporters-capitol>>. Último acesso em 12 fev. 2021.

PUSCHEL, André Felipe Silva; RODRIGUES, Roberto Tessis; VALLE, Vivian Cristina Lima López. O dilema ético da decisão algorítmica na administração pública. **A&C – Revista de Direito Administrativo & Constitucional**, Belo Horizonte, ano 22, n. 90, p. 207-226, out./dez. 2022.

SANTOS, Elaine Gomes dos; RIBEIRO; Raísa Duarte da Silva. Restrições à liberdade de expressão e crimes cibernéticos: a tutela penal do discurso de ódio nas redes sociais. **Revista dos Tribunais**, [s.l.], vol. 997/2018, p. 515-514, nov. 2018.

SANTOS, Gustavo Ferreira. Social media, disinformation, and regulation of the electoral process: a study based on 2018 Brazilian election experience. **Revista de Investigações Constitucionais**, Curitiba, vol. 7, n. 2, p. 429-449, maio/ago. 2020.

SARMENTO, Daniel. A liberdade de expressão e o problema do “hate speech”. **Revista de Direito do Estado**, Rio de Janeiro, ano 1, n. 4, p. 1-58, out./dez. 2006.





STROSSEN, Nadine. Hate: Why We Should Resist It With Free Speech, Not Censorship. **Oxford University Press**, New York. 2018.

TACCA, Adriano; ROCHA, Leonel Severo, Inteligência Artificial: reflexos no sistema de Direito. **Revista do Programa de Pós-Graduação em Direito da UFC**, [s.l.], vol. 38, n. 2, p. 53-68, jul./dez. 2018.

TOFFOLI, José Antonio Dias. Fake News, Desinformação e Liberdade de Expressão. **Interesse Nacional**, [s.l.], ano 12, n. 46, p. 1-15, jul./set. 2019. Disponível em: <<http://interessenacional.com.br/2019/07/11/fake-news-desinformacao-e-liberdade-de-expressao/>>. Último acesso em 25/06/2021.

UNITED STATES COURT OF APPEALS, Fourth Circuit. **Brian Davison v. Phyllis Randallet. AI**. [syllabus]. 2019. Disponível em: <<https://www.ca4.uscourts.gov/opinions/172002.P.pdf>>. Último acesso em 12 fev. 2021.

UNITED STATES COURT OF APPEALS, Ninth Circuit. **Prager University v. Google LLC, FKA Google Inc; Youtube, LLC** [opinion]. 2019. Disponível em: <<https://cdn.ca9.uscourts.gov/datastore/opinions/2020/02/26/18-15712.pdf>>.

UNITED STATES SUPREME COURT. **Knigh First Amendment Institute at Columbia University v. Donald J. Trump et. AI**. [syllabus]. 2018. Disponível em: <<https://www.law.cornell.edu/rio/citation/928F.3d226>>. Último acesso em 12 fev. 2021.

UNITED STATES SUPREME COURT. **Manhattan Community Access Corp. v. Halleck** [syllabus]. 2019. Disponível em: <<https://www.law.cornell.edu/supremecourt/text/17-1702>>. Último acesso em 12 fev. 2021.

VALLE, Vivian Cristina Lima López; FELISBERTO, Jéssica Heinzen. Administração Pública digital: limites e possibilidades em atenção à desigualdade social e ao custo dos direitos. **Revista Euro-latinoamericana de Derecho Administrativo**, Santa Fe, vol. 9, n. 1, p. 151-179, ene./jun. 2022.

VALLE, Vivian Cristina Lima López; GALLO, William Ivan. Inteligência artificial e capacidades regulatórias do Estado no ambiente da administração pública digital. **A&C – Revista de Direito Administrativo & Constitucional**, Belo Horizonte, n. 82, p. 67-86, out./ dez. 2020.

VINUESA, Ricardo et al. The role of artificial intelligence in achieving the Sustainable Development Goals. **Nature Communications**, [s.l.], vol 11, n. 1, artigo 233, p. 1-10. 2020. Disponível em: <<http://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-168566>>. Último acesso em 27/06/2021.

VOSOUGH, Soroush; ROY, Deb; ARAL, Sinan. The spread of true and false news online. **Science**, [s.l.], vol. 359, n. 6380, p. 1146-1151. 2018.

WAROFKA, Alex. An Independent Assessment of the Human Rights Impact of Facebook in Myanmar. **About Facebook**. 2018. Disponível em: <<https://about.fb.com/news/2018/11/myanmar-hria/>>. Último acesso em 31 jan. 2021.

WILSON, Richard A.; LAND, Molly K. Hate speech on Social Media: Towards a Context-Specific Content Moderation Policy. **Connecticut Law Review**, Connecticut, vol. 52. 2020.