

### *Fake News* no Contexto de Pandemia e Emergência Social: os Deveres e Responsabilidades das Plataformas de Redes Sociais na Moderação de Conteúdo *Online*: entre a Teoria e as Proposições Legislativas<sup>1</sup>

### *Fake News during the Pandemics and Social Emergency: the Duties and Responsibilities of Social Media Platforms in Online Content Moderation: between Theory and Legislative Proposals*

**IVAR ALBERTO HARTMANN<sup>2</sup>**

Escola de Direito da Fundação Getúlio Vargas, Rio de Janeiro, Brasil.

**JULIA IUNES MONTEIRO<sup>3</sup>**

Escola de Direito da Fundação Getúlio Vargas, Rio de Janeiro, Brasil.

RESUMO: A disseminação de *fake news* durante a pandemia da Covid-19 ou outros momentos de emergência social representa desafio complexo para a regulação. Quais medidas o Estado e as plataformas de mídia social devem adotar, segundo a literatura sobre moderação de *fake news* ligada a questões de saúde pública ou na moderação de conteúdo abusivo em geral? O Poder Legislativo adota tais diretrizes fornecidas pelos estudos científicos? Para responder à primeira pergunta de pesquisa, adotamos metodologia de ampla revisão de literatura de modo a identificar diretrizes consensuais nos estudos acadêmicos e relatórios sobre o tema elaborados por entidades governamentais e privadas. Para responder à segunda pergunta de pesquisa, adotamos metodologia empírico-quantitativa, mediante análise documental individualizada do universo de todos os 49 projetos de lei no Congresso

---

1 O presente estudo é um dos resultados de projeto de pesquisa sobre Moderação de Conteúdo *Online* do Centro de Tecnologia e Sociedade da FGV Direito Rio, iniciado em 2019 e com previsão de conclusão em 2021, viabilizado por financiamento da própria Fundação Getúlio Vargas e por doação, sem qualquer contrapartida, do Facebook. O projeto de pesquisa conta com a participação de docente do Programa de Pós-Graduação em Direito da Regulação da FGV Direito Rio, bem como pesquisadores de nível de doutorado, mestrado e graduação.

2 Orcid: <<https://orcid.org/0000-0002-8497-6291>>.

3 Orcid: <<https://orcid.org/0000-0002-4795-3832>>.

Nacional, desde a entrada em vigor do Marco Civil da Internet, que propõem alterações ao regime de obrigações dos provedores de aplicação. Construímos uma tipologia para classificar tais proposições. Os resultados da análise doutrinária e da pesquisa empírica indicam uma profunda desconexão entre, de um lado, a teoria sobre como *fake news*, inclusive durante a pandemia, deveria ser regulada e, de outro, o conjunto das proposições legislativas discutidas no Congresso.

**PALAVRAS-CHAVE:** *Fake news*; Covid-19; moderação de conteúdo *online*; liberdade de expressão; autorregulação.

**ABSTRACT:** The dissemination of fake news during the Covid-19 pandemics and other situations of social emergency represents a complex challenge for regulation. What measures should the State and social media platforms adopt, according to the literature on moderation of fake news about health issues and other abusive online content? Does the Legislator Branch adopt such guidelines provided by scientific studies? In order to answer the research question we adopt a method of broad literature review to identify consensus on academic studies and reports on the subject produced by governmental and private entities. To address the second research question we adopt quantitative empirical methodology, individually analyzing the entire universe of all 49 draft bills proposed in Congress since the enactment of the Civil Rights Framework for the Internet in Brazil proposing changes to the legal regime of duties and responsibilities of application providers. We elaborate a typology to classify such proposals. The findings of the literature review and the empirical research indicate a profound disconnect between, on the one hand, the theory on how to regulate fake news, including during a time of pandemic, and, on the other, the group of draft bills currently discussed in Congress.

**KEYWORDS:** Fake news; Covid-19; online content moderation; freedom of expression; self-regulation.

**SUMÁRIO:** Introdução; 1 A moderação de *fake news* e conteúdo abusivo *online*; 1.1 Poder Judiciário; 1.2 Usuários; 1.3 Plataformas; 1.4 Poder Executivo; 2 Metodologia; 3 Resultados; 4 Discussão dos resultados e conclusões; Referências.

## INTRODUÇÃO

O Marco Civil da Internet (Lei nº 12.965), que entrou em vigor em 2014, define as principais regras para o uso da internet no Brasil, regulamentando também os deveres e as responsabilidades de plataformas de mídia social como Facebook e YouTube, os denominados provedores de aplicação. No entanto, seis anos após a promulgação do Marco, o constante desafio do equilíbrio entre a liberdade de expressão e outros direitos *online* apresenta novos elementos, demandando reflexão sobre que novas exigências normativas seriam necessárias para lidar com o cenário atual da internet. A disseminação de *fake news*, especialmente durante a pandemia da Covid-19, é um desses novos elementos, com importância central para a saúde pública e a integridade do debate público, tendo consequências diretas na sustentabilidade da democracia (Goldman; Baker, 2019, p. 125). Durante a pandemia, a disseminação de teorias da conspiração em redes

sociais e a politização de questões de saúde pública – como a conveniência do uso de determinados medicamentos para tratar a doença – são fatores que tornam o combate do vírus pela sociedade muito mais difícil (Bavel; Baicker; Boggio *et al.*, 2020, p. 465). No intuito de evitar a disseminação de conteúdos falsos relativos à pandemia, diversas plataformas de redes sociais promoveram alterações em seus termos de uso. O Twitter, por exemplo, adotou novos parâmetros para a remoção de *posts* que contenham “negação das recomendações de autoridades de saúde locais ou globais; [...] descrição de supostas curas alegadas para Covid-19; [...] descrição de tratamentos prejudiciais ou medidas de proteção ineficazes [e] [...] negação de fatos científicos estabelecidos”<sup>4</sup>. Em que pese a necessidade de medidas por parte dos provedores de aplicação com a finalidade de dirimir riscos à saúde coletiva, a remoção de conteúdos, ainda que falsos, pode custar a violação de outros direitos, como a liberdade de expressão, sendo também problemática a concentração do poder decisório nessas plataformas que, ao atuarem como agentes de censura privada, acabam por definir as pautas que permanecem ou são excluídas do debate público.

Três situações emblemáticas podem ser utilizadas para representar alguns desses dilemas. Um primeiro exemplo se refere às teorias conspiratórias amplamente divulgadas no início da pandemia, abrangendo desde alegações de que o coronavírus teria sido criado em laboratório pela China, no intuito de ser utilizado como arma biológica<sup>5</sup>, até manifestações que negavam a própria existência de uma pandemia<sup>6</sup>. Em resposta a casos como esses, as plataformas não só passaram a remover conteúdos, como também anunciaram outras medidas, como a notificação dos usuários que interagissem com boatos e a priorização, nos mecanismos de busca e recomendação algorítmica, de conteúdos provenientes de fontes oficiais, como os da Organização Mundial da Saúde (OMS)<sup>7</sup>. Um segundo caso que merece referência trata-se de um vídeo publicado pelo atual Presidente da República, Jair Bolsonaro, em seus perfis do Facebook, Twitter e Instagram, no

---

4 Uma atualização sobre nossa estratégia continua durante a Covid-19. Twitter, 16 de março de 2020. Disponível em: <[https://blog.twitter.com/pt\\_br/topics/company/2019/uma-atualizacao-sobre-nossa-estrategia-continua-durante-o-covid-19.html](https://blog.twitter.com/pt_br/topics/company/2019/uma-atualizacao-sobre-nossa-estrategia-continua-durante-o-covid-19.html)>.

5 MORAES, Maurício. Na web, teorias da conspiração apontam China e EUA como criadores da Covid-19. Piauí, 4 de agosto de 2020. Disponível em: <<https://piaui.folha.uol.com.br/lupa/2020/08/04/coronaverificado-origem-covid/>>.

6 BERGAMO, Mônica. Google remove vídeos negando pandemia após intervenção do Ministério Público. *Folha de São Paulo*, 8 de maio de 2020. Disponível em: <<https://www1.folha.uol.com.br/colunas/monicabergamo/2020/05/google-remove-vidEOS-negando-pandemia-apos-intervencao-do-ministerio-publico.shtml>>.

7 Facebook will add anti-misinformation posts to your News Feed if you liked fake coronavirus news. *The Verge*, 16 de abril de 2020. Disponível em: <<https://www.theverge.com/2020/4/16/21223456/facebook-coronavirus-misinformation-fake-news-warning-update-who>>.

qual se posiciona de maneira contrária às recomendações de isolamento social, fato que gerou a exclusão do conteúdo pelas plataformas<sup>8</sup>. Nessa situação em especial, verifica-se que, apesar dos riscos que tal pronunciamento possa desencadear à saúde pública nacional, por outro lado, parece razoável argumentar que, independente do mérito da postagem, existe um interesse social na publicidade dos atos promovidos por agentes políticos, especialmente quando se trata do líder do Poder Executivo. Casos como este revelam não só os potenciais riscos de censura na tarefa de moderação de conteúdo operada pelas plataformas, bem como a necessidade de que sejam estabelecidos critérios democraticamente legítimos destinados a orientar e fiscalizar essa prática.

Um terceiro caso emblemático, também envolvendo conduta de autoridade pública, refere-se à ex-Secretária de Cultura Regina Duarte, que publicou, em seu perfil do Instagram, que o medicamento hidroxicloroquina teria sido liberado pela Anvisa para tratamento da Covid-19<sup>9</sup>. Adotando alternativa diversa à remoção do conteúdo, a plataforma inseriu um indicador de que a postagem seria “parcialmente falsa”, direcionando os usuários a um esclarecimento fornecido por uma agência de checagem de fatos<sup>10</sup>.

Em todos os casos acima mencionados, o desafio urgente é a dispersão de notícias falsas sobre a Covid-19. Há comum acordo entre todos de que esse tipo de informação deveria ser, em algum grau, restringido pelo Poder Público e pelas plataformas, mas o problema é que não há acordo sobre o conceito de *fake news* e até pessoas que adotam o mesmo conceito discordam sobre como classificar *posts* e mensagens específicas. Segundo quais regras e mediante qual procedimento remover mensagens abusivas que circulam nas plataformas sobre a pandemia e seus efeitos? Considerando a pluralidade de interesses em questão e a necessária compatibilidade com a liberdade de expressão, será que a pura e simples remoção do conteúdo falso pelas plataformas seria a única atitude possível ou recomendada?

---

8 Facebook e Instagram removem vídeo de Jair Bolsonaro por violação de regras. *O Globo*, 30 de março de 2020. Disponível em: <<https://g1.globo.com/economia/tecnologia/noticia/2020/03/30/facebook-e-instagram-removem-video-de-jair-bolsonaro-por-violacao-de-regras.ghtml>>.

9 Instagram avisa que Regina Duarte publicou notícia parcialmente falsa. *Folha de São Paulo*, 1º de abril de 2020. Disponível em: <<https://www1.folha.uol.com.br/colunas/monicabergamo/2020/04/instagram-avisa-que-regina-duarte-publicou-noticia-parcialmente-falsa.shtml>>.

10 Na ocasião, foi esclarecido pela agência de checagem que a Anvisa teria liberado a realização de pesquisas sobre o efeito da substância, mas não o uso do medicamento para fins de tratamento da doença.

O conceito que adotamos de *fake news* é aquele defendido por mais de uma dezena de pesquisadores das mais prestigiadas universidades do mundo em artigo na revista *Science* intitulado “*The Science of Fake News*”: “*We define ‘fake news’ to be fabricated information that mimics news media content in form but not in organizational process or intent*”<sup>11</sup> (Lazer et al., 2018, p. 1094). Ou seja, a identificação e mitigação da disseminação de *fake news* passa longe do arbitramento de uma “verdade real”, como se fosse possível ou democraticamente legítimo que um órgão público ou privado centralizasse a decisão daquilo que é verdade ou falso no debate público. Pelo contrário, *fake news* diz respeito a procedimento de produção de notícias e intenção na sua disseminação. Segundo os autores mencionados, a caracterização de uma manifestação, *post*, montagem, clipe audiovisual etc. como *fake news* exige o preenchimento de dois requisitos. Primeiro, essa manifestação se apresenta como se notícia fosse. Adota a aparência de matéria jornalística, com um verniz proposital de autenticidade. Segundo, a elaboração dessa manifestação ou autoproclamada matéria jornalística não respeitou padrões jornalísticos. A imprensa discute e respeita, há muitas décadas, conjuntos de padrões de conduta na redação e publicação de matérias jornalísticas.

O rigor conceitual no tratamento da questão das *fake news* durante a pandemia da Covid-19 é importante, mas com certeza não suficiente para resolver o problema. O estudo dessa questão e, portanto, o presente artigo se justificam em termos de relevância por pelo menos duas razões. Primeiro, porque traz um elemento urgente de saúde pública a um debate relevante e preexistente sobre os limites da liberdade de expressão na internet. Até então, pouquíssimos estudos sobre a manifestação do pensamento *online* focaram o risco à saúde pública causado por determinados discursos ou declarações. Segundo, porque, no contexto justamente da complexidade da determinação do que é *fake news* sobre o coronavírus, ocorreu um fato que possivelmente trará maior atenção da sociedade civil brasileira para a questão da atuação das plataformas digitais na tomada de decisão sobre o que os brasileiros podem ou não dizer sobre saúde pública. Conforme mencionado anteriormente, um *post* do Presidente da República foi removido por redes sociais por ser enquadrado como notícia falsa sobre o vírus<sup>12</sup>.

---

11 Em tradução livre: “A Ciência das Notícias Falsas”: “Nós definimos ‘notícias falsas’ como sendo informações fabricadas que imitam o conteúdo de novas mídias com relação a sua forma, mas não com relação ao processo organizacional ou sua intenção”.

12 Sobre o assunto, ver notícia disponível em: <<https://www1.folha.uol.com.br/poder/2020/03/depois-do-twitter-facebook-tambem-apaga-post-de-bolsonaro.shtml>>. Acesso em: 2 maio 2020. Ver, sobre o episódio, HARTMANN, Ivar. Combate a *fake news* requer critérios democraticamente legítimos. Opinião. *Folha de S.*

Naturalmente, os deveres e as responsabilidades das plataformas digitais no contexto da disseminação de notícias falsas sobre o coronavírus são regulamentados pelo Marco Civil da Internet. O problema contém, no entanto, dimensões e nuances que não eram previsíveis para o legislador antes da entrada em vigor do Marco. Conforme relatado, há consenso sobre a danosidade da disseminação de *fake news* durante a pandemia, mas grande nível de discordância sobre como reagir a esse problema. Forçar as plataformas a realizar filtragem prévia? Submeter todo e qualquer *post* sobre Covid-19 em rede social à análise do Poder Judiciário? Há algum papel regulatório para o Poder Executivo? O melhor é apenas confiar na autorregulação?

Buscamos, neste estudo, responder a duas perguntas de pesquisa. A primeira é: quais medidas o Estado e as plataformas de mídia social devem adotar, segundo a literatura sobre moderação de *fake news* ligadas a questões de saúde pública ou na moderação de conteúdo abusivo em geral? A segunda pergunta é: o Poder Legislativo adota tais diretrizes fornecidas pelos estudos científicos? Ou seja, o objetivo é avaliar se o Direito brasileiro está andando no caminho apontado pelas pesquisas sobre como combater *fake news* e conteúdo abusivo na internet. Especialmente no contexto da tramitação conturbada do PL 2.630/2020, que foi apelidado de “PL das *Fake News*”, qualquer que seja a avaliação que se possa fazer sobre o mérito de um projeto de lei isolado, é muito diferente de uma avaliação empírica do conjunto das propostas legislativas no Congresso. A tentativa, portanto, é de identificar padrões gerais (e não o conteúdo de dispositivos em propostas específicas) e contrastá-los com as recomendações de estudos científicos sobre moderação de *fake news*.

Para responder à primeira pergunta de pesquisa, adotamos metodologia de ampla revisão de literatura de modo a identificar diretrizes consensuais nos estudos acadêmicos e relatórios sobre o tema elaborados por entidades governamentais e privadas. Para responder à segunda pergunta de pesquisa, adotamos metodologia empírica quantitativa, mediante análise documental individualizada do universo de todos os 49 projetos de lei no Congresso Nacional, desde a entrada em vigor do Marco Civil da Internet, que propõem alterações ao regime de obrigações dos provedores de aplicação. Para permitir a análise dos padrões, construímos uma tipologia para classificar os projetos legislativos.

Além do amplamente discutido PL 2.630/2020, já existe evidência anedótica do tipo de proposta atualmente em consideração pelo Congresso em reação ao problema das *fake news* na pandemia. O PL 693/2020, proposto em março de 2020 na Câmara dos Deputados, pretende responsabilizar subsidiariamente as plataformas que houverem hospedado mensagens com recomendações contrárias àquelas dos órgãos de saúde. Rompendo com o critério importante do Marco Civil de responsabilizar as redes sociais apenas se descumprem ordem judicial de remoção, a proposta faria com que a mera notificação da autoridade de saúde fosse suficiente para gerar obrigação legal das plataformas de censurar *posts* de brasileiros. A obrigação seria de remover “notícias falsas e que estejam contra as orientações oficiais das autoridades nacionais e internacionais e aos consensos científicos durante emergência à saúde decretada pela Organização Mundial de Saúde”<sup>13</sup>. E o que deveria fazer o Facebook ou Google se as orientações oficiais das autoridades nacionais são contrárias às orientações oficiais das autoridades internacionais?

Esse é apenas um exemplo que demonstra o inevitável alto custo para a liberdade de expressão quando os estudos sobre o tema são ignorados e, por consequência, critérios de impossível operacionalização são usados para obrigar as plataformas a remover manifestações potencialmente legítimas e legais. Mas o PL 693 é representativo do que tramita no Congresso? E o PL 2.630? Qualquer que seja o destino desses projetos, se arquivados ou aprovados e sancionados, é importante identificar o perfil das propostas atualmente em consideração no Legislativo Federal de modo a permitir a avaliação da qualidade da intervenção legislativa e previsões sobre as contribuições do Congresso para o problema das *fake news* nos próximos anos.

## 1 A MODERAÇÃO DE *FAKE NEWS* E CONTEÚDO ABUSIVO ONLINE

As plataformas digitais elaboram suas próprias regras e realizam sua própria avaliação casuística da compatibilidade das manifestações em suas

---

13 Conforme redação do art. 5º do referido projeto, que insere o art. 21-A ao Marco Civil da Internet: “Art. 21-A. O provedor de aplicações de internet que disponibilize conteúdo gerado por terceiros será responsabilizado subsidiariamente pela segurança sanitária decorrente da divulgação e propagação de notícias falsas e que estejam contra as orientações oficiais das autoridades nacionais e internacionais e aos consensos científicos durante emergência à saúde decretada pela Organização Mundial de Saúde, quando após o recebimento de notificação pela autoridade sanitária ou nos termos do § 5º do art. 319 do Código de Processo Penal, deixar de promover, de forma diligente, no âmbito e nos limites técnicos do seu serviço, a indisponibilização desse conteúdo” (BRASIL. Projeto de Lei nº 693/2020. Disponível em: <[https://www.camara.leg.br/proposicoesWeb/prop\\_mostrarintegra;jsessionid=E394A340020C245B4B22F88F6E520ED1.proposicoesWebExterno1?codteor=1867119&filename=PL+693/2020](https://www.camara.leg.br/proposicoesWeb/prop_mostrarintegra;jsessionid=E394A340020C245B4B22F88F6E520ED1.proposicoesWebExterno1?codteor=1867119&filename=PL+693/2020)>).

redes com tais regras (Caplan, 2018, p. 12). Essa atividade de regulação privada é altamente sofisticada e, apesar de guardar certas similaridades com o exercício da função jurisdicional no sentido de exigir decisões individuais sobre o mérito de determinadas manifestações, possui características que a literatura julgou justificar o reconhecimento de uma nova categoria de atuação: a moderação de conteúdo. James Grimmelman (2015) oferece a descrição mais detalhada dos elementos, contextos e efeitos das “virtudes da moderação”. Com isso, essas empresas, como Facebook, Twitter ou YouTube, já exercem papel central na regulação do ambiente e do debate público *online*, inclusive influenciando a definição dos contornos da liberdade de expressão em diversos países (Pasquale, 2016) por meio não apenas da criação de regras sobre conteúdo supostamente abusivo, mas, também, da aplicação dessas regras (Denardis, 2012).

Apesar de essa dinâmica refletir uma tendência da nova era digital, a literatura especializada aponta que a concentração da regulação nesses organismos privados (Klonick, 2018; Suzor; Van Geelen; Myers West, 2018) é problemática sob o ponto de vista da proteção de direitos fundamentais, especialmente quando são relatados problemas de transparência no processo de moderação do conteúdo (Jhaver; Bruckman; Gilbert, 2019) e na operação dos algoritmos (Tufekci, 2015) utilizados por essas empresas. Diante disso, surgem novos questionamentos sobre os deveres e responsabilidades que as plataformas deveriam assumir perante a sociedade, tendo em vista o impacto social e político que circunda suas atividades. A principal lição fornecida por esses estudos é que ampliar o já imenso poder dessas plataformas para definir o que é *fake news* não é desejável.

Em paralelo a isso, a literatura sobre governança da internet já desenvolve há muito tempo o conceito da atuação multissetorial (Kleinwächter, 2009; Mueller, 2007; Belli, 2015), que, no caso da resposta às *fake news* durante a pandemia, implica reconhecer que as plataformas digitais são apenas um dos tipos de atores interessados em um amplo ecossistema que envolve a participação das instituições estatais e da sociedade civil e, portanto, atuação conjunta dos Poderes Executivo, Legislativo e Judiciário e também dos próprios usuários das redes. Diante disso, as abordagens regulatórias acerca dos deveres e responsabilidades a serem assumidos pelas plataformas digitais devem levar em consideração o fato de que a atuação desses organismos privados se insere em uma dinâmica multiparticipativa – consagrada no art. 24, I, do Marco Civil –, na qual governos, provedores de aplicação e sociedade civil desempenham papéis diversos e, idealmente, complementares.

Esse modelo multissetorial de moderação do conteúdo *online* importa em papéis e atuações de diferentes tipos, âmbitos e características conforme as instituições ou grupo participante.

### 1.1 PODER JUDICIÁRIO

Uma premissa importante é de que as Cortes possuem um papel proeminente na ponderação de direitos em sistemas de avaliação de discursos (Sarlet, 2016) pelo fato de que, dentre outras razões, o exercício da ponderação significa que, geralmente, regras gerais sobre quando o discurso deva prevalecer estão fora de questão, e as respostas só podem ser alcançadas após um exame das características de cada caso em particular (Barcellos, 2014; Barroso, 2004) – ainda que os tribunais constitucionais que exercem a ponderação possam ser descritos como partes que, com mais frequência, prefiram pender em favor da liberdade de expressão (Stern, 2006; Starck, 2005; Isensee; Kirchhof, 1989).

No entanto, em uma sociedade em rede, as Cortes não são mais as únicas instituições capazes de punir socialmente os discursos e manifestações abusivas como as *fake news* sobre a pandemia, tendo em vista que, nesse novo cenário, plataformas de conteúdo privado estabelecem regras sobre o conteúdo e fazem com que estas sejam cumpridas (Brown; Marsden, 2013), enquanto que os próprios usuários adquirem o papel de criticar, denunciar, rejeitar e boicotar discursos (ou seus próprios autores), fornecendo, assim, uma resposta rápida e, geralmente, mais efetiva do que juízes jamais poderiam dar na maioria dos casos. Mais importante ainda, o número de decisões individuais sobre liberdade de expressão que as Cortes teriam que emitir, de maneira a permanecer como único ou principal ente avaliador de *fake news*, é simplesmente impensável. E a porção de casos nos quais as Cortes encontram uma maneira de decidir as coloca, em grande parte das vezes, em uma posição de facilitadoras da censura privada<sup>14</sup>.

As transformações empreendidas pelo digital no exercício da liberdade de expressão demandam, portanto, uma transformação no papel do Poder Judiciário por ao menos duas razões: i) o intenso fluxo de informações na rede, o que torna inviável uma avaliação individual dos

---

14 No Brasil, por exemplo, o Relatório de Transparência do Google (2018) revela que requisições judiciais de remoção de conteúdo são mais numerosas que requisições provenientes do Poder Executivo em quase todos os anos na última década. No último relatório de junho de 2018, as requisições judiciais consistiram em 63% do total, em comparação com 43% de solicitações da Administração. Nesse período, difamação foi a maior causa dos pedidos de remoção de conteúdo, representando 46% do total. Relatório disponível em: <<https://transparencyreport.google.com/government-removals/overview?hl=en>>.

casos, e ii) a diminuição do desequilíbrio na capacidade comunicativa entre as pessoas, de modo que, atualmente, são cada vez menores os motivos para que o Judiciário decida disputas de difamação entre indivíduos (Hartmann, 2018).

Dessa maneira, no lugar de avaliar casos particulares de liberdade de expressão no que concerne ao mérito do conteúdo, os juízes deveriam passar a avaliar os elementos procedimentais dos sistemas de moderação de conteúdo criados e administrados pelas plataformas. Isto é, verificar se o esquema de moderação estabelecido pela plataforma respeita regras mínimas de devido processo, especialmente em sistemas legais nos quais os direitos fundamentais vinculam não apenas o Estado, mas também entes privados (Sarlet; Hartmann, 2019).

A ideia de que o foco do controle deve passar do mérito do conteúdo para a qualidade do procedimento e as garantias sobre ele não é nova. Na verdade, a noção de que sistemas virtuais de tomada de decisão pública – total ou parcialmente automatizados – precisam respeitar um devido processo tecnológico básico foi lançada de forma pioneira por Danielle Citron (2008) há mais de uma década. No contexto de plataformas privadas de redes sociais, Rebecca Tushnet (2008) criticou o papel atribuído a essas empresas pelo regime de responsabilidade “*safe harbor*”, que resultou em um desalinhamento dos incentivos em direção a uma censura privada legalmente protegida. Uma alternativa mais desejável seria “algum tipo de devido processo legal, autogovernança democrática, ou regra de não discriminação [...] [que deveria incluir] no mínimo, alternativas para empoderar os usuários dos grandes provedores de serviço de internet substancialmente e procedimentalmente”<sup>15</sup> (Tushnet, 2008, p. 131).

Esse é o papel que as Cortes deveriam assumir no combate às *fake news* durante a pandemia: fazer com que sejam cumpridas regras procedimentais para a moderação de conteúdo e avaliar a arquitetura e as regras básicas do fluxo de informação nessas plataformas. Dessa maneira, o Judiciário deveria ser o vetor de implementação de um sistema de autorregulação regulada – no qual, idealmente, o Legislativo irá fornecer as diretrizes procedimentais para a autorregulação do conteúdo a ser realizada pelas plataformas, permitindo que os juízes, por sua vez, verifiquem se essas re-

---

15 Tradução livre a partir do original: “Some type of procedural due process, democratic self-governance, or nondiscrimination rule [...] [which should include] at a minimum, alternatives for empowering users of major ISPs substantively and procedurally [...]” (TUSHNET, Rebecca. Power without responsibility: intermediaries and the first amendment. *The George Washington Law Review*, v. 76, p. 131, 2008).

gras básicas foram seguidas e, caso sim, se abstenham de intervir. David Kaye propõe elementos desse novo sistema, que deveria incluir, a partir de uma perspectiva procedimental, a tomada de decisão descentralizada e o aprimoramento radical da transparência – tanto no processo de elaboração das normas quanto no processo decisório (Kaye, 2019, p. 137).

## 1.2 USUÁRIOS

Quem deveria então realizar a análise de mérito de cada caso de *fake news* em particular? Em razão também do volume das manifestações que precisam ser submetidas a algum crivo, este trabalho é o novo papel para os usuários. Os usuários devem ser os filtros (Hartmann, 2017) pelo fato de que a vasta maioria dos discursos flui entre agentes que não possuem uma capacidade comunicativa especialmente dominante, como a de grandes jornais, em comparação com pessoas comuns. Salvos nas situações excepcionais de redes artificiais de disseminação de *fake news*, trata-se de pessoas comuns se expressando em um ambiente que nunca será totalmente como o ideal do livre mercado de ideias, mas que se assemelha mais a essa utopia desejável do jamais antes na história dos meios de comunicação em massa. Um conjunto perfeitamente objetivo e estático de normas substantivas sobre o que é ou não permitido em termos de expressão não é necessariamente uma precondição para que este tipo de autogovernança ocorra *online* (Risch, 2009), já que a autorregulação espontânea dos mundos *online* tem sido demonstrada há décadas (Humphreys, 2008). Essa solução de incorporar os usuários como moderadores do conteúdo oferece ainda a vantagem adicional de garantir compatibilidade com normas locais ou regionais de manifestação sobre saúde pública.

As condições que justificam a autorregulação (Ogus, 1995) do discurso estão presentes e tal modelo está mais alinhado com os ideais da democracia liberal, na medida em que significa que mais pessoas possam influenciar o debate público (Baker, 2001). A participação central dos próprios usuários na regulação do conteúdo evitaria alguns dos problemas de conflito de interesse típicos da regulação tradicional, os quais seriam reduzidos pela autorregulação (Stiglitz, 2009).

## 1.3 PLATAFORMAS

Diante do amplo cenário que caracteriza a moderação de conteúdo *online*, o papel das plataformas não seria o de exercer uma avaliação de casos individuais de *fake news*, mas o de possibilitar e capacitar – empode-

rar – os usuários para tanto. Neste sentido, as plataformas de mídia privada não podem ocupar o lugar de editores, que, como ocorria nas mídias tradicionais, eram responsáveis por selecionar individualmente os conteúdos que seriam impressos para seus leitores. Um sistema legal de garantia à liberdade de expressão não deve ser interpretado de forma a autorizar que as plataformas privadas possuam completa autonomia em decidir os discursos que serão filtrados (Balkin, 2018).

Muito pelo contrário: “Na medida em que mecanismos de busca, redes sociais e outras novas mídias predominantes simplesmente permitem conexões entre audiências e conteúdo, eles se encaixam ainda mais na extremidade ‘menos expressiva’” de um espectro que inclui veículos de imprensa tradicionais, rádio, televisão e redes por cabo (Pasquale, 2016: 507). Apesar disso, a concentração da regulação nesses organismos privados tem permitido que essas empresas filtrem mais conteúdo do que o legalmente necessário, assumindo, assim, uma posição de uma instituição quase judicial, ao arbitrar quais discursos são aceitáveis ou não na maioria das sociedades (Klonick, 2018). Há um agravante. As pessoas contratadas indiretamente pelas plataformas para realizar uma espécie de trabalho intelectual fungível de colarinho azul que consiste em visualizar centenas de *posts* diariamente para decidir se constituem violação dos “termos de serviço” sofrem com condições de trabalho precarizadas e falta de garantias de seguridade social e atendimento psicológico (Barrett, 2020). Como consequência, há muitos casos documentados de sequelas e distúrbios psicológicos que acometem aqueles que, dia após dia, exercem a tarefa de filtrar o que há de mais abjeto nas mensagens em redes sociais (Roberts, 2019).

Diante disso, os legisladores devem estabelecer obrigações procedimentais básicas para um “controle descentralizado” (Barzilai-Nahon, 2008; Shaw, 2012) das *fake news* sobre saúde pública, obrigações que as plataformas privadas devem cumprir, e o Judiciário, por sua vez, deve controlar. Dessa maneira, plataformas privadas de liberdade de expressão devem ser gradualmente pressionadas para serem menos como editores e mais como a Wikipedia, administrando “geografias altamente desiguais de participação” (Graham; Straumann; Hogan, 2015, p. 1174) e focando na complexa arquitetura de um sistema de autorregulação pelos usuários que emprega ferramentas de moderação jamais disponíveis às cortes na época da mídia de massa.

Cabe aos gestores da plataforma garantir que nenhum conteúdo seja removido sem que antes seja avaliado por uma quantidade suficiente de

usuários – mais do que um, pois se presume que uma decisão tomada por várias pessoas é melhor do que uma decisão tomada por apenas algumas (Ladha, 1992; Austen-Smith; Banks, 1996). De qualquer forma, pesquisas acerca da qualidade dos resultados da moderação de conteúdo *online* demonstram que a qualidade está mais frequentemente relacionada com a eficácia de ferramentas de coordenação coletiva do que com o número geral de editores (Kittur, 2008).

As plataformas deveriam, em qualquer caso, buscar por usuários moderadores atuantes com diferentes *backgrounds*, evitando assim “filtros bolha” (Flaxman; Goel; Rao, 2016) e “radicalização” de certas posições (Sunstein, 2009), o que é especialmente característico das *fake news*. As iniciativas de publicidade personalizada, já praticadas por essas empresas, permitem, por exemplo, que possam escolher usuários de diferentes áreas de formação – médicos, por exemplo – e inclinações políticas – nem só com valores progressistas e nem só com valores conservadores – para avaliar postagens marcadas como *fake news*, tanto no intuito de restringir a sua divulgação para uma audiência ampliada quanto no intuito de evitar uma avaliação enviesada sob o ponto de vista cultural. Da mesma forma, empresas de redes sociais deveriam também garantir que casos de notícias falsas (Klein; Wueller, 2017) não sejam avaliados apenas com base na opinião de um grupo demográfico específico, e, em vez disso, deveriam impulsionar um debate interideológico que, na internet, pode ser menor do que gostaríamos, mas não é inexistente (Hargittai, 2008).

Para além disso, algoritmos que influenciam em elementos de moderação de conteúdo que foram automatizados precisam ser desenvolvidos e aprimorados para reduzir vieses (Davidson, 2017), mas também para permanecerem responsáveis, ou seja, passíveis de *accountability* (Annany; Crawford, 2013; Diakopoulos, 2016). A identificação e classificação de *fake news* em redes sociais feita por aplicações de inteligência artificial, mais do que em outros campos, precisa produzir evidências acerca dos elementos sobre os quais foi conferido maior peso em determinado processo decisório (Desai; Kroll, 2017) de modo a tornar a decisão explicável.

Iniciativas normativas no sentido de impedir que as plataformas assumam o papel de tomadores de decisão direta automatizada sobre o mérito de cada postagem produzida por seus usuários seriam um passo no sentido de desconstituir, ao menos no campo da moderação das *fake news* em situações de pandemia ou emergência social, o cenário precisamente descrito por Frank Pasquale (2015), como o de uma “sociedade caixa preta”.

## 1.4 PODER EXECUTIVO

A existência de um papel para a Administração Pública no combate das *fake news* é altamente contraintuitivo. Isso parece estar em desacordo com a sabedoria tradicional das democracias modernas de que o Poder Executivo não deve estar de forma alguma ligado à avaliação das manifestações dos cidadãos. No entanto, “a proteção dos valores da liberdade de expressão na era digital será [...] cada vez mais um problema de regulação tecnológica e administrativa” (Balkin, 2009, p. 441).

Certamente, o governo deve permanecer proibido de exercer moderação de conteúdo baseada no mérito do discurso, de modo a poder arbitrar casos individuais de *fake news* sobre a pandemia. No entanto, como demonstramos, essa não é a única peça do quebra-cabeças que requer uma ação pública e privada. Enquanto o Legislativo estabelece as diretrizes procedimentais gerais para os sistemas de autorregulação do conteúdo em que as plataformas privadas e os usuários operam, e o Judiciário avalia até que ponto as plataformas respeitam essas diretrizes, a Administração deve desempenhar um papel intermediário (Balkin, 2018). O seu papel é o de auditar permanentemente esses sistemas de governança privada baseados em *crowdsourcing*, com a finalidade de fiscalizar elementos procedimentais consignados como escolhas de arquitetura, que regulam *ex ante* (Lessig, 1996).

O código fonte não pode ser totalmente aberto sempre, de modo que uma fiscalização técnica é necessária para assegurar que as plataformas privadas não estão se utilizando do código para escapar da regulação (Wu, 2003). A Administração é, portanto, o melhor ator para fiscalizar o uso de inteligência artificial na moderação de conteúdo em geral e inclusive das *fake news*, tendo em vista que as Cortes são geralmente incapazes de reunir todas as evidências e decidir sobre temas de vieses algorítmicos (Grimmelmann, 2014), tarefa para a qual alguns já argumentam ser necessária a criação de uma agência regulatória autônoma (Tutt, 2017).

Esta seção colacionou o estado da arte de estudos científicos sobre a complexa questão da moderação de conteúdo *online*, especialmente as *fake news*. Em suma, a delimitação dos contornos da liberdade de expressão *online* atualmente demanda: *i*) uma reformulação dos papéis das instituições estatais com relação à garantia da liberdade de expressão; *ii*) o estímulo à participação ativa dos usuários no processo de identificação e supressão de *fake news*; e *iii*) a imposição de deveres procedimentais a serem seguidos pelas plataformas digitais enquanto administradoras desses sistemas de mo-

deração descentralizada. A investigação empírica apresentada na seção seguinte tem o objetivo de identificar se algumas dessas diretrizes amparadas pelas pesquisas científicas fazem parte das abordagens consideradas pelo Poder Legislativo brasileiro na regulamentação do tema.

## 2 METODOLOGIA

A pesquisa empírica parte deste estudo consistiu em levantamento de dados nos *sites* da Câmara dos Deputados e do Senado Federal a partir das seguintes palavras-chave: “provedores de aplicação”, “altera o Marco Civil da Internet”, “Marco Civil da Internet” e “Lei 12.965”. A partir dos resultados apresentados pelo mecanismo de busca, foi feita uma análise individual de cada projeto, no intuito de selecionar aqueles que efetivamente se enquadrassem nos parâmetros da pesquisa, qual seja: *i*) projetos que visassem alterar o Marco Civil da Internet (Lei nº 12.965) e que *ii*) dispusessem sobre deveres e/ou responsabilidades de provedores de aplicação.

Os resultados foram compilados em uma planilha, utilizada tanto para tabulação quanto para a geração de análises estatísticas apresentadas nos resultados. A partir dessa avaliação quantitativa dos dados, foi criada uma tipologia de seis grupos para enquadrar os projetos segundo suas principais características. O levantamento dos dados foi realizado durante o mês de março de 2020, sendo a última busca realizada em 31 de março de 2020, de modo que a pesquisa abrange os projetos de lei propostos desde o ano de 2014 até a referida data. O ano de 2014 foi escolhido como data inicial da pesquisa por ser o ano de promulgação do Marco Civil da Internet.

## 3 RESULTADOS

A partir das buscas realizadas nos sites da Câmara e do Senado, foram encontrados 49 projetos de lei ao todo, com a lista completa e seus respectivos detalhes disponibilizada pelos autores desse estudo na internet<sup>16</sup>. O maior número de propostas se concentra no ano de 2019, sinal de que, recentemente, tem aumentado a preocupação do Congresso sobre essa matéria, como se pode verificar na figura 1.

---

16 A lista de projetos encontrados e suas principais informações podem ser acessados nesta tabela: <[https://docs.google.com/spreadsheets/d/e/2PACX-1vRou4cg0-oeaWZmBmYTWDoUApeCvHkrpgNChILz0DsWRIQq7sRe4hCn8k\\_jqdRLluitrffdmjUH8s8t/pubhtml](https://docs.google.com/spreadsheets/d/e/2PACX-1vRou4cg0-oeaWZmBmYTWDoUApeCvHkrpgNChILz0DsWRIQq7sRe4hCn8k_jqdRLluitrffdmjUH8s8t/pubhtml)>.

### PROPOSTAS DE ALTERAÇÃO DO MCI NO QUE CON CERNE A PROVIDORES DE APLICAÇÃO A CADA ANO

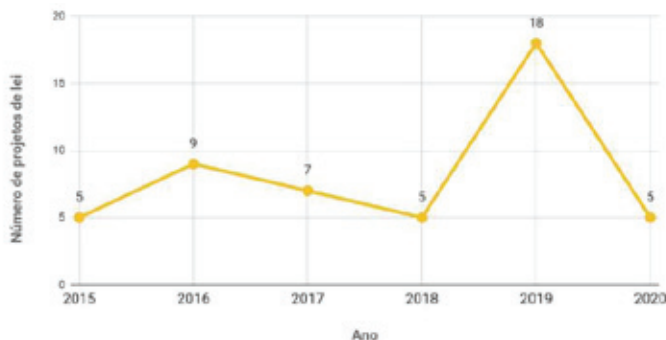


Figura 1

De modo geral, os assuntos tratados pelas propostas encontram-se distribuídos em dois grandes temas: liberdade de expressão e dados pessoais. A figura 2 indica os principais assuntos identificados dentro destes temas, bem como o número de projetos de lei que tratam de cada um deles.

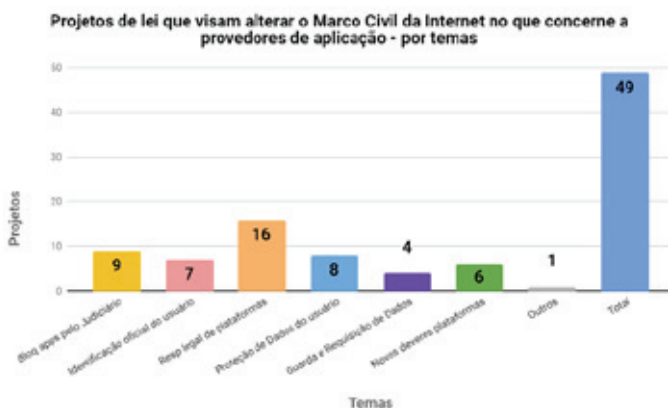


Figura 2<sup>17</sup>

17 Conforme indicado pelo gráfico, foram identificados 49 projetos de lei acerca dos deveres e responsabilidades dos provedores de aplicação, os quais foram divididos, por aproximação temática, em seis assuntos principais: *i)* 9 (nove) projetos de lei buscam restringir o bloqueio de aplicativos pelo Poder Judiciário; *ii)* 7 (sete) projetos estabelecem a obrigatoriedade de identificação oficial do usuário da plataforma; *iii)* 16 (dezesesseis) projetos modificam o regime legal de responsabilização de plataformas por conteúdos de terceiros, atualmente adotado pelo Marco Civil; *iv)* 8 (oito) projetos versam sobre proteção de dados dos usuários; *v)* 4 (quatro) projetos tratam dos deveres de guarda e requisição de dados pelas plataformas; *vi)* 6 (seis) projetos elencam novos tipos de deveres aos provedores de aplicação, que não constam na redação original do Marco Civil, restando, por fim, um projeto que não se encaixou em nenhuma das categorias citadas.

No tema de liberdade de expressão, a maioria dos projetos tem como foco: *i)* impedir ou restringir o bloqueio de aplicações por parte do Poder Judiciário; *ii)* vedar o anonimato nas redes, estabelecendo a obrigatoriedade de as plataformas exigirem o cadastramento de documento oficial de identidade dos usuários; e *iii)* modificar o regime de responsabilização das plataformas por conteúdos gerados pelos usuários.

O segundo grande grupo de propostas trata da coleta e tratamento de dados pessoais dos usuários, bem como das regras de guarda e requisição dos registros de acesso. Além destes, existe um terceiro grupo de projetos que busca estabelecer novos deveres aos provedores de aplicação, que não constam na redação original do Marco Civil. Com relação aos dados do gráfico, é importante esclarecer que, como há proposições que tratam de mais de um assunto, a soma dos números referentes a cada área temática é maior que o número total de 49 projetos.

Analisando cada um dos assuntos citados, verifica-se que a primeira pauta, sobre o bloqueio de aplicações pelo Poder Judiciário, é reflexo de uma série de episódios de bloqueio do WhatsApp, que foram determinados judicialmente, ao longo dos anos de 2015 e 2016. Os bloqueios, que atingiram todos os usuários da ferramenta no País, eram adotados como medida coercitiva para que as plataformas revelassem à Justiça o conteúdo das conversas e os dados de determinados usuários sob investigação criminal. Verifica-se que as proposições visando impedir o bloqueio de aplicativos (9 PLs) se concentraram no ano de 2016, momento em que veio à tona este debate, e compõem aproximadamente 18% dos projetos.

Mais recentemente, o Congresso tem se preocupado com a questão do anonimato nas redes e, a partir de 2019, cresceu o número de projetos que exigem das plataformas a identificação de seus usuários através de documento oficial, como RG ou CPF. Assim, 14% dos PLs destinados a alterar o Marco Civil da Internet investem em medidas de autenticação do usuário de aplicações (7 PLs), alguns deles, inclusive, estabelecendo sanções às plataformas que descumprirem essa exigência. Os principais argumentos apresentados como justificativa dessas proposições são os de que *i)* a Constituição assegura a liberdade de manifestação do pensamento, sendo vedado o anonimato (art. 5º, IV) e que *ii)* a medida de identificação visa facilitar eventual responsabilização de usuários que vierem a praticar atos ilícitos.

No que se refere à responsabilidade das plataformas pelos conteúdos postados por seus usuários, o Marco Civil adota o modelo de responsabilização apenas após ordem judicial, determinando que os provedores de aplicação só podem ser responsabilizados caso, após receberem notificação judicial específica determinando a remoção do conteúdo, não tomem as providências para tornar indisponível o material apontado como ilícito pelo Poder Judiciário. A lei, no entanto, estabeleceu exceções para casos envolvendo direito autoral (art. 19, § 2º) e a divulgação não consensual de imagens íntimas (art. 21), categoria conhecida como pornografia de vingança (Hartmann, 2018b).

Essa é a temática que mais recebeu propostas de alteração legislativa desde a aprovação do MCI até os tempos atuais. Cerca de 33% dos projetos (16 PLs) propõem, em alguma medida, alterar o regime de responsabilização das plataformas – seja eliminando por completo a dinâmica de notificação judicial estabelecida pelo art. 19, seja inserindo novas exceções a este regime, além dos casos de direito autoral e pornografia de vingança, já indicados no art. 19, § 2º e no art. 21. Importante salientar, inclusive, que algumas dessas propostas (PL 7.458/2017, PL 6.989/2017, PL 7.460/2017 e PL 7.918/2017) tramitam em regime de urgência na Câmara, isto é, foram incluídas automaticamente na Ordem do Dia para discussão e votação imediata em plenário, nos termos do art. 155 do Regimento Interno da Casa, por terem sido consideradas matérias de relevante e inadiável interesse nacional. De modo geral, as principais preocupações dos parlamentares veiculadas nos projetos são a propagação de notícias falsas, de atentados terroristas e de conteúdos ofensivos, que instiguem lesão corporal ou suicídio.

Com base nisso, os principais argumentos apresentados como justificativa para mudanças no modelo de responsabilização são os de que situações como essas demandariam uma resposta mais rápida por parte das plataformas para a remoção de conteúdos ilícitos e que, tendo em vista a rapidez da disseminação de informações *online*, o processo judicial seria burocrático e moroso para garantir a proteção das vítimas a tempo. No entanto, esses projetos parecem ignorar outras maneiras de promover *accountability* das plataformas, como, por exemplo, caminhos que as tornem responsáveis não pelos conteúdos disseminados por terceiros, mas pelos procedimentos e pela infraestrutura que administram, conforme pontuado na revisão de literatura objeto da primeira seção deste estudo.

O segundo grande tema endereçado pelos PLs se refere aos dados pessoais dos usuários. A maioria dos projetos nessa temática trata de assuntos relativos à *i*) coleta e tratamento de dados pessoais operado pelas plataformas, correspondendo a 16% do total das propostas (8 PLs) e ao *ii*) dever de guarda que as plataformas possuem sobre os registros de acesso dos usuários, bem como as regras de disponibilização desses registros a terceiros.

Pela redação atual do Marco Civil, a plataforma somente será obrigada a disponibilizar os registros de conexão e de acesso, bem como outras informações que possam contribuir para a identificação do usuário ou do terminal, mediante ordem judicial. Quanto a esse assunto, verificou-se que 8% dos projetos analisados (4 PLs) visam eliminar a exclusividade de ordem judicial para acesso a esses dados pessoais, permitindo que a autoridade policial também possa solicitar essas informações, as quais a plataforma deverá oferecer, independente da análise do Poder Judiciário. Ainda nessa temática, 4% das propostas (2 PLs) têm como objetivo estender o prazo a ser obedecido pelas plataformas para a guarda desses registros. A principal justificativa indicada nessas proposições é a necessidade de tornar mais célere o processo e de garantir o sucesso da investigação criminal.

Um último grupo de projetos (6 PLs) propõe a adoção de novos deveres pelas plataformas, que não constam na redação original do Marco Civil. Dentre eles, encontram-se desde determinações para que sejam criados centros de atenção aos usuários compulsivos de serviços de internet e de redes sociais a determinações para que os sítios de busca divulguem fotos de pessoas desaparecidas. Essas abordagens representam 12% das propostas de alteração ao MCI.

Dentre elas, destaca-se o PL 533/2018, ao estabelecer deveres procedimentais a serem adotados pelos provedores de aplicação, no intuito de prevenir a disseminação de informações inverídicas. Em vez de investir na responsabilização das plataformas pelo conteúdo postado pelos usuários, como é o caso da maioria dos projetos, esse PL adota caminho diverso, ao estabelecer deveres como:

- [i] a adoção de] medidas efetivas e transparentes para combater a publicação e a disseminação de perfis e notícias falsos [...];
- [ii] a disponibilização de] funcionalidade de fácil acesso que permita ao usuário avaliar o grau de confiabilidade das notícias acessadas e denunciar os conteúdos disponibilizados [...];
- [iii] tornar disponível e facilitar o acesso aos critérios utilizados para

identificação, bloqueio e remoção de notícias falsas [...]; [iv] encaminhar ao órgão competente, na forma de regulamentação, relatórios que demonstrem o grau de efetividade das medidas adotadas [...]. (Brasil, 2018)<sup>18</sup>

Conforme já mencionado, encontra-se em tramitação no Congresso, quando da elaboração do presente estudo, o PL 2.630/2020, ou “PL das *Fake News*”<sup>19</sup>, destinado a tratar o problema da desinformação *online*, que tem atraído a atenção do debate público e acadêmico. O PL é motivado em parte pela urgência de combater as *fake news* ligadas à Covid-19. Ainda que contenha algumas inconsistências merecedoras de críticas, pode-se dizer que ele se diferencia positivamente das outras dezenas de projetos analisados por este trabalho por trazer propostas antes não veiculadas no debate legislativo brasileiro e perfeitamente aderentes às recomendações dos estudos científicos analisados na primeira seção.

A proposta investe no estabelecimento de diretrizes procedimentais e recomendações de transparência acerca do processo de moderação de conteúdo empreendido pelas plataformas<sup>20</sup>. Alguns dos pontos positivos presentes na redação do PL que foi aprovada no Senado e enviada à Câmara dos Deputados no final de junho de 2020: *i*) o estabelecimento de mecanismos básicos de devido processo para a moderação de conteúdo, indicados no art. 12 da proposta, como, por exemplo: *a*) a notificação e exposição de motivos ao usuário autor de *post* removido ou conta punida e *b*) a possibilidade de interposição de recurso sobre tais decisões tomadas pelas plataformas; *ii*) a produção de relatórios de transparência determinada no art. 13 da proposta com informações sobre: *a*) os critérios utilizados na moderação do conteúdo e os *b*) dados sobre os conteúdos denunciados e removidos; *iii*) a criação do Conselho de Transparência e Responsabilidade na Internet

---

18 BRASIL. Projeto de Lei nº 533/2018, art. 5º, p. 4. Disponível em: <<https://legis.senado.leg.br/sdleg-getter/documento?dm=7896618&ts=1582134007061&disposition=inline>>.

19 A proposta foi protocolada no dia 1º de abril de 2020, tanto na Câmara dos Deputados como no Senado Federal, correspondendo aos Projetos de Lei nºs 1.429/2020 (novo 2.927/2020), de autoria dos Deputados Felipe Rigoni e Tábata Amaral, e 1.358/2020 (novo 2.630/2020), de autoria do Senador Alessandro Vieira – ambos, com igual redação. Observa-se que esses PLs foram protocolados um dia após o levantamento final de dados realizado por esta pesquisa. A proposta veiculada por esses documentos foi aprovada pelo Senado Federal sob o nº 2630/2020 e encaminhada para tramitação na Câmara dos Deputados. Para mais informações, ver: <<https://www25.senado.leg.br/web/atividade/materias/-/materia/141944>>.

20 Essa foi a inovação mais relevante trazida pela legislação alemã de 2017 sobre o tema, a *Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken* (NetzDG) ou Lei para Aprimoramento da Aplicação do Direito em Redes Sociais. Para além de dispositivos de responsabilidade civil por danos a terceiros, a norma criou obrigações de transparência para as plataformas, exigindo que publiquem relatórios periódicos com números sobre requisições de remoção feitas pelos usuários que se sentem ofendidos e os totais de remoções efetivamente realizadas pela empresa.

indicada no art. 25 da proposta, com o papel de fiscalizar o cumprimento de regras de procedimento e transparência por parte das redes sociais.

Iniciativas como essas buscam tornar as plataformas responsáveis pelos serviços e pela infraestrutura que administram e, mesmo que esse tipo de abordagem ainda seja incipiente no debate legislativo brasileiro, é interessante verificar que ao menos dois projetos em um universo de 50 propostas<sup>21</sup> tenham optado por essa alternativa regulatória.

#### 4 DISCUSSÃO DOS RESULTADOS E CONCLUSÕES

É lamentável que uma parcela tão grande das propostas de lei em tramitação adote o caminho da repressão mais rigorosa das *fake news* e do conteúdo abusivo em redes sociais, seja forçando pessoas a entregar dados pessoais para poderem expressar sua opinião, seja tornando mais frágil a posição das plataformas no que diz respeito à responsabilidade civil, incentivando assim um número maior de manifestações censuradas sem apreciação do Judiciário. Especialmente no período da pandemia e outros tempos de emergência social, essa censura centralizada privada abrangeria *posts* sobre Covid-19, mas obviamente também sobre os mais diversos temas.

Os projetos em sua maioria aumentariam as chances de que uma rede social preferisse apagar todo e qualquer *post* sobre coronavírus de modo a evitar o risco de um processo judicial e uma indenização com valor difícil de prever. Nesse cenário, seriam removidas manifestações talvez ilegítimas como aquelas que veiculam números errados sobre a dispersão e mortalidade do vírus. Mas também poderiam ser censurados *posts* nos quais as pessoas criticam a postura de seu governo estadual, do governo federal ou da Organização Mundial da Saúde diante da pandemia. As propostas em tramitação em sua maioria criariam um ambiente hostil para discussões complexas sobre os formatos possíveis de distanciamento social ou as previsões de duração da fase mais mortífera da Covid-19 – questões sobre as quais não há uma única resposta correta.

O movimento plural de defesa e discussão do que mais tarde se tornou o Marco Civil da Internet iniciou justamente na oposição a projetos

---

21 Trata-se de 50 projetos de lei, contando com o PL 2.630/2020, que, em um primeiro momento, não havia sido contabilizado no universo dos 49 projetos de lei analisados, por ter sido protocolizado em momento posterior ao levantamento inicial.

de lei que inaugurariam a regulação da internet brasileira pela via da criminalização de condutas. A força motriz das discussões do Marco Civil foi a noção de que o primeiro tratamento legislativo da questão deveria ter o objetivo de estabelecer direitos básicos – inclusive de liberdades individuais contra o Estado.

Ao investigar o perfil e padrão das propostas que tramitam no parlamento brasileiro, este estudo realizou uma reflexão sobre que novos caminhos regulatórios são necessários para lidar com os atuais desafios da internet brasileira no contexto da disseminação de *fake news* sobre questões de saúde pública. Quaisquer que sejam as alternativas de atualização do Marco Civil ou de novos diplomas legais adotadas pelo legislador, é preciso que os projetos sejam avaliados segundo as recomendações de uma ampla literatura sobre moderação de conteúdo e *fake news online*, que aponta já diretrizes muito claras sobre quais caminhos são viáveis, quais são promissores e quais resultariam em censura excessiva sem qualquer resultado colateral positivo.

## REFERÊNCIAS

ANNANY, Mike; CRAWFORD, Kate. Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*, 2013. Disponível em: <<https://journals.sagepub.com/doi/abs/10.1177/1461444816676645>>.

AUSTEN-SMITH, David; BANKS, Jeffrey S. Information aggregation, rationality, and the condorcet jury theorem. *American Political Science Review*, v. 90, n. 1, 1996. Disponível em: <<https://authors.library.caltech.edu/67312/1/2082796.pdf>>.

BAKER, C. Edwin. *Media, markets, and democracy: communication, society and politics*. Cambridge University Press, 2001. Disponível em: <[encurtador.com.br/lrtB4](http://encurtador.com.br/lrtB4)>.

BALKIN, Jack. The future of free expression in a digital age. *Pepperdine Law Review*, v. 36, 2009. Disponível em: <[encurtador.com.br/iDT29](http://encurtador.com.br/iDT29)>.

BALKIN, Jack. Free speech in the algorithmic society: big data, private governance, and new school speech regulation. *UC Davis Law Review*, v. 51, 2018. Disponível em: <[encurtador.com.br/fiq47](http://encurtador.com.br/fiq47)>.

BARCELLOS, Ana Paula de. Intimidade e pessoas notórias. Liberdades de expressão e de informação e biografias. Conflito entre direitos fundamentais. Ponderação, caso concreto e acesso à justiça. Tutelas específica e indenizatória. *Revista Direito Público*, n. 5, 2014. Disponível em: <<https://www.portaldeperiodicos.idp.edu.br/direitopublico/article/view/2372>>.

BARRETT, Paul. Who moderates the social media giants? A call to end outsourcing. NYU Stern. Center for Business and Human Rights, 2020.

BARROSO, Luís Roberto. Colisão entre liberdade de expressão e direitos da personalidade. Critérios de ponderação. Interpretação constitucionalmente adequada do Código Civil e da Lei de Imprensa. *Revista de Direito Administrativo*, n. 235, jan./mar. 2004. Disponível em: <<https://bibliotecadigital.fgv.br/ojs/index.php/rda/article/view/45123>>.

BARZILAI-NAHON, Karine. Toward a theory of network gatekeeping: a framework for exploring information control. *Journal of The American Society For Information Science and Technology*, n. 59, 2008. Disponível em: <[encurtador.com.br/cmEM2](http://encurtador.com.br/cmEM2)>.

BAVEL, J. J. V.; BAICKER, K.; BOGGIO, P. S. et al. Using social and behavioural science to support Covid-19 pandemic response. *Nature Human Behavior*, v. 4, 460-471, 2020.

BELLI, Luca. A heterostakeholder cooperation for sustainable internet policymaking. *Internet Policy Review*, n. 4, 2015. Disponível em: <[encurtador.com.br/huDHS](http://encurtador.com.br/huDHS)>.

BONELLI, Maria Gloria. Profissionalismo, gênero e significados da diferença entre juízes e juízas estaduais e federais. *Contemporânea – Revista de Sociologia da UFSCar*, v. 1, p. 103-123, 2011. Disponível em: <[encurtador.com.br/beqv7](http://encurtador.com.br/beqv7)>.

BROWN, Ian; MARSDEN, Christopher T. *Regulating Code*. Good governance and better regulation in the information age. MIT Press, 2013. Disponível em: <[encurtador.com.br/beisz](http://encurtador.com.br/beisz)>.

CAPLAN, Robyn. Content or context moderation. Artisanal, community-reliant, and industrial approaches. *Data & Society*, New York, nov. 2018. Disponível em: <<https://datasociety.net/library/content-or-context-moderation/>>.

CITRON, Danielle Keats. Technological due process. *Wash. U. L. Rev.*, v. 85, 2008. Disponível em: <[encurtador.com.br/ipJL3](http://encurtador.com.br/ipJL3)>.

DAVIDSON, Thomas et al. Automated hate speech detection and the problem of offensive language. *Proceedings of The 11th International AAAI Conference on Web and Social Media (ICWSM-17)*, 2017. Disponível em: <<https://arxiv.org/abs/1703.04009>>.

DESAI, Deven; KROLL, Joshua. Trust but verify: a guide to algorithms and the law. *Harvard Journal of Law and Technology*, v. 31, no. 1, 2017. Disponível em: <<https://jolt.law.harvard.edu/assets/articlePDFs/v31/31HarvJLTech1.pdf>>.

DENARDIS, Laura. Hidden levers of internet control. *Information, Communication & Society*, v. 15, n. 5, p. 720-738, 2012. Disponível em: <[encurtador.com.br/isUXY](http://encurtador.com.br/isUXY)>.

DIAKOPOULOS, Nicholas. Accountability in algorithmic decision making. *Communications of the ACM*, v. 59, No. 2, 2016. Disponível em: <<https://dl.acm.org/doi/10.1145/2844110>>.

FLAXMAN, Seth; GOEL, Sharad; RAO, Justin M. Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly*, v. 80, n. S1, 2016. Disponível em: <<https://5harad.com/papers/bubbles.pdf>>.

GOLDMAN, Alvin; BAKER, Daniel. Free speech, fake news, and democracy. *First Amendment Law Review*, v. 18, 2019.

GOOGLE. Transparency Report, 2018. Disponível em: <[encurtador.com.br/etET0](http://encurtador.com.br/etET0)>.

GRAHAM, Mark; STRAUMANN, Ralph K.; HOGAN, Bernie. Digital divisions of labor and informational magnetism: mapping participation in Wikipedia. *Annals of the Association of American Geographers*, v. 105, n. 6, 2015. Disponível em: <[encurtador.com.br/ouLT6](http://encurtador.com.br/ouLT6)>.

GRIMMELMANN, James. Speech engines. *Minnesota Law Review*, n. 98, 2014. Disponível em: <<https://james.grimmelman.net/files/articles/speech-engines.pdf>>.

\_\_\_\_\_. The virtues of moderation. *Yale J.L. & Tech*, 2015. Disponível em: <<https://scholarship.law.cornell.edu/facpub/1486/>>.

HARGITAI, Eszter et al. Cross-ideological discussions among conservative and liberal bloggers. *Public Choice*, n. 134, p. 67-86, 2008. Disponível em: <[encurtador.com.br/fwAER](http://encurtador.com.br/fwAER)>.

HARTMANN, Ivar A. Let the users be the filter? Crowdsourced filtering to avoid online intermediary liability. *Journal of the Oxford Centre for Socio-Legal Studies*, v. 2017, n. 1, 2017. Disponível em: <[encurtador.com.br/hjowR](http://encurtador.com.br/hjowR)>.

\_\_\_\_\_. Liberdade de expressão e capacidade comunicativa. Um novo critério para resolver conflitos entre direitos fundamentais informacionais. *Direitos Fundamentais e Justiça*, v. 12, n. 39, 2018.

\_\_\_\_\_. Regulação da internet e novos desafios da proteção de direitos constitucionais: o caso do *revenge porn*. *Revista de Informação Legislativa*, v. 55, n. 219, 2018b.

HUMPHREYS, Sal. Ruling the virtual world. Governance in massively multiplayer online games. *European Journal of Cultural Studies*, n. 11, 2008. Disponível em: <<https://journals.sagepub.com/doi/abs/10.1177/1367549407088329>>.

ISENSEE, Josef; KIRCHHOF, Paul. *Handbuch des Staatsrechts*. Band IV – Freiheitsrechte. Heidelberg: C.F. Müller Juristischer Verlag, 1989. Disponível em: <[encurtador.com.br/moAU4](http://encurtador.com.br/moAU4)>.

JHAVER, S.; BRUCKMAN, A.; GILBERT, E. Does transparency in moderation really matter?: user behavior after content removal explanations on Reddit. *Proceedings of the ACM on Human-Computer Interaction*, 2019. Disponível em: <<https://dl.acm.org/doi/10.1145/3359252>>.

KAYE, David. *Speech Police: the global struggle to govern the internet*. Columbia Global Reports, 2019. Disponível em: <[encurtador.com.br/vFLQV](http://encurtador.com.br/vFLQV)>.

KITTUR, Aniket; KRAUT, Robert E. Harnessing the Wisdom of Crowds in Wikipedia: quality through coordination. *CSCW'08*, 8-12 nov., San Diego, California, EUA, 2008.

KLEIN, David O.; WUELLER, Joshua R. Fake news: a legal perspective. *Journal of Internet Law*, v. 20, n. 10, 2017. Disponível em: <[encurtador.com.br/bMOP8](http://encurtador.com.br/bMOP8)>.

KLEINWÄCHTER, Wolfgang. Internet co-governance. Towards a multilayer multiplayer mechanism of consultation, coordination and cooperation (M3C3). In: MANSELL, Robin (Org.). *The Information Society*, v. III (Democracy, governance and regulation). New York: Routledge, 2009. Disponível em: <[encurtador.com.br/gluxY](http://encurtador.com.br/gluxY)>.

KLONICK, Kate. The new governors: the people, rules, and processes governing online speech. *Harvard Law Review*, n. 131, 2018. Disponível em: <[encurtador.com.br/svKPW](http://encurtador.com.br/svKPW)>.

LADHA, Krishna K. The condorcet jury theorem, free speech, and correlated votes. *American Journal of Political Science*, v. 36, n. 3, 1992. Disponível em: <<https://www.jstor.org/stable/2111584?seq=1>>.

LESSIG, Lawrence. The zones of cyberspace. *Stanford Law Review*, v. 48, 1996. Disponível em: <<https://www.jstor.org/stable/1229391>>.

MUELLER, Milton et al. The internet and global governance: principles and norms for a new regime. *Global Governance*, n. 13, 2007. Disponível em: <[encurtador.com.br/afgwF](http://encurtador.com.br/afgwF)>.

OGUS, Anthony. Rethinking self-regulation. *Oxford Journal of Legal Studies*, v. 15, 1995. Disponível em: <[encurtador.com.br/pyLSY](http://encurtador.com.br/pyLSY)>.

PASQUALE, Frank. *The Black Box Society*, Harvard University Press, 2015. Disponível em: <<https://www.hup.harvard.edu/catalog.php?isbn=9780674970847>>.

\_\_\_\_\_. Platform neutrality: enhancing freedom of expression in spheres of private power. *Theoretical Inquiries L.*, v. 17, 2016. Disponível em: <[encurtador.com.br/avAFV](http://encurtador.com.br/avAFV)>.

RISCH, Michael. Virtual rule of law. *West Virginia Law Review*, n. 112, 2009. Disponível em: <[https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1463583](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1463583)>.

ROBERTS, Sarah. *Behind the screen: content moderation in the shadows of social media*. Yale University Press, 2019. Disponível em: <[encurtador.com.br/iyLO5](http://encurtador.com.br/iyLO5)>.

SARLET, Ingo; ROBL FILHO, Ilton. Estado Democrático de Direito e os limites da liberdade de expressão na Constituição Federal de 1988, com destaque para o problema da sua colisão com outros direitos fundamentais, em especial, com os direitos de personalidade. *Revista da Academia Brasileira de Direito Constitucional*, v. 08, n. 14, Constituição, Economia e Desenvolvimento, 2016. Disponível em: <<http://abdconst.com.br/revista15/estadollton.pdf>>.

SARLET, Ingo; HARTMANN, Ivar. Direitos fundamentais e direito privado: a proteção da liberdade de expressão nas mídias sociais. *Revista Direito Público*, v. 16, n. 90, 2019.

SHAW, Aaron. Centralized and decentralized gatekeeping in an open online collective. *Politics & Society*, n. 40, 2012. Disponível em: <[encurtador.com.br/ekoLQ](http://encurtador.com.br/ekoLQ)>.

STARCK, Christian. *Kommentar zum Grundgesetz*. Band I. Band 1, Präambel, Artikel 1 bis 19. Munique: Franz Vahlen GmbH, 2005. Disponível em: <[encurtador.com.br/uAGJW](http://encurtador.com.br/uAGJW)>.

STERN, Klaus. *Das Staatsrecht Der Bundesrepublik Deutschland*. Band IV/1. Die einzelnen Grundrechte. Munique: C.H. Beck, 2006. Disponível em: <[encurtador.com.br/dhsQX](http://encurtador.com.br/dhsQX)>.

STIGLITZ, Joseph. Regulation and failure. In: MOSS, David A.; CISTERNINO, John A. *New perspectives on regulation*. Cambridge: The Tobin Project, 2009. Disponível em: <<https://www.tobinproject.org/books-papers/new-perspectives-regulation>>.

SUNSTEIN, Cass R. *Republic.com 2.0*. Princeton University Press, 2009. Disponível em: <<https://www.jstor.org/stable/j.ctt7tbsw>>.

SUZOR, N.; VAN GEELLEN, T.; MYERS WEST, S. Evaluating the legitimacy of platform governance: a review of research and a shared research agenda. *International Communication Gazette*, v. 80, n. 4, p. 385-400, 2018. Disponível em: <[encurtador.com.br/bpxS8](http://encurtador.com.br/bpxS8)>.

TUFEKCI, Zeynep. Algorithmic harms beyond Facebook and Google: emergent challenges of computational agency. *Colo. Tech. L.J.*, v. 13, n. 203, 2015. Disponível em: <<https://ctlj.colorado.edu/wp-content/uploads/2015/08/Tufekci-final.pdf>>.

TUSHNET, Rebecca. Power without responsibility: intermediaries and the first amendment. *The George Washington Law Review*, v. 76, 2008. Disponível em: <[encurtador.com.br/wAOS9](http://encurtador.com.br/wAOS9)>.

TUTT, Andrew. An FDA for algorithms. *Administrative Law Review*, 69, v. 1, 2017. Disponível em: <[encurtador.com.br/gfFJR](http://encurtador.com.br/gfFJR)>.

WU, Tim. When code isn't law. *Virginia Law Review*, n. 89, 2003. Disponível em: <[encurtador.com.br/cRVW2](http://encurtador.com.br/cRVW2)>.

**Sobre o autor e a autora:**

**Ivar Alberto Hartmann** | *E-mail:* ivarhartmann@gmail.com

Doutorado em Direito Público pela UERJ, Mestre em Direito Público pela PUCRS, Mestre em Direito (LL.M.) pela Harvard Law School, Professor e Coordenador do Centro de Tecnologia e Sociedade da FGV Direito Rio, Professor do Programa de Pós-Graduação em Direito da Regulação (Mestrado e Doutorado) da FGV Direito Rio.

**Julia Iunes Monteiro** | *E-mail:* juliaiunes@gmail.com

Pesquisadora do Centro de Tecnologia e Sociedade da FGV Direito Rio no Projeto Moderação de Conteúdo Online, também desenvolve pesquisas na área de Democracia Digital e Governo Aberto, Doutoranda em Teoria do Estado e Direito Constitucional pela Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio), Mestre em Direito Público pela Universidade do Estado do Rio de Janeiro (UERJ), Graduada pela Faculdade Nacional de Direito da Universidade Federal do Rio de Janeiro (FND-UFRJ), Advogada.

Data de submissão: 31 de julho de 2020.

Data do aceite: 4 de novembro de 2020.