

Mídias sociais, regulação processual e dever de cuidado: evidências empíricas do debate no Brasil

Ivar Hartmann, Ramon Costa, Bruno Oliveira

02-12-2024

Resumo: Este artigo investiga a regulação de plataformas digitais no Brasil, com foco na discussão do dever de cuidado nesse contexto. O objetivo é analisar como essa abordagem é apresentada e debatida na consulta pública sobre regulação de plataformas promovida pela CGI.br e na audiência pública do STF sobre a constitucionalidade do art. 19 do Marco Civil da Internet. A metodologia empregada é a análise documental qualitativa das contribuições nesses fóruns, buscando referências ao "dever de cuidado". Além disso, há uma revisão bibliográfica comparando a abordagem centrada no mérito do conteúdo com a abordagem centrada no mérito do procedimento. Os resultados demonstram o estado incipiente do debate sobre o dever de cuidado em plataformas e a importância de um processo regulatório para plataformas digitais que equilibre a liberdade de expressão com a prevenção de danos e a proteção do usuário.

Palavras-chave: mídias sociais; liberdade de expressão; dever de cuidado; governança de plataformas; regulamentação processual

Introdução

É um fato bem conhecido que as plataformas que operam em mercados bilaterais (Parker et al., 2016) regulam o comportamento dos usuários em ambos os lados com um grau de poder e eficácia comparável à ação governamental (Belli e Zingales, 2017). No caso de plataformas de expressão, como as empresas de mídia social, o principal comportamento que regulam é a expressão. Essa ação regulatória das plataformas pode, à primeira vista, parecer semelhante à autorregulação tradicional, um modelo que existe há décadas (Ogus, 1995). No entanto, difere porque as empresas de mídia social não impõem limites à sua própria ação, mas sim às ações de seus usuários em ambos os lados do mercado (Hartmann, 2022).

A expressão online é um exemplo perfeito de comportamento humano regulado por quatro elementos distintos: normas sociais, a lei, o mercado e a arquitetura (Lessig, 1998). O principal incentivo das empresas de redes sociais para censurar a expressão, contudo, é fomentar um ambiente agradável e familiar, propício ao aumento da receita publicitária (mercado) (Cohen, 2017; Klonick, 2017). Isso leva a casos de censura excessiva. Há muitos exemplos notórios: o Facebook, em determinado momento, removeu fotos de beijos gays (Hudson, 2011); plataformas costumavam remover fotos de amamentação (Locatelli, 2017); a foto da "Menina do Napalm", uma obra de arte, também foi censurada (Ibrahim, 2017). A sociedade civil reagiu e garantiu revisões nas regras de moderação ou em sua aplicação. Isso demonstra que a moderação pode falhar em ambos os casos: por ser insuficiente ou excessiva. Atualmente, a tendência é de pressão da sociedade civil por uma moderação mais rigorosa.

Essa é a realidade no Brasil de 2024. Assolado por notícias falsas (Ribeiro e Ortellado, 2018) e discursos de ódio político (Silva, 2020; Ozawa et al., 2023) (MCI) (Lemos e Souza, 2016), o Congresso aprovou o projeto de lei multissetorial (Magalhães, 2015), rejeitando a prática de notificação e remoção de conteúdo devido aos seus efeitos

inibidores perversos (Seltzer, 2010; Zingales, 2015). O Supremo Tribunal Federal (STF) também demonstrou disposição para reprimir as mídias sociais (Da Ros e Taylor, 2022) e começou a analisar se o art. 19 é uma solução constitucional para equilibrar a liberdade de expressão do usuário e os direitos das vítimas de conteúdo ilegal. Um processo judicial, em vez de um diálogo multissetorial que desencadeasse ação legislativa, coroou anos de desenvolvimentos no debate sobre moderação de conteúdo online no Brasil.

Em determinado momento nos últimos anos, o projeto de lei 2630/20, apelidado de "projeto de lei das notícias falsas", mobilizou atenção e negociações como uma via para uma regulamentação sofisticada. Alguns anos depois, em vez de um esforço para explorar os benefícios e desafios de múltiplas estratégias possíveis para a regulamentação de plataformas digitais, o país se depara com uma escolha superficial, exercida por um tribunal, entre notificação judicial e remoção de conteúdo e a notificação e remoção de conteúdo à moda antiga.

Existem duas desvantagens. Uma delas é o custo de retornar a um padrão de responsabilidade que induz à censura. Dados sobre remoção de conteúdo e valor de mercado antes e depois da MCI mostram que voltar ao sistema de notificação e remoção agora significaria uma perda anual de bem-estar do consumidor de até US\$ 740 milhões (Hartmann et al, 2023). A segunda desvantagem é menos óbvia e mais difícil de compreender. Trata-se do impacto negativo na qualidade do debate sobre a responsabilidade nas redes sociais e as soluções desejáveis.

O debate e a literatura sobre liberdade de expressão online no Brasil nunca foram particularmente ricos, frequentemente focando nas vantagens e desvantagens de diferentes padrões de responsabilidade. Por exemplo, a grande maioria dos projetos de lei propostos no Congresso entre 2014 e 2020 sobre responsabilização nas redes sociais adotou a resposta simplista de simplesmente substituir o art. 19 por um padrão de responsabilidade mais rigoroso (Hartmann e Iunes, 2020).

A abordagem do dever de cuidado foi raramente mencionada. O conceito está longe de ser estranho ao direito brasileiro. As obrigações do dever de cuidado derivam de interpretações de leis brasileiras, como o Estatuto da Criança e do Adolescente (ECA), o Código Civil, o Código do Consumidor e até mesmo de disposições constitucionais (Costa, 2008; Castro, 2017). O ordenamento jurídico brasileiro já estabelece um marco de deveres para a proteção de grupos vulneráveis, de modo que as plataformas também estão sujeitas ao cumprimento do dever de cuidado devido à natureza e à operação de seus negócios (Frazão, 2024). Especialmente após o fracasso do projeto de lei 2630/20, com a atenção voltada para o julgamento do STF, a regulamentação com base no mérito do conteúdo monopolizou o debate. Isso contrasta fortemente com a regulamentação com base no mérito do procedimento, como previsto pelos Princípios de Santa Clara ¹.

Os Princípios de Santa Clara são cruciais neste contexto, pois estabelecem diretrizes para auxiliar na regulamentação da moderação de conteúdo por plataformas digitais. Esses princípios foram desenvolvidos por um grupo de acadêmicos, especialistas e defensores dos direitos digitais em 2018, com o objetivo de proporcionar maior transparência e equidade nos processos de moderação de conteúdo. Assim, existe uma relação relevante entre os Princípios de Santa Clara e a discussão sobre a regulamentação e a responsabilização das plataformas digitais pelo conteúdo. Isso porque os princípios são guiados por aspectos como transparência, liberdade de

¹ Disponível em: <https://santaclaraprinciples.org/>

expressão e proteção dos direitos do usuário.

Este artigo apresenta os fundamentos da regulação com base no mérito processual e, em seguida, descreve como a sociedade civil tem inserido essa estratégia regulatória no debate. Realizamos uma análise qualitativa das discussões em dois fóruns-chave: a consulta pública sobre plataformas digitais organizada pelo Comitê Diretivo da Internet e a consulta pública sobre o art. 19 e sua constitucionalidade, promovida pelo STF. Diante da diversidade da regulação com base no mérito processual, apresentada na primeira seção deste artigo, optamos por focar no dever de cuidado para compreender como ele é apresentado, construído e argumentado nessas duas arenas cruciais para a participação da sociedade civil.

A pesquisa faz parte do projeto "Regulamentação de plataformas no Brasil e no Reino Unido: concepção e aplicação de marcos de 'dever de cuidado'", conduzido por meio de uma parceria entre a Universidade de Sussex, no Reino Unido, e o Insper, no Brasil, financiado pela Academia Britânica, com o apoio do Fundo de Parcerias Científicas Internacionais do governo britânico. Esta pesquisa teve como objetivo explorar a experiência do Reino Unido com a Lei de Segurança Online de 2023. O projeto produzirá evidências para subsidiar a abordagem brasileira à regulamentação de plataformas, especialmente considerando a falta de discussão sobre o princípio do dever de cuidado no desenvolvimento e implementação de um marco regulatório para plataformas digitais.

O artigo está estruturado da seguinte forma. Primeiramente, descrevemos a metodologia de pesquisa empregada para desvendar empiricamente o perfil do debate sobre o dever de cuidado como uma alternativa viável para a regulação de plataformas. Em seguida, apresentamos uma revisão da literatura sobre o contraste entre a regulação das mídias sociais com base nos méritos do conteúdo (em grande parte, padrões de responsabilidade) e com base nos méritos das garantias processuais, que refletem um dever de cuidado das plataformas em relação à liberdade de expressão dos usuários. Por fim, apresentamos e discutimos os resultados de nossa análise qualitativa.

Metodologia

A metodologia deste artigo consiste em uma análise documental qualitativa das contribuições para a Consulta sobre a Regulação de Plataformas Digitais promovida pela CGI.br e para a audiência pública do STF sobre o Tema 987, que aborda a constitucionalidade do artigo 19 da Lei nº 12.965/2014 (Quadro Civil Brasileiro da Internet). Este artigo estabelece a necessidade de uma ordem judicial prévia e específica para a remoção de conteúdo, a fim de que provedores de serviços de Internet, websites e administradores de aplicativos de redes sociais sejam responsabilizados civilmente por danos decorrentes de atos ilícitos cometidos por terceiros.

Para atingir esse objetivo, utiliza-se a técnica de análise de comentários públicos em ambos os fóruns, com foco no tema do dever de cuidado. A audiência pública realizada pelo STF em 2023 para fundamentar sua decisão sobre o art. 19 contou com 47 participantes. Para identificar as declarações referentes ao dever de cuidado, buscamos o termo "*cuidado*" no documento contendo as transcrições da audiência (STF, 2023). Após identificar o uso dessa palavra, verificamos se ela foi mencionada em referência ao conceito de dever de cuidado. Em caso afirmativo, anotamos qual participante utilizou esse conceito e separamos sua apresentação para leitura e análise posteriores. Ao final desse processo, identificamos 12 participantes que se referiram ao "dever de cuidado".

A consulta pública sobre a regulamentação das plataformas digitais, organizada em 2023 pelo Comitê Diretivo da Internet do Brasil, recebeu 1.306 comentários. Utilizando um arquivo csv contendo todas as contribuições (CGI.br, s.d.), realizamos uma busca automatizada com código Python por comentários contendo as palavras “dever” e “cuidado”, permitindo qualquer combinação de caracteres entre esses dois termos. Essa busca retornou doze comentários enviados por oito colaboradores diferentes. Uma busca similar pelos termos “dever” e “cuidado” em inglês não apresentou resultados. Após identificar aqueles que mencionaram o dever de cuidado em cada um desses fóruns, procedemos à sumarização de suas contribuições. Em seguida, organizamos e sistematizamos as contribuições, que são apresentadas na seção de resultados.

Esta análise e categorização de discursos em dois importantes espaços de debate público sobre plataformas digitais visa destacar como um potencial regime de dever de cuidado está sendo concebido para um processo regulatório no Brasil. Cabe ressaltar que os fóruns possuem responsabilidades e procedimentos distintos, o que permite um escopo analítico mais amplo sobre como diferentes setores da sociedade civil estão se posicionando. As análises também possuem uma conexão especial com a revisão bibliográfica apresentada neste artigo, que detalha os procedimentos de moderação de conteúdo e a presença do dever de cuidado em um contexto regulatório para esses procedimentos, incluindo os direitos dos usuários e as responsabilidades das plataformas na mitigação e prevenção de riscos e danos causados por conteúdo de terceiros.

Regulação processual das redes sociais

O âmbito da regulação pode ser o mérito do conteúdo ou o mérito dos procedimentos de moderação. Este último é norteado por uma preocupação com o devido processo legal (Citron, 2008). O primeiro é problemático, especialmente pela falta de precisão.

A qualidade da análise é insuficiente e os danos colaterais são consideráveis devido às abordagens adotadas pela legislação da grande maioria dos países em relação a quem decide sobre o mérito de casos específicos de discurso. Os governos depositaram toda a sua confiança em decisões privadas automatizadas (Wu, 2019; OFCOM, 2019) ou em decisões tomadas por profissionais não oficiais (Barrett, 2020; Roberts, 2019), bem como em decisões públicas (judiciais) (Epstein, Parker e Segal, 2013; Keller, 2020) sujeitas a erros. O resultado até agora é que essas decisões pressionam ainda mais as plataformas contra a responsabilidade legal, enquanto a percepção da sociedade civil é de que a disseminação de desinformação (Wardle e Derakhshan, 2017) e discurso de ódio só aumentou.

Dever de cuidado

O escopo da autorregulação da liberdade de expressão pode se restringir aos méritos do procedimento de moderação de conteúdo de uma plataforma. A supervisão regulatória pode se abster de focar nos méritos de uma publicação específica e, em vez disso, avaliar a qualidade dos processos implementados para sinalizar, detectar, revisar, tomar decisões e recorrer de decisões sobre o conteúdo do usuário. Enquanto muitos países continuam pressionando as empresas por resultados perfeitos de remoção de conteúdo, eles subestimam o custo para a liberdade de expressão decorrente da imprecisão

inerente aos processos de moderação das plataformas.

Se o conteúdo ilegal for mantido em níveis suficientemente baixos, os legisladores não se sentirão inclinados a questionar o impacto na saúde dos moderadores humanos, o nível de precisão da inteligência artificial que toma ou auxilia nas decisões de moderação, ou a transparência de todo o sistema para os usuários. O devido processo legal na moderação de conteúdo deve ser reconhecido como um fim em si mesmo, e não como um luxo.

Abandonar a avaliação baseada no resultado do procedimento de moderação de uma plataforma em um caso específico resolve muitos problemas das alternativas tradicionais para analisar o mérito da expressão. Os tribunais não seriam sobrecarregados com processos sobre publicações individuais, e suas análises sobre o que a empresa deveria ter feito em um caso específico teriam pouca ou nenhuma importância. Os litígios teriam que questionar as falhas processuais gerais com base em evidências de múltiplos casos. A responsabilidade não dependeria unicamente do frágil julgamento de um tribunal sobre a legalidade de uma expressão isolada. Se uma empresa de mídia social cumpre as regras sobre o procedimento de moderação, ela não pode ser punida por causa de um único caso que passou despercebido e resultou em censura excessiva ou insuficiente.

As vantagens de abandonar as obrigações de resultado e focar no dever de cuidado, criando diretrizes processuais obrigatórias, são talvez mais visíveis no caso da moderação automatizada – algo que os governos têm incentivado recentemente as plataformas a aprimorar (Gorwa, Binns e Katzenbach, 2020). Atualmente, os governos analisam apenas uma pequena fração do que a caixa-preta do algoritmo de moderação produz (Tufekci, 2015). Não há uma forma de mensurar o desempenho do software com base em um conjunto de decisões representativo do todo. Uma decisão automatizada que não detectou discurso de ódio pode ser a exceção ou a regra, mas os formuladores de políticas, os usuários e terceiros afetados não têm a menor ideia. A responsabilidade pelo resultado em casos de moderação selecionados arbitrariamente não oferece incentivo para que as plataformas melhorem a precisão de seus aplicativos de inteligência artificial.

O desalinhamento fundamental de incentivos prejudica, portanto, os usuários, mas também as plataformas. A empresa não pode ser responsabilizada por decisões específicas emitidas por inteligência artificial desenvolvida propositalmente para operar de forma autônoma, decidindo sobre novas formas de expressão que os engenheiros poderiam ter previsto. A responsabilidade objetiva pelos resultados da moderação automatizada é equivocada e catastrófica para a inovação (Thierer, 2016). como responsabilidade objetiva para veículos autônomos².

Em vez de impor um dever de resultado – zero conteúdo ilegal ou zero censura ilegal –, os governos podem definir diretrizes básicas para o procedimento de moderação e aplicar deveres de cuidado. Este é o modelo proposto pelo Livro Branco sobre Danos Online do governo do Reino Unido, baseado na constatação de que um dever de resultado aliado à responsabilização se mostrou uma solução ineficiente: “o foco na responsabilização pela presença de conteúdo ilegal não incentiva as melhorias sistêmicas nos processos de governança e gestão de riscos que consideramos necessárias”. O Livro Branco detalha um modelo com deveres de cuidado sob supervisão governamental.

² No mínimo, deveria haver um sistema diferenciado de responsabilidade que proteja as empresas que adotaram proativamente as melhores práticas recomendadas (Scherer, 2016).

Uma forma alternativa de responsabilização, no âmbito do dever de cuidado, poderia permitir que usuários e outras partes lesadas ainda responsabilizassem as empresas de mídia social, mas somente se os demandantes fossem além da discussão de uma única publicação e apresentassem alegações procedentes sobre as inadequações do procedimento como um todo. Essa distinção tem raízes no direito privado francês, que separa a *obrigação de meio* da *obrigação de resultado*. A *obrigação de meio* também é descrita como uma obrigação processual, segundo a qual uma empresa está vinculada a um plano de vigilância que estabeleceu e pode ser acusada de não cumpri-lo (Cossart, 2017).

Neste contexto, a autorregulação regulamentada significa que as empresas de redes sociais têm alguma, mas não total, independência para definir os detalhes de seus procedimentos. Apresentaremos alguns elementos do devido processo legal para a moderação de conteúdo, mas já existem propostas na literatura, especialmente no que diz respeito à transparência. Os tribunais ou mesmo a Administração Federal de um determinado país não devem estabelecer esses elementos. É papel inescapável dos legisladores definir as diretrizes processuais gerais para a moderação. Um ponto de partida notavelmente avançado para legisladores em qualquer lugar do mundo é o conjunto de garantias processuais dos Princípios de Santa Clara.

Sem dúvida, uma abordagem baseada no dever de cuidado inclui responsabilizar os intermediários pelo conteúdo gerado por terceiros em suas plataformas. Ao passar de uma análise caso a caso para uma análise sistêmica, os governos poderiam aumentar a obrigação das plataformas de monitorar, identificar e remover ou mitigar conteúdo prejudicial, mesmo que gerado por terceiros. A implementação desse dever de cuidado varia de acordo com a jurisdição, mas, em geral, os intermediários devem tomar medidas razoáveis para evitar danos aos usuários (Machado; Aguiar, 2023).

Essa abordagem também deve considerar a necessidade de equilibrar a proteção dos direitos dos usuários com a preservação da liberdade de expressão. Há preocupação com o risco de censura caso as plataformas adotem uma abordagem excessivamente cautelosa na moderação de conteúdo, resultando na supressão indevida da liberdade de expressão. Portanto, o dever de cuidado deve ser equilibrado para preservar os direitos e mitigar danos e riscos, sem conceder carta branca para decisões de moderação de conteúdo que careçam de critérios objetivos e juridicamente sólidos.

Garantias processuais como deveres de cuidado

Transparência

A transparência das políticas substantivas e processuais da plataforma, bem como de sua aplicação efetiva, é fundamental. A legislação alemã já estabeleceu um conjunto claro de requisitos de acesso à informação (Wagner et al., 2020), que estão atualmente servindo de base para projetos de lei em outros países, como o Brasil.

A primeira dimensão da transparência é a informação sobre o todo. O primeiro dos três Princípios de Santa Clara, “números”, refere-se à transparência abrangente sobre a moderação efetivamente realizada, a fim de informar não apenas os usuários da rede social, mas também a sociedade civil em geral e as autoridades públicas. De acordo com esse princípio, a empresa é obrigada a fornecer relatórios periódicos em formato legível por máquina com o número total de decisões de moderação. Os relatórios devem detalhar as estatísticas descritivas de acordo com quatro variáveis: objeto da decisão

(Postagem ou usuário), tipo de decisão (denúncia ou remoção), origem da denúncia (decisão automatizada, governo, usuários etc.) e localização geográfica (do usuário que fez a denúncia e das partes afetadas).

O segundo princípio de Santa Clara também contém uma obrigação sobre moderação em abstrato: as regras substantivas e processuais adotadas pela plataforma devem ser acessíveis a todos. Nos dias de hoje, isso pode parecer óbvio, mas não faz muito tempo, nem mesmo as maiores redes sociais forneciam suas políticas de conteúdo aos seus próprios usuários.

A segunda dimensão da transparência diz respeito à informação sobre a moderação no caso específico. O segundo princípio de Santa Clara trata da exigência de informações mais detalhadas para o usuário afetado, cuja expressão foi restringida. As plataformas devem informar o autor da publicação sobre o motivo específico da remoção, assim como um tribunal é obrigado a fornecer o fundamento jurídico de sua decisão. Elas também devem oferecer detalhes sobre o processo pelo qual a publicação foi identificada e considerada em conflito com a política de conteúdo. O usuário tem o direito de saber, por exemplo, se sua expressão foi restringida por um humano, por uma decisão automatizada ou por uma combinação de ambos. Além disso, a empresa também deve instruir o usuário sobre suas opções para questionar a decisão.

Os princípios de Santa Clara não contêm diretrizes específicas sobre a responsabilização ou explicabilidade da moderação algorítmica. A OCDE foi pioneira na elaboração de um conjunto de princípios sobre inteligência artificial que devem orientar os legisladores na definição das obrigações processuais das redes sociais em relação aos seus sistemas automatizados de moderação ³. Ao longo dos anos, surgiram diversas listas de princípios e recomendações para o uso ético da inteligência artificial (Fjeld et al., 2020).

Princípios e recomendações éticas por si só, no entanto, não são suficientes. O uso da IA é generalizado em todo o mundo, em diferentes mercados, tanto por governos quanto por empresas privadas. Há ampla evidência de falhas que afetam desproporcionalmente as minorias. Os estudos em ciência da computação e engenharia oferecem conhecimento suficiente sobre o processo para permitir que os formuladores de políticas comecem a elaborar regras concretas, especialmente em setores onde o dano é irreparável (Black e Murray, 2019). Um requisito processual óbvio de transparência na moderação automatizada é a elaboração proativa de dados sobre a precisão. As plataformas devem começar configurando testes para verificar a precisão de seus modelos de aprendizado de máquina na identificação de diferentes tipos de conteúdo e anunciando periodicamente os índices de precisão (Davidson et al., 2017).

Recursos

O terceiro princípio de Santa Clara estabeleceu a garantia fundamental do devido processo legal, ou seja, o direito de recorrer da remoção de conteúdo ou de uma punição aplicada à conta do usuário. A revisão da decisão de moderação de conteúdo deve ser feita por pessoas, após o usuário ter tido a oportunidade de apresentar sua defesa. Assim como na decisão original, a revisão precisa fornecer ao usuário a fundamentação

³ Os princípios de IA da OCDE mais úteis no contexto da moderação de conteúdo são: “Deve haver transparência e divulgação responsável em relação aos sistemas de IA para garantir que as pessoas compreendam os resultados baseados em IA e possam contestá-los” e “As organizações e os indivíduos que desenvolvem, implementam ou operam sistemas de IA devem ser responsabilizados pelo seu funcionamento adequado, em conformidade com os princípios acima”. Disponível em: <https://oecd.ai/en/ai-principles>.

política que a sustenta.

Assim como a transparência em geral, isso pode significar mais informações sobre as estratégias de moderação da plataforma para aqueles que persistentemente tentam burlá-las, como terroristas. No entanto, esse é um preço razoável a se pagar por uma redução da arbitrariedade da plataforma. Da mesma forma que a liberdade de expressão é crucial justamente para os discursos com os quais mais discordamos, o devido processo legal é fundamental, sobretudo, para aqueles que detestamos, como terroristas e supremacistas brancos. Uma garantia de apelação não está de forma alguma relacionada a uma avaliação dos méritos daquele caso específico de discurso e não deve ser aplicável apenas a expressões online que a sociedade civil considera menos nocivas. Versões anteriores do projeto de lei 2630/20 incluíam essa diretriz processual como uma obrigação para as empresas de mídia social que operam no país.

Os tribunais como executores das garantias processuais da liberdade de expressão

Regras processuais para moderação de conteúdo podem ser o novo porto seguro para empresas de mídia social. Em vez de serem estritamente responsáveis, responsáveis após notificação ou após notificação judicial, as plataformas deveriam ser responsabilizadas apenas quando não cumprirem as diretrizes processuais consagradas em lei. Tais diretrizes exigirão regulamentação adicional por um órgão público independente com poder para consolidá-las, definindo regras e padrões para o procedimento, mas nunca para o mérito (Reino Unido, 2019). Essa autoridade poderia decidir, por exemplo, que discursos de ódio devem ser submetidos à revisão de cinco usuários e notícias falsas apenas a três. Mas jamais poderia, em hipótese alguma, arbitrar o mérito de uma publicação específica (Hartmann, 2020).

O novo papel dos tribunais seria o de apontar omissões importantes do legislador quanto às garantias processuais necessárias que devem ser concedidas aos usuários. Os tribunais também poderiam exercer uma supervisão limitada do trabalho da autoridade independente na definição de padrões para os procedimentos de moderação e, eventualmente, na aplicação de multas. No entanto, o Judiciário gradualmente deixaria de tomar decisões sobre o mérito de casos específicos de discurso. Uma possível transição seria a adoção inicial de um sistema de licenciamento, no qual a plataforma que cumpre todas as normas sobre os procedimentos de moderação estaria sujeita a um padrão de responsabilidade diferente daquele da empresa que não cumpriu ou simplesmente optou por não aderir. Essa é uma solução atualmente em discussão para a regulamentação de aplicações de inteligência artificial, como veículos autônomos (Tutt, 2017).

É evidente que diferentes tipos de danos online exigem soluções específicas. Regular os procedimentos de moderação de conteúdo é mais eficaz do que apostar em padrões de responsabilidade para moderação de conteúdo baseados em uma análise centralizada do mérito da expressão. Um país pode preferir uma regulamentação mais rigorosa de notícias falsas do que de difamação, mas, em ambos os casos, o governo não deve impor às redes sociais uma obrigação de apresentar resultados. Os esforços de moderação de pornografia infantil e pornografia de vingança de uma empresa certamente devem estar sujeitos a requisitos mais rigorosos, mas também não podem ser considerados totalmente inadequados e passíveis de multas simplesmente porque uma entre milhares de publicações prejudiciais não é detectada e removida.

Dever de cuidado na audiência pública do STF

Mais do que aqueles que mencionaram o conceito de dever de cuidado, chama a atenção aqueles que não o fizeram: representantes das principais redes sociais. Representantes do Facebook, Google, Twitter, Bytedance (TikTok Brasil) e Mercado Livre participaram da audiência pública, e nenhum deles discutiu o dever de cuidado. Por outro lado, o conceito foi mencionado por diversos órgãos governamentais ⁴, organizações da sociedade civil ⁵ e representantes de associações empresariais de outros setores ligados à Internet ⁶. Esse fato parece indicar uma estratégia das empresas de mídia social para desviar a discussão sobre o Artigo 19 do Marco Civil da Internet no Supremo Tribunal Federal (STF) dos debates sobre a adoção do dever de cuidado ou do entendimento de que esse dever já possa existir de alguma forma no ordenamento jurídico brasileiro. Em contrapartida, diversos outros participantes buscaram defender a linha contrária, argumentando que as plataformas já deveriam aderir a um dever de cuidado.

Flávio Dino, então Ministro da Justiça e Segurança Pública (MJSP), argumentou que as plataformas violavam um dever de cuidado inerente a qualquer atividade econômica, conforme previsto no Código de Defesa do Consumidor. De forma semelhante, Estela Aranha, também falando em nome do MJSP, afirmou que o dever de cuidado era uma premissa fundamental das relações de consumo, enraizada no princípio da boa-fé. Isis Menezes Taboas, representante do Ministério da Mulher, também adotou a visão de que as plataformas já deveriam estar sujeitas a um dever de cuidado. No entanto, argumentou que esse dever decorria dos termos de uso das plataformas, que implicavam uma responsabilidade de supervisão e promoção de um ambiente seguro. As diferenças entre essas declarações, e o fato de outros representantes do governo não terem adotado uma posição semelhante quanto à existência de um dever de cuidado, indicam que não havia uma abordagem governamental coordenada sobre o que deveria ser defendido sobre esse tema.

Alguns representantes de organizações não governamentais também apoiaram interpretações que implicariam o reconhecimento de um dever de cuidado já existente para as plataformas de mídias sociais no Brasil. Isabella Henriques, do Instituto Alana, defendeu uma interpretação sistemática e harmoniosa que reconheça que as normas do Marco Civil da Internet não podem se sobrepor a um dever geral de cuidado com o conteúdo consumido por crianças e adolescentes. Nessa perspectiva, mesmo em relação aos danos causados por conteúdo de terceiros, a interpretação de que a remoção de conteúdo ilícito depende de uma ordem judicial explícita deve ser rejeitada. Em vez disso, o dever de cuidado exigiria uma postura mais proativa, que vá além da mera remoção de conteúdo prejudicial assim que a plataforma tome conhecimento inequívoco dele. Uma vez que o dano está associado ao próprio modelo de negócios da plataforma, o dever de cuidado deveria se estender ao desenvolvimento da plataforma e de seus algoritmos.

Outros foram menos detalhistas, mas ainda reconheceram alguma forma de dever de cuidado já existente. João Quinelato, representando o Instituto Brasileiro de Direito Civil

⁴ Ministério da Justiça e Segurança Pública, Secretaria de Comunicação Social da Presidência da República, Escritório Nacional de Defesa da Democracia da Procuradoria-Geral da República – PNDD/AGU e Ministério da Mulher.

⁵ Comitê de Tecnologia e Inovação – OAB/SP, Instituto Brasileiro de Políticas e Direito do Consumidor, Instituto Alana, Instituto Brasileiro de Direito Civil – IBDCIVIL e Instituto de Tecnologia e Sociedade do Rio – ITS Rio.

⁶ Associação Brasileira de Internet (Abranet) e Associação Brasileira de Rádio e Televisão (Abratel).

(IBDCIVIL), argumentou que a proteção da dignidade humana exige uma regulamentação que imponha responsabilidades proativas aos provedores, presumindo, portanto, deveres de cuidado. Da mesma forma, Aislan Vargas Basílio, da Associação Brasileira de Rádio e Televisão (Abratel), sugeriu similarmente um dever de cuidado baseado no princípio da boa-fé objetiva, como também proposto por Estela Aranha.

As demais declarações referentes ao dever de diligência podem ser divididas em dois grupos: as que defendem a adoção desse conceito e as que o criticam.⁷ No primeiro grupo, João Brant, representando a Secretaria de Comunicação Social da Presidência da República, argumentou que a ausência de um dever de diligência permite que as plataformas digitais permaneçam passivas. Essa escolha regulatória pode favorecer a liberdade de expressão, mas tornou-se evidente que essa abordagem é desequilibrada, visto que a proteção dos direitos autorais é mais forte na Internet do que a proteção contra ataques à democracia. Por essa razão, devem ser buscadas reformas no modelo atual, com cautela em relação aos potenciais efeitos inibidores que algumas alternativas regulatórias podem acarretar.

Marcelo Eugênio Feitosa Almeida, representando o Escritório Nacional de Defesa da Democracia da Procuradoria-Geral da República (PNDD/AGU), afirmou brevemente que a regulação das mídias sociais está passando de um modelo de ação estatal negativa para um modelo focado em ações positivas das plataformas, no qual o dever de diligência se situaria. Por fim, Carlos Affonso Souza, em nome da Associação Brasileira da Internet (Abranet), apresentou o dever de cuidado como uma das alternativas para aprimorar o Artigo 19 do Marco Civil da Internet brasileiro.

Entre os críticos do dever de cuidado, Ronaldo Lemos, falando em nome da Comissão de Tecnologia e Inovação da Ordem dos Advogados do Brasil, Seção São Paulo (OAB/SP), e Fabio Steibel, do Instituto de Tecnologia e Sociedade do Rio (ITS Rio), defenderam maior cautela na adoção de modelos estrangeiros, incluindo o dever de cuidado. Lemos argumentou que o aprimoramento do modelo brasileiro deve ser guiado por experiências bem-sucedidas do passado, como a regulação da pornografia de vingança, onde esse tipo de conteúdo foi definido como uma exceção que justifica o desvio da regra geral do artigo 19. Ele sugeriu a criação de outras exceções claramente definidas. Conceitos abertos como o dever de cuidado, segundo ele, seriam subjetivos e difíceis de aplicar, concedendo discricionariedade excessiva às plataformas e diminuindo o papel do judiciário ao transferi-lo para as corporações. Como modelo alternativo, Lemos propôs a autorregulação regulamentada, com órgãos de supervisão tendo autoridade vinculante para fiscalizar as atividades de moderação da rede e estabelecer regras de transparência, orientando as plataformas para uma melhor tomada de decisão. Steibel expressou preocupações semelhantes, destacando os riscos de conceder ainda mais poder às plataformas e levar à moderação excessiva.

Dever de cuidado na consulta pública do CGI.br

A análise do dever de cuidado no contexto brasileiro da regulação de plataformas envolve essencialmente o debate público, considerando a participação de atores multissetoriais na regulação de plataformas digitais no país. Nesse sentido, por meio do Grupo de Trabalho de Regulação de Plataformas, o CGI.br realizou a “Consulta sobre a

⁷ A declaração de Guilherme Magalhães Martins, em nome do Instituto Brasileiro de Políticas e Direito do Consumidor, menciona o dever de cuidado, mas não aprofunda o tema. Portanto, não está agrupada com nenhum conjunto de declarações.

Regulação de Plataformas Digitais”, com o objetivo de reunir diversas perspectivas sobre a regulação de plataformas. Essa consulta possui significativa relevância e impacto para a sociedade civil, visto que o Comitê tem desempenhado um papel importante em discussões sobre questões relacionadas à Internet, incluindo seu engajamento e participação em processos legislativos como o Marco Civil da Internet e a Lei Geral de Proteção de Dados (LGPD) (Cappi; Oms, 2023).

Todas as menções ao dever de cuidado vieram de participantes associados a organizações especializadas em regulamentação da Internet ⁸ ou com experiência profissional na área ⁹. Essas menções não se concentraram em nenhuma pergunta específica feita na consulta pública. Em vez disso, apareceram em respostas a nove perguntas diferentes, com não mais do que dois comentários mencionando o dever de cuidado em resposta à mesma pergunta.

Embora esta amostra seja pequena em comparação com o total (0,9%), limitando inferências e comparações mais amplas entre grupos de comentários que mencionam ou não este conceito, é importante notar que os comentários que mencionam o dever de cuidado são significativamente mais longos do que a média dos comentários submetidos na consulta. Os comentários que mencionam o dever de cuidado apresentaram uma média de 5.298 caracteres e uma mediana de 2.487 caracteres, enquanto aqueles que não o mencionam apresentaram médias de 1.310 e 590 caracteres, respectivamente.¹⁰ Se a extensão dos comentários puder ser usada como uma métrica para estimar sua complexidade ou completude, esses dados quantitativos indicam uma correlação entre comentários que mencionam o dever de cuidado e comentários mais abrangentes e complexos – possivelmente explicada pelo perfil dos participantes que abordaram esse conceito, como mencionado anteriormente.

Outra descoberta quantitativa que corrobora essa observação é que, com exceção dos dois comentários mais curtos que mencionam o dever de cuidado, os outros dez estão muito próximos¹¹ ou dentro do quartil dos comentários mais longos. Além disso, quatro comentários (um terço dos que mencionam o dever de cuidado) estão entre os 5% de comentários mais longos, incluindo o segundo mais longo no geral.

Em relação ao conteúdo dos comentários, quatro participantes manifestaram apoio à incorporação do dever de cuidado no ordenamento jurídico brasileiro, dois se opuseram e, em dois casos, não ficou claro se os participantes eram a favor ou contra esse modelo de responsabilidade. Por exemplo, Flávio Rech Wagner (id nº 715)¹² manifestou-se

⁸ Alana Institute, Digital Rights Coalition, Electronic Frontier Foundation e Access Now, ABERT - Associação Brasileira de Emissoras de Rádio e Televisão, ITS Rio e CTS-FGV.

⁹ Flávio Rech Wagner é professor do Instituto de Informática da UFRGS. Foi presidente do Capítulo Brasileiro da Internet Society, fundou e presidiu a Sociedade Brasileira de Computação e foi membro do Comitê Diretivo da Internet – CGI.br e NIC.br. Flávia Lefèvre Guimarães foi coordenadora do departamento jurídico do Idec, membro do Conselho Consultivo da Anatel e conselheira do Comitê Diretivo da Internet. Atualmente, trabalha como advogada.

¹⁰ Essa diferença persiste mesmo se excluirmos da comparação os comentários da consulta pública que respondem a outros comentários e, portanto, podem ser, em média, mais curtos, já que tendem a apenas concordar ou complementar brevemente os comentários anteriores. Sem esse subconjunto de comentários, o comprimento médio dos comentários aumentou ligeiramente para 1.358 caracteres e a mediana para 625 caracteres.

¹¹ Como existem 1.306 comentários, o primeiro quartil vai até a posição 326. O décimo comentário mais longo que menciona o dever de cuidado está na posição 327.

¹² Indica o número do comentário no documento da consulta de plataformas do CGI.br que baseou a pesquisa apresentada neste texto: <https://dialogos.cgi.br/documentos/debate/consulta-plataformas/> O documento também está em anexo na publicação no site da poliTICs.

apenas contra a atuação da CGI.br na avaliação de relatórios de conformidade com o dever de cuidado elaborados por plataformas digitais, argumentando que a CGI.br não deveria assumir responsabilidades típicas de órgãos estatais. Essa posição não endossa nem rejeita o dever de cuidado, mas apenas afirma que a CGI.br não deveria ter tal função.

A outra posição ambígua veio de Carlos Affonso Souza, em nome do ITS Rio. Por um lado, o ITS Rio (id nº 1119) argumentou que a adoção de mecanismos de notificação e ação sem responsabilidade direta por conteúdo específico de terceiros poderia levar a ineficiências, potenciais violações da liberdade de expressão e suscetibilidade a abusos por parte dos usuários. Por outro lado, em um comentário separado (id nº 1124) que pressupunha a adoção do dever de cuidado sem criticá-lo explicitamente, o ITS Rio defendeu que as sanções administrativas por descumprimento do dever de cuidado deveriam se concentrar no conteúdo como um sistema para evitar o cerceamento do debate público. Segundo essa visão, seria possível alinhar o atual marco do Marco Civil da Internet com disposições para sanções administrativas por descumprimento dos deveres de risco sistêmico.

Entre os comentários favoráveis à adoção do dever de cuidado, o Centro de Tecnologia e Sociedade da FGV Rio fez a observação mais sucinta, mencionando-o apenas como uma medida de mitigação sugerida para operações de influência em plataformas (id nº 1069). Flávia Lefèvre Guimarães enfatizou que o dever de cuidado não seria tão inovador quanto possa parecer, argumentando que se alinha ao princípio da responsabilidade previsto no Artigo 3º do Marco Civil da Internet e às disposições existentes no Código de Defesa do Consumidor, na Lei Geral de Proteção de Dados, no Código Eleitoral, no Código Civil e no Estatuto da Criança e do Adolescente (id nº 859). Ela defendeu a regulamentação desse dever por meio do Projeto de Lei 2630, destacando sua inspiração na Lei de Serviços Digitais, e sugerindo que o Código de Defesa do Consumidor e o Marco Civil da Internet já poderiam servir de base para uma maior responsabilização.

João Coelho, representando o Instituto Alana, e Rodolfo Salama, em nome da ABERT – Associação Brasileira de Emissoras de Rádio e Televisão – compartilharam opiniões semelhantes às de Flávia Lefèvre Guimarães. A ABERT também apoiou o modelo proposto pelo Projeto de Lei 2630 (id nº 920), elogiando particularmente o Artigo 11 do projeto, que lista os crimes a serem prevenidos e mitigados no âmbito do regime de dever de cuidado. Sugeriu ainda a inclusão de crimes eleitorais nessa lista. O Instituto Alana argumentou que a legislação vigente já impõe um dever de cuidado às plataformas (id nº 805), especialmente no que diz respeito aos direitos de crianças e adolescentes. O instituto incluiu um parecer jurídico da Professora e Advogada Ana Frazão, também citado em outro comentário (id nº 806), que sugeriu a extensão do dever de cuidado quando as plataformas amplificam ou monetizam conteúdo gerado pelo usuário.

Em oposição ao dever de cuidado, Bia Barbosa, representando a Digital Rights Coalition, e Veridiana Alimonti, em nome da Electronic Frontier Foundation e da Access Now, apresentaram dois comentários cada. A Digital Rights Coalition expressou preocupação com o potencial de vigilância decorrente da imposição de um dever de cuidado, particularmente se este se estender a serviços de mensagens privadas (id nº 1267). Em outro comentário (id nº 1328), a coalizão argumentou explicitamente contra a atribuição de um dever de cuidado às plataformas, defendendo que estas não devem julgar o que constitui conteúdo criminoso nem decidir o que deve circular. Contudo, no mesmo comentário, a coalizão propôs um meio-termo: um regime em que as plataformas teriam obrigações de avaliação de risco e ação apenas em casos específicos,

como ameaças graves e iminentes à integridade física, à saúde pública ou à democracia, incluindo pandemias, ataques terroristas ou outros perigos sérios.

De forma semelhante, a Electronic Frontier Foundation e a Access Now defenderam o abandono das obrigações de diligência para mitigar os riscos de abuso em regulamentações baseadas em conteúdo, privilegiando as avaliações de impacto sistêmico (id nº 1125). Em relação à diligência proposta no Projeto de Lei 2630/2020, as organizações criticaram o amplo escopo de atividades ilegais a serem controladas pelas plataformas, abrangendo seis leis e 40 diferentes crimes, muitos dos quais com exceções (id nº 1129). Essa abrangência dificultaria a aplicação da lei, agravada pela falta de clareza sobre como e quais instituições avaliariam a conformidade. Essa regulamentação colocaria as plataformas em um papel judicial, o que elas consideraram inadequado.

Considerações Finais

O caminho a seguir após a decisão do STF sobre o art. 19 e a responsabilidade dos intermediários é incerto. Dependerá de como a responsabilização nas redes sociais se consolidará à medida que os tribunais inferiores aplicarem a decisão do STF e as plataformas adaptarem seus termos e condições, bem como seus sistemas de moderação de conteúdo. Uma coisa permanece certa: simplificar o debate sobre alternativas de responsabilização não beneficia nenhuma das partes envolvidas.

O dever de cuidado e outras abordagens de regulação processual podem coexistir com diferentes padrões de responsabilidade, se necessário. As obrigações legais das plataformas quanto à qualidade dos procedimentos de moderação podem complementar eficazmente até mesmo padrões de responsabilidade mais rigorosos, como a notificação e remoção de conteúdo. A experiência do Reino Unido pode oferecer lições úteis para o Brasil, além de exemplos de advertência. Para distinguir as aplicações improdutivas ou ineficientes do dever de cuidado daquelas que poderiam promover a proteção dos direitos dos usuários no Brasil, é necessário um debate mais aprofundado sobre as características e os desafios dessa abordagem.

Demonstramos que, em dois fóruns cruciais para a participação da sociedade civil na regulamentação das redes sociais, a proporção do debate dedicada ao dever de cuidado foi muito pequena. As raras contribuições, contudo, desempenharam um papel importante ao fornecer mais contexto e nuances às discussões sobre como coibir conteúdo abusivo online, garantindo, ao mesmo tempo, que as plataformas digitais não descartem o essencial junto com o supérfluo.

Nos próximos anos, à medida que a sociedade brasileira honra sua tradição de genuína participação multissetorial na governança da Internet, não deve permitir que uma única decisão judicial monopolize os esforços para encontrar opções viáveis de regulação das redes sociais. Seja por meio do projeto de lei 2630/20 em versão reformulada ou de uma proposta totalmente nova, a sociedade civil deve instar o Congresso a construir uma substituição democraticamente responsável para o art. 19 da MCI.

Referências

- BARRETT, Paul M. Quem modera os gigantes das mídias sociais? Um apelo para acabar com a terceirização. *Relatório NYU Stern*, 2020. Disponível em: <https://www.stern.nyu.edu/experience-stern/faculty-research/who-moderates-social-media-giants-call-end-outsourcing>.
- BELLI, Luca; ZINGALES, Nicolo (orgs.). *Regulamentação de plataformas: como as plataformas são reguladas e como elas nos regulam*. Resultado oficial da Coalizão Dinâmica do Fórum de Governança Intergovernamental da ONU sobre Responsabilidade de Plataformas. FGV Direito Rio, 2017. Disponível em: <https://hdl.handle.net/10438/19402>.
- BLACK, Julia; MURRAY, Andrew. Regulamentação da IA e do aprendizado de máquina: definindo a agenda regulatória. *European Journal of Law and Technology*, v. 10, n. 3, 2019.
- CASTRO, Thamis de. A Função da Cláusula de Bons Costumes no Direito Civil e a Teoria Tríplice da Autonomia Privada Existencial. *Revista Brasileira de Direito Civil*, v. 14, 2017, p. 99-125.
- CITRON, Danielle Keats. Devido Processo Tecnológico. *Washington University Law Review*, v. 85, 2008.
- COHEN, Julie E. Direito para a Economia de Plataformas. *UCD Law Review*, v. 51, 2017, p. 165.
- COSSART, Sandra; et al. A Lei Francesa sobre o Dever de Cuidado: Um Passo Histórico para Fazer a Globalização Funcionar para Todos. *Business and Human Rights Journal*, v. 2, p. 321, 2017.
- COSTA, Judith H. Martins. Um Aspecto Da Obrigação De Indenizar: Notas Para Uma Sistematização Dos Deveres Pré-negociais De Proteção No Direito Civil Brasileiro. *Revista dos Tribunais*, v. 867, 2008, p. 11-51.
- CAPPI, Juliano; OMS, Juliana. Sistematização das contribuições à consulta sobre regulação de plataformas digitais. São Paulo: Núcleo de Informação e Coordenação do Ponto BR, 2023.
- DA ROS, Luciano; TAYLOR, Matthew M. Bolsonaro e o Judiciário: entre acomodação e confronto. In: BIRLE, Peter; SPECK, Bruno (eds.). *Brasil sob Bolsonaro. Quão ameaçada está a democracia?*. Instituto Ibero-Amerikanisches Preußischer Kulturbesitz, 2022. p. 1-20.
- DAVIDSON, Thomas; et al. Detecção automatizada de discurso de ódio e o problema da linguagem ofensiva. In: *Anais da 11ª Conferência Internacional da AAAI sobre Web e Mídias Sociais*, 2017, p. 1.
- DAVIDSON, Thomas; et al. Detecção automatizada de discurso de ódio e o problema da linguagem ofensiva. In: *Anais da 11ª Conferência Internacional da AAAI sobre Web e Mídias Sociais*, 2017.
- DIÁLOGOS CGI.BR. Consulta sobre Regulação de Plataformas Digitais. Disponível em: <https://dialogos.cgi.br/documentos/debate/consulta-plataformas/>.
- EPSTEIN, Lee; PARKER, Christopher M.; SEGAL, Jeffrey A. Os juízes defendem o discurso que odeiam? Viés intragrupal, oportunismo e a Primeira Emenda. In: *Encontro Anual da Associação Americana de Ciência Política*, 2013.
- 'O Dilema das Redes' erra. 2020. Disponível em: <https://about.fb.com/wp-content/uploads/2020/10/What-The-Social-Dilemma-Gets-Wrong.pdf>

FJELD, Jessica; et al. Inteligência Artificial com Princípios: Mapeando o Consenso em Abordagens Éticas e Baseadas em Direitos para Princípios de IA. *Berkman Klein Center*, Publicação de Pesquisa nº 2020-1, 2020.

FRAZÃO, Ana. A responsabilidade civil das plataformas digitais por conteúdos de terceiros diante do artigo 19 do Marco Civil da Internet. In: FRAZÃO, Ana; MULHOLAND, Caitlin; POLIDO, Fabrício Bertini Pasquot (Coords.). *Marco Civil da Internet: impactos, evoluções e perspectivas*. São Paulo: Thomson Reuters Brasil, 2024. p. 17-41.

GORWA, Robert; BINNS, Reuben; KATZENBACH, Christian. Moderação algorítmica de conteúdo: desafios técnicos e políticos na automação da governança de plataformas. *Big Data & Society*, 2020, p. 1-15.

HARTMANN, I.; et al. Análise econômica da responsabilidade das redes sociais. *Economic Analysis of Law Review*, v. 14, n. 2, 2023, pp. 17-41.

HARTMANN, I.; IUNES, J. 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. *Revista Direito Público*, v. 94, 2020, pp.

HARTMANN, Ivar A. Uma nova estrutura para moderação de conteúdo online. *Computer Law & Security Review*, v. 36, 2020.

HARTMANN, Ivar A. Autorregulação em plataformas de conteúdo online e a proteção dos direitos da personalidade. In: ALBERS, Marion; SARLET, Ingo (orgs.). *Proteção dos Direitos da Personalidade e Proteção de Dados Online*. Springer, 2022.

HUDSON, John. A controvérsia sobre a proibição de beijos gays no Facebook não acabou.

The Atlantic, 2011. Disponível em:

<https://www.theatlantic.com/technology/archive/2011/04/controversy-over-facebooks-gay-kissing-ban-isnt-over/349921/>.

IBRAHIM, Yasmin. Facebook e a Garota do Napalm: Reenquadrando o Icônico como Pornográfico. *Social Media + Society*, out-dez 2017, p. 1.

KELLER, Clara Iglesias. Política por judicialização: o marco institucional da responsabilidade dos intermediários no Brasil. *Revista Internacional de Direito, Computadores e Tecnologia*, 2020.

KLONICK, Kate. Os Novos Governadores: As Pessoas, as Regras e os Processos que Regem a Liberdade de Expressão Online. *Harvard Law Review*, v. 131, 2018.

LEMO, Ronaldo; SOUZA, Carlos Affonso. *Marco Civil da Internet – Construção e Aplicação*. Juiz de Fora: Editar, 2016.

LESSIG, Lawrence. A nova escola de Chicago. *The Journal of Legal Studies*, v. 27, n. 2, 1998.

LOCATELLI, Elisabetta. Imagens de amamentação no Instagram: autorrepresentação, publicidade e gestão da privacidade. *Social Media + Society*, abril de 2017.

MACHADO, Caio CV; AGUIAR, Thaís Helena. Regulamentações emergentes sobre moderação de conteúdo e políticas de desinformação em plataformas de mídia online: incorporando o dever de cuidado em modelos de responsabilidade intermediária. *Business and Human Rights Journal*, v. 8, p. 244–251, 2023.

MAGALHÃES, João Carlos. Imaginando criticamente a governança da Internet: uma análise de conteúdo da Consulta Pública do Marco Civil da Internet. *Série de dissertações de mestrado MEDIA@LSE*, 2015.

OFCOM. Uso de IA na moderação de conteúdo online. Cambridge Consultants/OFCOM, 2019. Disponível em: <https://www.ofcom.org.uk/siteassets/resources/documents/research-and-data/online-research/other/cambridge-consultants-ai-content-moderation.pdf?v=324081>.

OFCOM. Uso de IA na moderação de conteúdo online. Relatório, 2019. Disponível em: https://www.ofcom.org.uk/__data/assets/pdf_file/0028/157249/cambridge-consultants-ai-content-moderation.pdf.

OGUS, Anthony. Repensando a autorregulação. *Oxford Journal of Legal Studies*, v. 15, 1995.

OZAWA, João VS; et al. Como a desinformação no WhatsApp passou de arma de campanha a propaganda governamental no Brasil. *Social Media + Society*, v. 9, n. 1, 2023, p. 1-12.

PARKER, Geoffrey; VAN ALSTYNE, Marshall W.; CHOUDARY, Sangeet Paul. *Revolução das Plataformas: Como os Mercados em Rede Estão Transformando a Economia e Como Fazê-los Funcionar para Você*. WW Norton & Company, 2016.

RIBEIRO, Márcio Moretto; ORTELLADO, Pablo. Notícias falsas: o que são e como lidar com elas. *Sur*, v. 15, n. 27, 2018, p. 1-12.

ROBERTS, Sarah T. *Por trás da tela: Moderação de conteúdo nas sombras das mídias sociais*. Yale University Press, 2019.

SCHERER, Matthew U. Regulamentação de Sistemas de Inteligência Artificial: Riscos, Desafios, Competências e Estratégias. *Harvard Journal of Law and Technology*, v. 29, p. 357, 2016.

SELTZER, Wendy. Liberdade de expressão desvinculada da proteção do direito autoral: efeitos inibidores da DMCA sobre a Primeira Emenda. *Harvard Journal of Law and Technology*, v. 24, 2010.

SILVA, Daniel N. A Pragmática do Caos: Analisando a Linguagem Antidemocrática de Bolsonaro. *Trabalhos em Linguística Aplicada*, v. 1, 2020, pág. 507-537.

SUPREMO TRIBUNAL FEDERAL - STF. RE 1.037.396 e RE 1.057.258 – Responsabilidade civil dos provedores por conteúdo ilícito gerado por terceiros. Transcrições da 38ª audiência pública. Ministros-Relatores Dias Toffoli e Luiz Fux. Brasília: Supremo Tribunal Federal/Secretaria Judiciária/Coordenadoria de Processamento Final/Gerência de Transcrição, 2023. Disponível em: https://www.stf.jus.br/arquivo/cms/audienciasPublicas/anexo/RE1037396_RE1057258_Transcricoes_Audiencia_Publica_38.pdf.

THIERER, Adam. Inovação sem permissão e políticas públicas: um plano de 10 pontos. 2016. p. 11. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2761139.

TUFEKCI, Zeynep. Danos algorítmicos além do Facebook e do Google: desafios emergentes da agência computacional. *Colorado Technology Law Journal*, v. 13, 2015, p. 207.

TUTT, Andrew. Uma FDA para algoritmos. *Administrative Law Review*, v. 69, n. 1, 2017.

GOVERNO DO REINO UNIDO. Livro Branco sobre Danos Online. 2019. p. 45-46. Disponível em: <https://www.gov.uk/government/consultations/online-harms-white-paper>.

WAGNER, Ben; et al. Regulamentando a Transparência? Facebook, Twitter e a Lei Alemã de Fiscalização de Redes. In: FAT* '20, 27 a 30 de janeiro, Barcelona, Espanha, 2020, p. 9.

WARDLE, Claire; DERakhshan, Hossein. DESORDEM DA INFORMAÇÃO: Rumo a uma estrutura interdisciplinar para pesquisa e formulação de políticas. *Relatório do Conselho da Europa* , 2017.

WU, Tim. A Inteligência Artificial Vai Devorar o Direito? A Ascensão de Sistemas Híbridos de Ordenação Social. *Columbia Law Review* , v. 119, n. 1, 2019.

ZINGALES, Nicolo. A abordagem brasileira à responsabilidade dos intermediários da Internet: modelo para um regime global? *Internet Policy Review* , v. 4, n. 4, 2015.

IVAR HARTMANN -- Professor associado de direito do Insper. É membro fundador da Platform Governance Research Network e afiliado ao International Panel on the Information Environment. Sua pesquisa foca a moderação de conteúdo online e tecnologia aplicada à pesquisa quantitativa em Direito.

RAMON COSTA -- Professor de Direito do Insper. Doutor em Direito pela Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio). Pesquisador do Núcleo Legalite: Direito e novas tecnologias, da PUC-Rio.

BRUNO OLIVEIRA -- Bacharel em Direito pela Universidade de São Paulo. Mestre em Direito e Desenvolvimento pela Escola de Direito de São Paulo da Fundação Getulio Vargas. Pesquisador do Núcleo de Saúde do Centro de Gestão e Políticas Públicas do Insper.