



A CORREGULAÇÃO COMO ALTERNATIVA À MODERAÇÃO DE CONTEÚDO ON-LINE

THE CO-REGULATION AS AN ALTERNATIVE TO ONLINE CONTENT MODERATION

GUILHERME AMORIM CAMPOS DA SILVA

Doutor em Direito do Estado (2010) e Mestre em Direito Constitucional (2002) pela PUC/SP. Pós-doutor em Direito Econômico, Financeiro e Tributário pela USP (2024). Professor dos cursos de graduação, mestrado e doutorado da Faculdade Autônoma de Direito (Fadisp). Advogado. Bolsista de projeto de pesquisa apoiado pela Funadesp – Fundação Nacional de Desenvolvimento do Ensino Superior Particular. E-mail: guilherme.silva@unialfa.com.br. ORCID: <https://orcid.org/0000-0002-1018-7468>

JOHNNY OLIVEIRA DE PASSOS

Gestor Master da 1ª Unidade de Processamento Judicial (UPJ) das Varas Criminais da comarca de Trindade (GO). Analista Judiciário do Tribunal de Justiça do Estado de Goiás (TJGO). Bacharel em Direito pela Faculdade Santa Rita de Cássia (Unifasc). Especialista em Direito Penal e Processual Penal. Mestrando em Função Social do Direito pela Faculdade Autônoma de Direito (Fadisp) e Escola Judicial do TJGO (Ejug). E-mail: jopassos@tjgo.jus.br.

RESUMO:

O presente artigo investiga a regulação da moderação de conteúdo em plataformas digitais sob a perspectiva do constitucionalismo brasileiro, focalizando a tensão entre a proteção preferencial da liberdade de expressão — consagrada pela Constituição de 1988 — e a necessidade de enfrentamento do discurso de ódio, da desinformação massiva e dos ataques coordenados às instituições democráticas. Parte-se da constatação de que o modelo brasileiro vigente, fundamentado no artigo 19 do Marco Civil da Internet e recentemente reinterpretado pelo Supremo Tribunal Federal nos temas de repercussão geral 987 e 533, revela-se estruturalmente insuficiente para responder aos desafios do ambiente digital contemporâneo, gerando insegurança jurídica, fragmentação normativa e judicialização excessiva. Mediante abordagem jurídico-teórica qualitativa, com análise documental, jurisprudencial e comparativa, examinam-se os principais modelos estrangeiros de regulação de plataformas — norte-americano, europeu e alemão —, evidenciando-se seus fundamentos, virtudes, limitações e os riscos de transplante acrítico para o contexto brasileiro. Demonstra-se que a correção constitui a alternativa regulatória mais adequada à realidade constitucional e institucional nacional, por articular responsabilidades entre Estado, plataformas digitais e sociedade civil mediante princípios estruturantes de transparência, devido processo legal, proporcionalidade e supervisão independente. Conclui-se que a correção oferece superior legitimidade democrática, maior efetividade na proteção de direitos fundamentais e capacidade adaptativa às





transformações tecnológicas e comunicacionais, superando tanto as deficiências da autorregulação privada quanto os excessos da regulação estatal mandatória.

PALAVRAS-CHAVE: liberdade de expressão; moderação de conteúdo; constitucionalismo digital; plataformas digitais; correção.

ABSTRACT:

This article investigates the regulation of content moderation on digital platforms from the perspective of Brazilian constitutionalism, focusing on the tension between the preferential protection of freedom of expression — enshrined in the 1988 Constitution — and the need to address hate speech, mass disinformation, and coordinated attacks on democratic institutions. The analysis begins with the finding that the current Brazilian model, based on Article 19 of the Civil Rights Framework for the Internet (Marco Civil da Internet) and recently reinterpreted by the Federal Supreme Court in general repercussion themes 987 and 533, proves structurally insufficient to respond to the challenges of the contemporary digital environment, generating legal uncertainty, normative fragmentation, and excessive judicialization. Through a qualitative legal-theoretical approach, employing documentary, jurisprudential, and comparative analysis, the article examines the main foreign models of platform regulation — American, European, and German — evidencing their foundations, virtues, limitations, and the risks of uncritical transplantation to the Brazilian context. It is demonstrated that co-regulation constitutes the most adequate regulatory alternative for the national constitutional and institutional reality, as it articulates responsibilities among the State, digital platforms, and civil society through structuring principles of transparency, due process of law, proportionality, and independent supervision. The conclusion establishes that co-regulation offers superior democratic legitimacy, greater effectiveness in protecting fundamental rights, and adaptive capacity to technological and communicational transformations, overcoming both the deficiencies of private self-regulation and the excesses of mandatory state regulation.

KEYWORDS: freedom of expression; content moderation; digital constitutionalism; digital platforms; co-regulation.

1 INTRODUÇÃO

A regulação da moderação de conteúdo digital constitui desafio estrutural às democracias contemporâneas, transcendendo a dimensão técnico-corporativa para configurar-se como problemática de poder, legitimidade e efetivação de direitos fundamentais. Bilhões de indivíduos exercem direitos basilares — liberdade de expressão, acesso à informação, participação política — mediante intermediação de plataformas digitais privadas que estabelecem normas próprias de visibilidade, amplificação e remoção de conteúdo, com impacto direto sobre a esfera pública.





No Brasil, a escalada de conflitos nessa arena intensificou-se nos últimos anos, culminando em decisões do Supremo Tribunal Federal nos temas de repercussão geral 987 e 533, proferidas em junho de 2025, que reconheceram expressamente a inconstitucionalidade parcial do regime vigente, estabelecendo inflexão do debate doutrinário para o reconhecimento judicial da insuficiência do modelo atual.

O problema central consiste em determinar como regular a moderação de conteúdo em plataformas privadas preservando a proteção preferencial da liberdade de expressão consagrada na Constituição de 1988, respondendo simultaneamente aos riscos de discurso de ódio, desinformação massiva e ataques coordenados às instituições democráticas. Três modelos regulatórios competem nesse cenário: a autorregulação privada, centrada no artigo 19 do Marco Civil da Internet; a regulação estatal mandatória, exemplificada pela NetzDG alemã; e a corregulação, regime híbrido articulando Estado, plataformas e sociedade civil mediante transparência, devido processo e mitigação de riscos sistêmicos.

O objetivo geral da investigação consiste em reconstruir o debate sobre moderação de conteúdo digital no Brasil, articulando a proteção preferencial da liberdade de expressão com o enfrentamento do discurso de ódio, desinformação e ataques às instituições democráticas, de modo a justificar a corregulação como modelo adequado ao contexto constitucional brasileiro.

Para tanto, pretende-se analisar a liberdade de expressão no ambiente digital e sua relação com discurso de ódio e desinformação, examinar criticamente o modelo brasileiro vigente fundado no artigo 19 do Marco Civil da Internet e na jurisprudência recente do Supremo Tribunal Federal, comparar modelos regulatórios estrangeiros — notadamente o norte-americano, o europeu e o alemão — e propor modelo de corregulação baseado em transparência, devido processo, proporcionalidade e supervisão independente.

A investigação orienta-se por três hipóteses centrais. A primeira sustenta que a autorregulação concentra poder decisório sem responsabilização adequada, carecendo de garantias processuais e respostas estruturadas a riscos sistêmicos. A segunda postula que a regulação mandatória gera remoção excessiva de conteúdo e transfere decisões penais complexas a atores privados sem garantias processuais adequadas. A terceira defende que a corregulação oferece superior legitimidade democrática, proteção a direitos fundamentais e capacidade adaptativa frente aos desafios tecnológicos e comunicacionais emergentes.





Metodologicamente, adota-se abordagem qualitativa de natureza jurídico-teórica, valendo-se predominantemente de método dedutivo e de técnica de análise documental e comparativa, com base em legislação, jurisprudência e doutrina nacional e estrangeira sobre liberdade de expressão, moderação de conteúdo e constitucionalismo digital.

A justificativa da investigação decorre tanto da gravidade concreta dos problemas quanto de lacunas teóricas e normativas. Episódios recentes de desinformação eleitoral, ataques institucionais coordenados, disseminação de discursos de ódio contra minorias e campanhas anticientíficas evidenciam a insuficiência do arranjo regulatório atual. Ademais, a produção acadêmica nacional carece de estudos integrando dogmática da liberdade de expressão, constitucionalismo digital, governança privada de plataformas, direito comparado e jurisprudência recente do Supremo Tribunal Federal em arcabouço teórico unitário.

Espera-se que a investigação contribua em dois planos distintos. No plano teórico, mediante a construção de leitura crítica e integrada sobre constitucionalismo digital, governança privada e modelos regulatórios comparados, oferecendo subsídios para debates futuros. No plano prático, que a proposta de correção sirva de fundamento para reformas legislativas, decisões judiciais e formulação de políticas públicas voltadas à regulação de plataformas digitais.

2 LIBERDADE DE EXPRESSÃO, DISCURSO DE ÓDIO E DESINFORMAÇÃO NO AMBIENTE VIRTUAL

A liberdade de expressão constitui um dos elementos fundamentais do Estado Democrático de Direito. Quando pessoas e grupos têm direito de emitir opiniões, criticar governantes, divulgar informações e construir sentidos coletivos sobre a realidade, cria-se condição para que o poder público seja fiscalizado e decisões políticas sejam tomadas com base em debate informado. Em termos simples, trata-se do direito de manifestar pensamentos, ideias e informações, sem censura prévia e sem medo de represálias indevidas do Estado ou de atores privados (Barroso, 2025).

No ambiente digital, esse direito assume uma nova dimensão. A esfera pública tradicional, até então composta por ruas, praças e jornais impressos, tem sido parcial e gradativamente substituída por espaços digitais estruturados por algoritmos que





amplificam conteúdos, definem o que é visto, para quem e, por vezes, organizam a própria experiência de realidade dos usuários (Medeiros, 2013). Bilhões de pessoas passaram a participar da conversação pública não mais em espaços públicos fisicamente localizados, mas em plataformas como Instagram, TikTok, X (antigo Twitter), YouTube e aplicativos de mensagens como WhatsApp e Telegram (We Are Social, 2024). Essa transformação trouxe inúmeras vantagens às sociedades em geral: pessoas geograficamente dispersas podem se conectar instantaneamente; movimentos sociais se organizam em alta velocidade e custo reduzido; vozes antes marginalizadas pela mídia tradicional encontram espaço para se expressar (Serra Júnior; Rocha, 2013).

Contudo, ao mesmo tempo em que trouxe benefícios, também permitiu uma série de problemas. O primeiro deles é a desinformação, isto é, conteúdos falsos ou gravemente enganosos, muitas vezes produzidos e disseminados de forma estratégica para influenciar eleições, corroer a confiança em instituições públicas ou comprometer políticas de saúde pública (Torre; Jerónimo, 2023). Ao contrário de situações em que alguém simplesmente erra um dado ou compartilha uma informação equivocada de boa-fé, a desinformação tal como discutida hoje costuma envolver campanhas coordenadas, uso de perfis falsos ou automatizados, e exploração deliberada dos mecanismos de recomendação das plataformas para alcançar milhões de pessoas rapidamente (Torre; Jerónimo, 2023).

Durante a pandemia de Covid-19, por exemplo, campanhas de desinformação sobre eficácia de vacinas foram disseminadas em massa, levando pessoas a tomar decisões que comprometeram sua saúde (Falcão; Souza, 2021). Em eleições, a desinformação foi usada para enfraquecer a confiança em urnas eletrônicas e processos democráticos (Pimenta *et al.*, 2024). Esse tipo de informação falsa estruturada representa uma ameaça diferente de um erro isolado ou de uma opinião controversa legítima.

O segundo problema é o discurso de ódio, entendido como expressão dirigida contra pessoas ou grupos com base em características como raça, etnia, religião, gênero, orientação sexual ou condição de pessoa com deficiência, com o objetivo ou efeito de incitar hostilidade, discriminação ou violência contra esses grupos (Brugger, 2010). A diferença fundamental em relação a uma crítica dura, mas legítima, está na direção contra grupos vulneráveis e no conteúdo que estimula a supressão de direitos, inferiorização ou eliminação física dessas pessoas (Andrade, 2023). É possível fazer





crítica severa a políticas públicas, decisões de governantes e ideologias políticas sem que haja necessariamente discurso de ódio (Andrade, 2023). No entanto, quando, por exemplo, o conteúdo diz “pessoas dessa etnia deveriam ser mortas” ou “mulheres não merecem trabalhar”, está-se claramente em território de discurso que incita violência e discriminação (Andrade, 2023).

Conteúdos como esses colocam em tensão a liberdade de expressão com outros direitos fundamentais, como igualdade, dignidade humana, segurança e participação política em condições de não discriminação. A Constituição de 1988 protege tanto a liberdade de expressão quanto esses outros direitos. Quando entram em conflito, surge a pergunta de como equilibrá-los sem sacrificar um ou outro de forma injusta (Barroso, 2025).

A complexidade inerente ao ambiente virtual motivou o debate doutrinário sobre “constitucionalismo digital”, expressão que designa o esforço de adaptar normas do constitucionalismo contemporâneo ao contexto da vida em plataformas digitais e da atuação de empresas de tecnologia privadas na estruturação do espaço público (Celeste; Santarém, 2021). Trata-se de uma “constituição em construção” da vida online, na qual valores como liberdade, igualdade e democracia não são ignorados porque a conversação se desloca para o meio digital. Ao contrário, tornam-se ainda mais visíveis e urgentes. Afinal, a atmosfera digital tornou-se o espaço primário onde as conversações humanas ocorrem (Castro, 2023).

No constitucionalismo digital, há uma verdadeira tensão entre regulação privada e regulação pública. De um lado, plataformas estabelecem normas próprias de conduta por meio de termos de uso e políticas internas, decidindo unilateralmente o que é permitido em seus espaços. De outro, Estados definem parâmetros legais para garantir direitos fundamentais na rede, utilizando ferramentas como lei, jurisprudência e regulação administrativa (Castro, 2023). Nesse cenário, cortes constitucionais como o STF são chamadas a resolver conflitos que envolvem não apenas leis estatais, mas também regras de empresas privadas que, na prática, condicionam quem é visto e ouvido em escala massiva (Castro, 2023).

No constitucionalismo digital, manifesta-se verdadeira tensão entre regulação privada e regulação pública. De um lado, plataformas estabelecem normas próprias de conduta por meio de termos de uso e políticas internas, decidindo unilateralmente o que é permitido em seus espaços. De outro, Estados definem parâmetros legais para garantir direitos fundamentais na rede, utilizando ferramentas como lei,





jurisprudência e regulação administrativa (Castro, 2023). Nesse cenário, cortes constitucionais como o Supremo Tribunal Federal são chamadas a resolver conflitos que envolvem não apenas leis estatais, mas também regras de empresas privadas que, na prática, condicionam quem é visto e ouvido em escala massiva (Castro, 2023).

Essa tensão deriva da transformação das plataformas digitais em infraestruturas essenciais da comunicação pública contemporânea. Diferentemente dos meios de comunicação tradicionais, que operavam sob licenças estatais e regimes de concessão pública, as plataformas digitais desenvolveram-se inicialmente sob paradigma de autorregulação quase absoluta, fundado na compreensão de que constituiriam meros intermediários técnicos neutros. Tal concepção, entretanto, mostrou-se progressivamente insustentável diante da evidência empírica de que algoritmos de recomendação, sistemas de ranqueamento e políticas de moderação exercem função curatorial ativa, determinando não apenas quais conteúdos circulam, mas também sua amplitude, velocidade e impacto sobre a formação da opinião pública.

A regulação privada manifesta-se primordialmente através dos termos de serviço e das diretrizes comunitárias, documentos contratuais unilateralmente elaborados pelas plataformas que estabelecem proibições comportamentais frequentemente mais restritivas que os limites penais e constitucionais estatais. Tais normas privadas proíbem, exemplificativamente, discursos considerados ofensivos segundo padrões corporativos, independentemente de constituírem ilícitos penais, e impõem sanções que variam desde a remoção de conteúdo até a exclusão permanente de usuários, sem necessariamente observar garantias processuais mínimas. A legitimidade dessa normatividade privada fundamenta-se, segundo seus defensores, na autonomia empresarial, na liberdade de iniciativa e no direito de propriedade sobre a infraestrutura tecnológica, argumentos que encontram respaldo na tradição liberal de limitação do poder estatal sobre atores privados.

Contudo, a aplicação desse paradigma liberal clássico às plataformas digitais enfrenta objeções substanciais. Primeiramente, a concentração oligopolística do mercado de redes sociais reduz drasticamente a possibilidade de escolha genuína pelos usuários, tornando questionável o argumento contratualista de que a adesão aos termos de serviço representa manifestação livre de vontade. Quando poucas plataformas controlam a maior parte do fluxo comunicacional relevante, a exclusão de seus espaços equivale, na prática, a uma forma de exclusão da esfera pública digital,





com consequências para o exercício efetivo de direitos políticos e comunicativos. Em segundo lugar, o caráter transnacional dessas empresas permite que imponham padrões normativos desenvolvidos segundo contextos culturais e jurídicos específicos — frequentemente norte-americanos — a sociedades com tradições constitucionais distintas, gerando tensões entre soberania regulatória nacional e governança corporativa global.

A regulação pública, por sua vez, manifesta-se através de múltiplos instrumentos normativos que buscam vincular as plataformas a obrigações compatíveis com o ordenamento constitucional e legal de cada jurisdição. No Brasil, o Marco Civil da Internet estabeleceu, em seu artigo 19, regime de responsabilidade civil subsidiária das plataformas, condicionando sua responsabilização ao descumprimento de ordem judicial específica de remoção de conteúdo. Tal modelo privilegiou a liberdade de expressão e a inovação tecnológica, evitando obrigações genéricas de monitoramento que poderiam conduzir à censura privada preventiva. Entretanto, a ausência de mecanismos de transparência, devido processo e accountability deixou ampla margem para decisões discricionárias das plataformas, que removem conteúdo voluntariamente segundo critérios próprios, sem necessariamente observar os limites constitucionais aplicáveis ao poder público.

A intervenção das cortes constitucionais nesse campo decorre da necessidade de compatibilizar a autonomia privada das plataformas com a eficácia horizontal dos direitos fundamentais, doutrina segundo a qual direitos constitucionais irradiam efeitos também sobre relações entre particulares, especialmente quando há assimetria significativa de poder.

O Supremo Tribunal Federal tem sido progressivamente demandado a posicionar-se sobre conflitos que envolvem remoção de conteúdo, suspensão de perfis e aplicação de políticas de moderação, situações nas quais normas privadas colidem com garantias constitucionais de liberdade de expressão, acesso à informação e devido processo legal. Nos julgamentos dos temas de repercussão geral 987 e 533, a Corte reconheceu que o modelo de autorregulação irrestrita mostra-se constitucionalmente insuficiente, demandando maior transparência, previsibilidade e possibilidade de revisão judicial das decisões moderadoras.

Essa judicialização revela, em última análise, que a tensão entre regulação privada e pública não se resolve mediante simples prevalência de uma sobre a outra,





mas exige construção de arranjos institucionais complexos que articulem legitimidade democrática, eficiência operacional e proteção de direitos.

A regulação pública deve estabelecer parâmetros substantivos e procedimentais que limitem a discricionariedade corporativa sem inviabilizar a gestão cotidiana das plataformas, enquanto a regulação privada deve subordinar-se aos princípios constitucionais fundamentais, especialmente quando suas decisões produzem efeitos análogos aos de atos estatais restritivos de direitos.

Este equilíbrio constitui o núcleo do que se denomina corregulação, modelo que distribui responsabilidades entre Estado, plataformas e sociedade civil, preservando espaços de autonomia privada sob supervisão pública orientada por transparência, proporcionalidade e accountability democrática.

Nesse sentido, a liberdade de expressão não pode ser entendida de modo irrestrito, como se fosse um direito absoluto e sem qualquer possibilidade de responsabilização em eventual abuso seu. De fato, trata-se de um direito preferencial, ou seja, dotado de carga valorativa maior em situações de conflito, mas que, de toda forma, admite restrições proporcionais e necessárias quando está em jogo a proteção de outros bens fundamentais, como a vida, a integridade física, a igualdade ou a própria ordem democrática (Sarlet, 2019).

Como sustentam Cruz e Mendes (2025), liberdade e responsabilidade são vetores indissociáveis. Isso significa que não se pode admitir que a liberdade de expressão sirva como escudo para práticas ilícitas como apologia ao terrorismo e outros crimes graves, produção e distribuição de material de abuso infantil, ataques organizados à integridade das instituições democráticas ou difusão de conteúdo anticientífico que coloque em risco a saúde pública. Senão em hipóteses extremas como essas, deve haver alta proteção para opiniões críticas, sátiras, paródias e manifestações de dissenso político, a fim de preservar o papel da liberdade de expressão na formação da opinião pública e no controle social do poder.

Em um ambiente digital, o equilíbrio entre liberdade e responsabilidade torna-se ainda mais difícil porque (i) a escala de disseminação dos conteúdos é muito maior, (ii) o tempo de reação é curto e (iii) o poder de intermediação das plataformas é gigantesco. Enquanto em um ambiente tradicional um comentário ofensivo tende a alcançar apenas um público restrito, no ambiente digital esse mesmo comentário pode atingir milhões de pessoas em poucas horas, mormente quando amplificado por algoritmos de recomendação (Gomes, 2022).





Ao confeccionarem e imporem seus termos de uso, as plataformas digitais assumem uma função que se assemelha à de um “legislador privado”, estabelecendo regras sobre o que pode ou não permanecer em circulação, quase sempre com pouca transparência e com mecanismos de recurso pouco compreensíveis para os usuários comuns (Gomes, 2022). A esse fenômeno, atribui-se o nome de “governança privada da *internet*”; embora formalmente se trate de contratos de adesão que usuários “concordam” ao criar contas, na prática essas regras têm efeitos equivalentes ao de normas constitucionais, pois impactam direitos fundamentais de milhões de pessoas simultaneamente (Gomes, 2022).

A partir do final do século XIX e início do século XXI, em meio a globalização e expansão da *internet*, surgem os chamados “direitos fundamentais de 4ª geração”, idealizados por Bobbio e relacionados ao avanço da tecnologia e à consolidação da sociedade da informação, exigindo uma proteção mais ampla da pessoa no ambiente digital, especialmente no que se refere à privacidade, proteção de dados e liberdade de expressão (Bobbio, 2004).

Nesse contexto, o constitucionalismo digital reforça a necessidade de compreender a liberdade de expressão não apenas como uma garantia na relação entre indivíduo e Estado, mas também como um direito a ser assegurado nas relações entre indivíduos e empresas privadas de tecnologia (Castro, 2023). Embora os direitos fundamentais tenham surgido, originalmente, como instrumentos de defesa contra abusos do poder estatal, o poder e a influência das *big techs* sobre a coletividade demandam que os princípios constitucionais irradiem efeitos também sobre esses atores, à luz da Teoria da eficácia horizontal dos direitos fundamentais, de modo a garantir a efetiva proteção desses direitos nas relações entre usuários e plataformas digitais (Castro, 2023).

Conforme ensinam Mendes e Fernandes (2020), “o constitucionalismo digital [...] procura estender os princípios e valores do constitucionalismo clássico para o ambiente virtual, de modo a assegurar que o exercício do poder privado pelas grandes empresas de tecnologia não resulte na aniquilação de direitos fundamentais. A arquitetura da rede, longe de ser um espaço anárquico ou neutro, é moldada por códigos e algoritmos que exercem uma verdadeira função normativa, exigindo, portanto, mecanismos de controle e limitação que sejam compatíveis com o Estado Democrático de Direito”.



Sendo assim, o constitucionalismo digital não deve ser visto como algo que veio para substituir as leis que já temos – não é um rompimento com a história do direito –, mas como uma evolução necessária para lidar com o poder dos algoritmos e das *big techs* (Castro, 2023). Para Celeste e Santarém (2021), esse conceito representa uma forma de fazer valer os valores de liberdade e dignidade humana dentro do mundo virtual, impedindo que o ciberespaço se torne um território de exceção onde os direitos fundamentais não penetram. Se no passado as constituições serviam para limitar o poder do Estado, hoje também servem para limitar o poder de grandes corporações, notadamente aquelas que controlam as infraestruturas digitais que mediam a esfera pública contemporânea.

Se as redes sociais são hoje o principal espaço onde a política, a cultura e o debate público se desenvolvem, não podem ser tratadas como simples plataformas privadas regidas apenas por interesses privados. Sua função social e seu impacto direto na formação da opinião pública conferem-lhes uma dimensão pública inegável, exigindo limites e responsabilidades proporcionais ao poder que exercem sobre a vida coletiva. Diante disso, surge a necessidade de refletir sobre a natureza e os limites da atuação das plataformas digitais, especialmente no que diz respeito à moderação de conteúdo, uma vez que tais práticas interferem diretamente na circulação de informações e na efetividade dos direitos fundamentais no ambiente virtual.

3 REGULAÇÃO DE MODERAÇÃO DE CONTEÚDO: O CASO BRASILEIRO

Se por um lado a liberdade de expressão é condição *sine qua non* para que cidadãos participem da vida pública, organizem movimentos sociais e fiscalizem governantes, por outro, a ausência completa de qualquer forma de moderação pode transformar o espaço digital em terreno fértil para a difusão de discurso de ódio, desinformação e ataques às instituições democráticas (Gomes, 2022). Moderar conteúdo, portanto, não é sinônimo de censurar opiniões incômodas ou silenciar crítica legítima, mas de estabelecer critérios para determinados conteúdos, seja porque violam frontalmente a lei penal, seja porque representam ameaças sérias a direitos fundamentais de terceiros ou ao próprio funcionamento da democracia (Gomes, 2022).





Em suma, moderação de conteúdo pode ser entendida como o conjunto de decisões sobre remoção, bloqueio, rotulagem, redução de alcance, suspensão de contas e outras medidas que plataformas adotam em relação aos conteúdos publicados por usuários (Castro, 2023). Essas decisões podem ser (a) automáticas, quando um algoritmo identifica palavras-chave ou padrões de comportamento e toma uma providência sem intervenção humana, ou (b) humanas, quando equipes de revisão analisam denúncias ou casos recomendados pelo sistema. Em muitos casos, há a combinação de ambas: um filtro automático identifica suspeita, depois um ser humano avalia se realmente há ou não violação de norma (Castro, 2023).

No Brasil, até a data deste estudo, ainda não há legislação específica sobre o tema. O Projeto de Lei (PL) n. 2.630/2020, que visa a instituir a “Lei brasileira de liberdade, responsabilidade e transparência na internet”, encontra-se pendente de votação na Câmara dos Deputados, em meio a disputas políticas e à atuação de grupos de interesse (Machado; Toledo, 2024).

Isso não significa, porém, ausência completa de regulação. O Marco Civil da Internet (MCI), aprovado em 2014, é a primeira e principal referência normativa do país, sobrevivendo, no ano de 2018, em complementariedade, a Lei Geral de Proteção de Dados (LGPD). Em seu tempo, o MCI foi inovador ao estabelecer princípios, garantias, direitos e deveres para o uso da internet no Brasil. No art. 19, estabeleceu um modelo de responsabilidade condicionada à ordem judicial para provedores de aplicação de internet, nos seguintes termos: *“com o intuito de assegurar a liberdade de expressão e impedir a censura, o provedor de aplicações de internet somente poderá ser responsabilizado civilmente por danos decorrentes de conteúdo gerado por terceiros se, após ordem judicial específica, não tomar as providências para, no âmbito e nos limites técnicos do seu serviço e dentro do prazo assinalado, tornar indisponível o conteúdo apontado como infringente, ressalvadas as disposições legais em contrário”* (Brasil, 2014).

Como se vê, o referido dispositivo dispõe que redes sociais e outros serviços só podem ser responsabilizados civilmente por danos decorrentes de conteúdos gerados por terceiros se, após decisão judicial, deixarem de remover o conteúdo ilícito apontado. O intuito do legislador foi o de evitar que plataformas implementassem políticas de remoção preventiva em massa, protegendo, assim, a liberdade de expressão contra a censura privada e deslocando para o juiz a palavra final sobre a ilicitude do conteúdo. Em seu contexto original, a norma simbolizou uma necessária e





relevante iniciativa em defesa da liberdade de expressão por trás das telas (Keller; Mendes; Fernandes, 2023).

Ocorre que, a evolução tecnológica e o agravamento dos danos digitais logo revelaram a insuficiência do modelo previsto no MCI. Quando confrontado com conteúdo de gravidade extrema, cujos danos se mostram irreversíveis na relação tempo-espço – como a disseminação de imagens de abuso sexual infantil, promoção de atos terroristas ou campanhas estruturadas para sabotar processos eleitorais –, somado à morosidade estrutural do Judiciário e sua limitação técnica frente ao volume de conteúdos digitais e o caráter reativo da lógica “ordem-remoção”, o art. 19 simplesmente não oferece proteção adequada (Keller; Mendes; Fernandes, 2023).

Diante da ausência de uma legislação específica e atualizada – fruto da inércia/omissão estratégica do Legislativo brasileiro –, os conflitos envolvendo responsabilidade de plataformas e moderação de conteúdo são crescentemente entregues ao STF (Castro, 2023). O Supremo passa a ser chamado não apenas a resolver casos individuais, mas a redefinir, por via interpretativa, o escopo de normas como o art. 19 do MCI, desempenhando função próxima à de um legislador substitutivo. Tal dinâmica reflete o que Machado (2019) descreve como “política togada”: a tendência de que temas estruturalmente políticos, diante da paralisia do Parlamento, sejam decididos por órgãos judiciais, com menor participação social e controle democrático típico de processos legislativos. Com efeito, o cidadão perde o direito de participar de um debate público e plural; e a regulação é feita em gabinetes fechados ou em decisões de tribunais, onde o rito é técnico e não político. A “fuga da política”, predominante no Brasil, sobrecarrega o Judiciário e o coloca em uma estranha posição carecedora de representatividade (Machado, 2019).

Esse fenômeno prevalecente no Brasil configura o que a doutrina intitula como “estratégias de ativismo” (Machado, 2019). Muitas vezes, os próprios partidos políticos, incapazes de vencer pelo voto no Congresso, provocam o Judiciário para obter por via judicial o que não obtiveram por via legislativa, naturalizando o ativismo judicial. Daí que o Judiciário passa a ser visto como o único Poder capaz de agir, enquanto o Legislativo é visto como inoperante, retirando do povo o poder de decidir por meio de seus representantes (Machado, 2008). Machado (2019) adverte que a judicialização da política é uma verdadeira armadilha, porquanto oferece uma solução rápida para problemas complexos, mas sob o custo de violação à separação de poderes – uma das cláusulas pétreas da Constituição (art. 60, § 4º, III).





O grande risco de se atribuir ao Judiciário a tarefa de “legislar” sobre a internet é o *déficit* de legitimidade democrática. Afinal, ministros do STF não possuem mandato popular; não foram eleitos para decidir os rumos da liberdade de expressão da nação. Quando o Judiciário assume essa posição, usurpando a competência do Legislativo, ele o faz por meio de procedimentos que não permitem a mesma participação e o contraditório social encontrados no Congresso (Machado, 2019). Lamentavelmente, como denuncia Machado (2019), tal postura é, muitas vezes, incentivada pela própria “inércia estratégica” dos políticos, que preferem que o STF tome as ditas “decisões impopulares”, resultando, de um lado, em um Judiciário sobrecarregado e superpoderoso e, de outro, em um Legislativo inoperante e enfraquecido, pondo em xeque a independência e harmonia entre os Poderes (art. 2º, CF).

Exemplo disso é o julgamento conjunto pelo STF dos Recursos Extraordinários n. 1.037.396 (Tema 987) e n. 1.057.258 (Tema 533), em junho de 2025. Até então, como visto, as plataformas estavam sob uma proteção quase total: só respondiam por um conteúdo ilícito se um juiz mandasse remover e elas ignorassem. No entanto, o STF reconheceu a inconstitucionalidade parcial e progressiva desse modelo. Para o Supremo, a regra do art. 19 criou um “estado de omissão”, pois não protegia de forma suficiente os direitos fundamentais e a própria democracia contra os abusos cometidos em larga escala no ambiente virtual.

Em síntese, decidiu o STF que, enquanto o Congresso não elaborar nova lei capaz sobre o tema, o art. 19 do MCI deve ser interpretado de acordo com a Constituição, de modo a oferecer maior proteção às pessoas contra conteúdos criminosos, ilegais e danosos na internet. Também decidiu que provedores de aplicações de internet, como redes sociais e buscadores, podem ser responsabilizados sem necessidade de ordem judicial quando forem notificados extrajudicialmente sobre crimes ou atos ilícitos existentes nas suas plataformas e não removerem tais conteúdos. Tal interpretação amplia o modelo já previsto no art. 21 do MCI, originalmente aplicado a casos de divulgação não consentida de cenas de nudez privadas. Essa mesma lógica passa a valer para crimes e atos ilícitos em geral, inclusive para casos de contas inautênticas ou falsas.

Para crimes contra a honra (calúnia, difamação e injúria), entendeu o STF que a responsabilização das plataformas continuará a exigir ordem judicial, conforme o art. 19 do MCI. Segundo a Corte, essa diferenciação é importante para proteger a





liberdade de expressão, evitando censura e remoção de conteúdos que veiculem críticas, ainda que incômodas. Porém, se o Judiciário entender que um determinado caso é de crime contra a honra e determinar a remoção, os provedores devem remover publicações com conteúdo idêntico, a partir de simples notificação, sem necessidade de novas decisões judiciais.

Segundo o STF, a regra do art. 19 do MCI continua a valer integralmente para alguns tipos específicos de provedores neutros, que não interferem sobre os conteúdos, como serviços de *e-mail*, aplicativos para realizar reuniões fechadas e serviços de mensagens instantâneas (como o WhatsApp), exclusivamente quanto às comunicações interpessoais, que são protegidas por sigilo constitucional.

Ainda de acordo com o STF, as plataformas podem ser responsabilizadas mesmo sem ordem judicial ou notificação privada em duas hipóteses específicas: (1) em anúncios ou impulsionamento pago de conteúdos, já que nesses casos a plataforma aprova a publicidade; e (2) quando for detectado o uso de redes artificiais de distribuição ilícitas usando robôs. Nesses casos, há uma presunção de que a plataforma tinha conhecimento da ilicitude e ela somente poderá afastar sua responsabilidade se provar que agiu em tempo razoável e com diligência para remover o conteúdo.

Por fim, decidiu o STF que, nos casos de crimes gravíssimos específicos, a plataforma deve zelar para que tais conteúdos não sejam sequer publicados. Nesses casos, aplica-se o chamado dever de cuidado, de modo que a plataforma deve atuar de maneira diligente e proativa para que esses conteúdos não circulem, independentemente de qualquer notificação ou ordem judicial. Essa regra se aplica aos seguintes crimes: (i) terrorismo; (ii) indução ao suicídio ou à automutilação; (iii) pornografia infantil e crimes graves contra crianças e adolescentes e pessoas vulneráveis; (iv) tráfico de pessoas; (v) discriminação e discurso de ódio; (v) crimes contra mulheres em razão de gênero; e (vi) atos antidemocráticos. A responsabilização por descumprimento desse dever de cuidado ocorrerá apenas quando houver falha sistêmica do provedor, ou seja, quando ele deixar de adotar medidas adequadas para prevenir ou remover esses conteúdos. A mera existência de um conteúdo ilícito isolado não basta para gerar responsabilidade.

Em todos esses casos, a responsabilização é subjetiva, ou seja, demanda análise de culpa ou dolo da plataforma. Para dar efetividade às regras de responsabilização, o STF decidiu que as plataformas devem, ainda, criar regras





próprias para: (i) criar sistema de notificação para usuários fazerem denúncias de crimes e atos ilícitos; (ii) disponibilizar canais de atendimento amplamente divulgados; (iii) implementar um devido processo que permita que os usuários entendam os fundamentos das decisões de remoção e possam recorrer; e (iv) elaborar relatórios anuais de transparência com os dados da atuação de remoção de conteúdo. Por fim, decidiu que provedores estrangeiros que atuam no Brasil devem manter representante legal no país para permitir o cumprimento de decisões judiciais e que a decisão vale apenas para casos futuros, de modo a garantir segurança jurídica.

A decisão do Supremo ilustra bem um paradoxo persistente no Brasil. Por um lado, corrige a percepção de uma imunidade quase absoluta das plataformas, recolocando-as sob o império da responsabilidade em situações de ilícitos graves (o que é bom). Por outro, expõe os riscos de uma extensão excessiva da tarefa de regulação para o Judiciário (o que é ruim). Ao admitir, ainda que de forma cautelosa, novos deveres e hipóteses de responsabilidade não previstos pelo legislador, o STF assume um papel normativo que tensiona a separação de poderes, a despeito de seu apelo ao Congresso para que legisle sobre o tema.

É inegável a relevância do julgado do Supremo Tribunal Federal para a proteção imediata de direitos fundamentais. Contudo, no longo prazo, a ausência de uma legislação amplamente debatida no Congresso Nacional pode tornar essa regulação frágil e suscetível a desvios autoritários. O constitucionalismo digital não subsistirá sem a devida segurança jurídica, o que só se verifica plenamente quando as normas decorrem de um consenso democrático. Atribuir ao Judiciário a tarefa de legislar sobre regulação de moderação de conteúdo on-line implica conceder-lhe um poder praticamente sem freios, deslocando a Corte Constitucional para o papel de um "supercensor" da vida digital brasileira — função incompatível com o desenho institucional da República.

A preocupação com a judicialização excessiva da regulação de plataformas digitais fundamenta-se na teoria da separação de poderes e nos limites funcionais da jurisdição constitucional. Embora o Supremo Tribunal Federal exerça legitimamente o controle de constitucionalidade e a proteção de direitos fundamentais, a definição de políticas regulatórias complexas, que demandam ponderações técnicas, avaliações de impacto econômico e escolhas valorativas sobre o equilíbrio entre liberdade e segurança, constitui tarefa primordialmente legislativa. A atuação judicial nesse campo, quando ultrapassa a função de estabelecer balizas constitucionais mínimas e





passa a detalhar obrigações específicas de moderação, incorre em risco de ativismo judicial que compromete a legitimidade democrática da regulação e fragiliza sua estabilidade institucional.

A regulação estabelecida mediante decisões judiciais apresenta limitações estruturais que a tornam inadequada como solução definitiva. Primeiramente, decisões judiciais respondem a casos concretos, não a planejamento regulatório abrangente, resultando em regime normativo fragmentado e assistemático. Em segundo lugar, a fundamentação judicial vincula-se às peculiaridades fáticas de cada litígio, dificultando a generalização de seus princípios para situações não idênticas. Terceiro, decisões judiciais carecem da flexibilidade necessária para ajustes periódicos diante da evolução tecnológica acelerada, exigindo novos processos para cada modificação normativa. Quarto, o processo judicial não permite a participação democrática ampla que caracteriza o debate legislativo, excluindo segmentos sociais não representados nas partes processuais. Por fim, a regulação judicial concentra enorme poder discricionário em poucos magistrados, sem os mecanismos de accountability político próprios do Poder Legislativo.

Ademais, a governança cotidiana da moderação de conteúdo permanece, em grande medida, sob o controle das próprias plataformas digitais. Mesmo sob a vigência do artigo 19 do Marco Civil da Internet, essas empresas removem conteúdos com base em termos de uso e políticas internas de alcance global, independentemente de ordem judicial, exercendo um poder normativo privado que, na prática, define os limites do discurso aceitável no ambiente digital. A legislação estatal limita-se a disciplinar hipóteses de responsabilidade civil e alguns deveres mínimos, sem impedir que as plataformas, segundo critérios próprios, excluam conteúdos considerados incompatíveis com suas políticas de comunidade. Forma-se, assim, uma convivência entre dois regimes distintos: um regime jurídico estatal, voltado à responsabilização e a deveres básicos, e um regime privado, no qual as empresas detêm ampla autonomia para estabelecer o escopo do discurso permitido em seus serviços (Castro, 2023).

Esse dualismo regulatório gera zona de opacidade normativa particularmente problemática para a proteção de direitos fundamentais. Enquanto o regime estatal subordina-se aos princípios constitucionais de legalidade, publicidade e devido processo legal, o regime privado opera segundo lógica contratual que, embora formalmente consensual, impõe-se de forma unilateral a usuários desprovidos de





poder de negociação efetivo. A assimetria entre plataformas globais e usuários individuais torna fictícia a premissa liberal de que termos de serviço representam acordo genuíno de vontades, aproximando-se mais de imposição normativa análoga à exercida por autoridades públicas. Essa analogia funcional justifica a submissão das plataformas a obrigações de transparência, fundamentação e possibilidade de revisão de suas decisões moderadoras, ainda que sejam atores privados.

O principal problema desse regime privado reside em sua opacidade. Para o usuário comum, as regras são frequentemente obscuras e sua aplicação pode ser influenciada por interesses econômicos ou políticos pouco transparentes. Quando a remoção de conteúdo ocorre por decisão unilateral da plataforma — e não por ordem judicial —, o usuário raramente sabe qual norma foi violada, não dispõe de mecanismos claros de contraditório e encontra grandes dificuldades para recorrer de forma efetiva (Gomes, 2022). Gomes (2022) documenta casos emblemáticos dessa assimetria, como a suspensão das contas de Donald Trump no X, a remoção de páginas de organizações políticas brasileiras ligadas ao Movimento Brasil Livre no Facebook e a exclusão de conteúdos de Jair Bolsonaro e seus aliados em diversas plataformas ao longo de 2023 e 2024. Em todos esses episódios, as decisões basearam-se em critérios não públicos, sem garantia efetiva de defesa e sem fundamentação detalhada por parte das empresas.

Esses episódios revelam déficit de accountability que caracteriza a governança privada das plataformas. Diferentemente de autoridades públicas, que submetem suas decisões ao controle judicial e à fiscalização de órgãos externos, as plataformas operam segundo estruturas corporativas hierárquicas nas quais mecanismos de revisão interna constituem instâncias recursais desprovidas de independência funcional. Mesmo quando estabelecem conselhos de supervisão externos, como o Oversight Board do Facebook, tais órgãos carecem de poder vinculante efetivo e limitam-se a recomendar decisões que as plataformas permanecem livres para acatar ou rejeitar. A ausência de instância revisora verdadeiramente independente contrasta drasticamente com o regime de garantias processuais aplicável a restrições estatais à liberdade de expressão, que sempre admitem controle jurisdicional.

A complexidade da moderação de conteúdo é agravada pelo fato de que, hoje, essa tarefa é amplamente atribuída à Inteligência Artificial. Como observa Esposito (2022), há uma distinção fundamental entre transparência e explicação. A mera divulgação de códigos ou bases de dados não esclarece, de modo compreensível





para humanos, por que uma decisão específica foi tomada. Segundo a autora, a Inteligência Artificial opera como uma forma de "comunicação artificial", produzindo resultados a partir de padrões estatísticos, e não de uma compreensão semântica do sentido. Por essa razão, sistemas automatizados tendem a falhar diante de ambiguidade, ironia e contextos culturais, resultando na remoção indevida de conteúdos legítimos.

A automação da moderação mediante algoritmos de aprendizado de máquina introduz camada adicional de opacidade ao regime privado. Sistemas de Inteligência Artificial treinados em grandes volumes de dados identificam padrões correlacionais sem necessariamente compreender o significado contextual das expressões linguísticas. Tal limitação mostra-se particularmente problemática quando aplicada a discursos complexos que dependem de ironia, sarcasmo, citação crítica ou referência contextual para sua correta interpretação. Um algoritmo treinado para identificar termos associados a discurso de ódio pode classificar erroneamente denúncia de racismo como manifestação racista, simplesmente pela presença de vocabulário estatisticamente correlacionado, independentemente da intenção comunicativa ou do contexto enunciativo.

Essa limitação técnica gera uma tensão permanente com a liberdade de expressão. Incapaz de compreender o significado do discurso, o sistema pode confundir uma crítica ao racismo com o próprio discurso racista (Esposito, 2022). A moderação, portanto, não pode ser reduzida a um processo mecânico e opaco. A liberdade de expressão deve ser protegida contra o arbítrio algorítmico, o que exige a presença de mecanismos efetivos de revisão humana. Embora o combate ao discurso de ódio e à desinformação seja necessário — o que é indiscutível —, as ferramentas utilizadas para tanto não podem operar como "caixas-pretas". Nesse sentido, a exigência de que decisões automatizadas sejam explicáveis em linguagem acessível emerge como um dos novos direitos promovidos pelo constitucionalismo digital, voltado à proteção do cidadão contra erros e abusos da automação.

O direito à explicação de decisões automatizadas fundamenta-se no princípio da dignidade humana e no direito ao devido processo legal, aplicados ao contexto da governança algorítmica. Quando decisões que afetam direitos fundamentais são delegadas a sistemas automatizados, a impossibilidade de compreender os critérios decisórios impede o exercício efetivo do direito de defesa e inviabiliza o controle de legalidade e proporcionalidade da medida restritiva. A explicabilidade não se confunde





com mera transparência do código-fonte, que permaneceria ininteligível para não especialistas e, mesmo para especialistas, não revelaria necessariamente a lógica decisória de sistemas de aprendizado de máquina cujo funcionamento emerge de correlações estatísticas não programadas explicitamente. Exige-se, portanto, explicação substantiva que identifique os fatores determinantes da decisão em linguagem compreensível, permitindo ao afetado contestar sua aplicação ao caso concreto.

A necessidade de revisão humana de decisões algorítmicas constitui, assim, salvaguarda essencial contra a automatização irrefletida da moderação. Tal exigência não implica inviabilizar o uso de Inteligência Artificial como ferramenta auxiliar de triagem e identificação preliminar de conteúdos potencialmente violadores, função na qual a escala da moderação torna indispensável o apoio tecnológico. Implica, contudo, reservar a decisão final sobre remoção de conteúdo ou suspensão de usuário a avaliação humana capaz de considerar nuances contextuais, ponderar proporcionalidade e fundamentar a decisão de modo compreensível.

Essa reserva de humanidade na moderação alinha-se ao princípio geral do direito administrativo segundo o qual decisões gravosas exigem motivação adequada e possibilidade de contestação, extensível por analogia às plataformas digitais quando exercem poder análogo ao estatal de restringir direitos fundamentais.

4 REGULAÇÃO DE MODERAÇÃO DE CONTEÚDO EM PERSPECTIVA COMPARADA: ESTADOS UNIDOS DA AMÉRICA, EUROPA E ALEMANHA

Em perspectiva comparada, destacam-se três modelos regulatórios de moderação de conteúdo on-line, cada qual com suas próprias lógicas, vantagens e desvantagens, oferecendo lições importantes – positivas e negativas – para o cenário brasileiro. O primeiro deles é o norte-americano, associado à Seção 230 da Lei Federal de Decência nas Comunicações de 1996 (*Section 230, Communications Decency Act*). O segundo é o europeu, expressado pelo Ato de Serviços Digitais (*Digital Services Act – DSA*), regulamentado pela União Europeia em 2024. O terceiro é o alemão, representado pela lei conhecida como *Netzwerkdurchsetzungsgesetz* (NetzDG), promulgada em 2017. Cada um desses modelos reflete escolhas políticas





e constitucionais distintas, e compreendê-los permite visualizar os *trade-offs* envolvidos em diferentes desenhos regulatórios.

O modelo norte-americano merece atenção especial porque influenciou decisões legislativas em muitos países, incluindo o Brasil através do MCI. A Seção 230, em sua formulação original, é um texto curto e potente. Em essência, diz que provedores de conteúdo interativo não serão responsabilizados por conteúdo gerado por terceiros, nem por remoção de conteúdo que acreditam ser prejudicial (“*no provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider*”).

Em termos históricos, o efeito dessa proteção foi enorme. Quando a lei foi aprovada, criou-se proteção a intermediários permitindo que plataformas crescessem sem medo de serem enterradas por litígios contínuos de usuários e de terceiros afetados por conteúdo. Essa proteção acelerou a inovação da internet americana: permitiu que serviços como Yahoo, eBay e, posteriormente, redes sociais como MySpace, Facebook e Twitter crescessem sem estrutura legal asfixiante (Stroppa *et al.*, 2022).

Do ponto de vista da liberdade de expressão, sua vantagem é que o Estado não pressiona as plataformas com ameaças de responsabilização civil generalizada, o que poderia levar a uma remoção em massa de conteúdos controversos por puro medo. A ideia era criar espaço para conversações livres, sem que uma paranoia de responsabilidade levasse à censura. Nesse sentido, a Seção 230 protegeu expressão legítima ao desencorajar remoção precipitada (Stroppa *et al.*, 2022).

Porém, do ponto de vista da proteção de outros direitos fundamentais, o modelo tem sido duramente criticado. Plataformas gigantes – Meta (Facebook, Instagram, WhatsApp), Google (YouTube, Gmail), X (antigo Twitter) – cresceram a ponto de serem praticamente monopólios de fato em determinados segmentos, valendo-se da norma para justificar práticas de moderação opacas, algoritmos de amplificação que prejudicam democracia e saúde pública, e negligência em face de conteúdo genuinamente ilícito como material de abuso infantil e relacionado a terrorismo. A crítica é que, à época de sua criação, em 1996, a Seção 230 foi pensada para “pequenos provedores” emergentes, não para monopólios de fato que controlam como bilhões de pessoas se comunicam (Stroppa *et al.*, 2022).

Por sua vez, o modelo europeu, manifestado notadamente pelo DSA, adota uma visão mais sofisticada e multifacetada. O DSA não revoga, como regra, a





proteção de intermediários contra responsabilidade automática por conteúdo de terceiros, mas estabelece um conjunto de obrigações para plataformas que buscam equilibrar liberdade de expressão com outros direitos (Nettesheim, 2022).

Entre as obrigações instituídas pela União Europeia (UE), tem-se a obrigatoriedade de transparência. As plataformas devem publicar dados sobre suas práticas de moderação, inclusive números de decisões tomadas por categoria, quantas foram revertidas após recurso de usuários e explicações dos critérios utilizados. Também devem publicar informações sobre seus algoritmos de recomendação – embora não sejam obrigadas a expor o código-fonte completo por razões de segurança e propriedade intelectual, devem fornecer informações suficientes para que pesquisadores e reguladores entendam como funcionam (Nettesheim, 2022).

Além disso, o DSA estabelece mecanismos procedimentais de notificação e contestação. Quando uma plataforma intenta remover conteúdo, em situações não-emergenciais, deve previamente notificar o usuário, informar qual regra foi violada e oferecer oportunidade de resposta antes da remoção. Depois da remoção, deve haver possibilidade de recurso. Isso é o que se chama de “devido processo procedimental” – garantias mínimas que asseguram que a decisão não é arbitrária e que a pessoa afetada teve chance de se defender (Nettesheim, 2022).

O DSA também foca no chamado “risco sistêmico”, conceito que engloba não apenas o conteúdo individual isolado, como também dinâmicas agregadas que afetam bens coletivos (eleições, saúde pública, crianças e adolescentes, instituições democráticas, informações confiáveis etc.). Por exemplo, um vídeo falso sobre uma política de saúde pode ser conteúdo isolado, mas se campanhas coordenadas espalharem milhões de versões falsas e algoritmos as amplificarem, tem-se um risco sistêmico. Por isso, o DSA impõe que plataformas realizem avaliações periódicas desses riscos e implementem medidas estruturais para mitigá-los. Outra exigência é a realização de auditoria independente. As plataformas devem submeter-se a auditorias conduzidas por terceiros independentes, os quais têm direito a acesso a dados internos e podem questionar práticas (Nettesheim, 2022).

Como ponto positivo, o modelo europeu busca equilibrar liberdade de expressão com deveres de cuidado e transparência proporcionais ao tamanho e impacto da plataforma, reconhecendo que empresas controladoras de espaços públicos digitais têm responsabilidades comparáveis a atores públicos e reclamando





delas a adoção de um “dever de governança responsável” para além da simples remoção após notificação (Nettesheim, 2022). Contudo, sua efetividade ainda é incerta, já que depende de força e ética de atuação dos 27 Estados-membros da UE em fiscalizar as *big techs*, que, como cediço, detêm considerável poder (político e econômico, principalmente). Um outro risco está no fato de que o excesso de exigências leve as plataformas a adotarem filtros demasiadamente rígidos, com potencial de remoção de conteúdo legítimo por receio de sanção (Nettesheim, 2022).

Por seu turno, o modelo alemão exemplifica uma abordagem mais mandatária e dura. A lei alemã, o *Netzwerkdurchsetzungsgesetz* (NetzDG), foi aprovada em 2017 em resposta a preocupações sobre proliferação de discursos nazistas, racistas, antissemitas e antijudaicos em redes sociais. A referida lei reflete a tradição constitucional alemã que, marcada pela história do nazismo e do Holocausto, admite restrições mais rigorosas a certos tipos de discurso que ameaçam a ordem democrática e a dignidade de grupos protegidos (Sarlet, 2024).

A NetzDG impõe às plataformas que removam rapidamente – geralmente em 24 horas – conteúdo manifestamente ilícito, especialmente quando constitui, sob a lei penal alemã, discurso de ódio, impondo multas de até 50 milhões de euros na hipótese de descumprimento. Entretanto, a rigidez desse modelo tem ocasionado efeitos não intencionais nada saudáveis ao Estado Democrático, a exemplo do *overremoval* (“remoção excessiva”). Com receio de serem multadas, as plataformas passaram a remover conteúdo legítimo por mera precaução. Assim, críticas a políticas migratórias são frequentemente confundidas com discurso de ódio e removidas. O mesmo ocorre em relação a piadas e comentários humorísticos que, embora incluam referências sensíveis ao público, mas sem incitação à violência, são removidos sob o pretexto de configurarem discurso de ódio (Sarlet, 2024).

Ao considerar tal modelo, Sarlet (2024) pondera que não se deve levar em conta apenas os sucessos do NetzDG contra conteúdos genuinamente prejudiciais, mas também seus riscos de censura. Essa análise comparativa permite identificar as diferentes características de cada modelo. O norte-americano privilegia a liberdade de expressão e a inovação ao máximo, mas sacrifica responsabilização das plataformas e proteção contra danos difusos a grupos vulneráveis e a instituições democráticas. O europeu busca equilíbrio mais sofisticado, com obrigações de transparência, mecanismos de governança e direitos procedimentais, mas ainda enfrenta desafios na prática e risco de remoção excessiva. O alemão é deveras rígido, com imposição





de prazos curtos e multas elevadíssimas; combate duramente o discurso de ódio, mas sob o risco concreto de censura e privatização de decisões penais – na medida em que as plataformas decidem o que é crime de discurso de ódio, e não os tribunais.

A simples importação desses modelos para o Brasil, sem a devida filtragem contextual, pode gerar sérios problemas ao país. Mostra-se incompatível com a tradição de proteção preferencial da liberdade de expressão prevista na Constituição de 1988 se o Brasil adotasse integralmente o NetzDG. Igualmente inadequado seria simplesmente replicar a Seção 230 dos Estados Unidos, feito para um contexto político, econômico e social bastante diferente. Mesmo o DSA da União Europeia, apesar de mais equilibrado e sofisticado, foi pensado para uma estrutura regulatória ainda longe de existir no Brasil.

A constatação de que nenhum dos modelos analisados se ajusta integralmente à realidade nacional evidencia a necessidade de um arranjo próprio, que concilie proteção à liberdade de expressão com mecanismos eficazes de enfrentamento a danos digitais e legitimidade democrática. É nesse contexto que se insere a proposta de correção, apresentada a seguir como alternativa ao cenário brasileiro.

5 CORREGULAÇÃO: UMA ALTERNATIVA PARA O BRASIL

A expressão correção, tal como utilizada neste artigo, designa um arranjo híbrido em que Estado, plataformas e sociedade civil compartilham responsabilidades na definição e aplicação de regras sobre moderação de conteúdo, em conformidade com normas constitucionais, mecanismos de transparência, *accountability* (prestação de contas) e participação multissetorial (Libman, 2023).

Não se trata, pois, de deixar tudo nas mãos de empresas privadas, regime conhecido como autorregulação, nem de concentrar todo o poder de decisão no Estado através de lei rígida e agência estatal monopolista, regime conhecido como regulação mandatária. Em vez disso, a correção busca fazer com que o Estado defina objetivos e parâmetros mínimos de proteção a direitos, que as plataformas implementem soluções em atenção a esses objetivos e parâmetros, e que órgãos independentes, com participação multissetorial, monitorem o cumprimento dessas regras, com transparência contínua e possibilidade recursal aos usuários afetados por decisões abusivas (Libman, 2023).





A ideia de correção, entretanto, não é essencialmente nova. Também se verifica em áreas como direito ambiental, segurança alimentar e privacidade, cujas experiências evidenciam que o regime puramente governamental é frequentemente rígido e ineficiente, enquanto a autorregulação privada carece de *accountability*. Assim, parece funcionar melhor um modelo em que o governo estabelece objetivos e diretrizes mínimas, as empresas privadas atuam com segurança e responsabilidade e são continuamente monitoradas por órgãos independentes (Libman, 2023).

O modelo de correção aqui proposto se estrutura em cinco pilares fundamentais e interdependentes. O primeiro é a transparência, concretizada no dever imposto às plataformas para que publiquem relatórios claros e periódicos sobre suas práticas de moderação de conteúdo, os quais devem incluir dados sobre quantos conteúdos foram removidos por categoria, quantas contas foram suspensas, qual é a taxa de remoção em comparação com volume total de conteúdo publicado, quantos recursos foram apresentados por usuários contra decisões de moderação, e qual é a taxa de reversão dessas decisões quando usuários recorrem. As plataformas devem também explicar, em termos acessíveis, quais critérios subjetivos utilizados – quando há margem para julgamento – e quais mecanismos automatizados. Também devem indicar se certos grupos de usuários ou certas formas de expressão estão sendo desproporcionalmente afetados (Almeida, 2022).

Inexistindo tal transparência, impossível avaliar se há vieses sistemáticos nas práticas de moderação, se grupos específicos ou ideologias específicas são tratados de forma discriminatória, se há remoções injustificadas ocorrendo em escala ou se determinado tipo de conteúdo legítimo está sendo alvo de remoções excessivas por mero erro ou por preferência política. Essa transparência não exige das empresas a exposição de seus segredos industriais, mas tão somente que forneçam informações suficientes que expliquem o porquê da lógica adotada (Almeida, 2022).

O segundo é o devido processo, isto é, o conjunto de garantias mínimas que qualquer pessoa deve ter antes de sofrer uma decisão que a prejudique. Isso inclui, por exemplo, o direito de ser informado sobre eventuais acusações, à ampla defesa e ao contraditório, à decisão fundamentada e, claro, o de recorrer. Assim, em um ambiente digital, os usuários devem ser notificados acerca da remoção de seus conteúdos, da suspensão ou bloqueio de suas contas, saber qual regra supostamente foi violada, ter acesso à explicação da decisão em linguagem compreensível,



oportunidade de contestar a decisão, receber respostas dentro de prazos razoáveis e poder recorrer a uma instância independente da plataforma (Almeida, 2022).

O terceiro é a proporcionalidade. Proporcionalidade aqui pode ser entendida como a exigência de que as respostas regulatórias sejam graduadas de acordo com a gravidade do conteúdo e o grau de risco associado. Isso implica, por um lado, evitar tolerância irrestrita a discursos manifestamente criminosos e danosos, mas, por outro lado, evitar também sanções máximas para infrações de baixa gravidade ou para conteúdo apenas controverso. Comentários isolados e pouco difundidos que violem levemente uma norma de conduta não devem receber a mesma resposta que uma campanha coordenada para incentivar ataques violentos contra um grupo étnico ou para disseminar mentiras estruturadas sobre funcionamento de eleições (Almeida, 2022).

O quarto é a supervisão independente e multisetorial. Isso significa criar ou fortalecer órgãos que incluam representantes do Estado, plataformas, academia científica e sociedade civil, com a função precípua de monitorar indicadores de moderação, revisar práticas, emitir recomendações e resolver disputas em instância externa às plataformas. Tais organismos podem funcionar como instâncias de apelação para usuários que se sintam injustiçados pelas decisões de moderação, oferecendo alternativa menos custosa e mais especializada que o Poder Judiciário tradicional, que não tem expertise em dinâmicas de plataformas digitais (Almeida, 2022). Nesse ponto, a experiência europeia com mecanismos de resolução extrajudicial de disputas, a exemplo de sistemas de *ombudsman* ou conselhos externos de supervisão, serve de referência para o desenho de tais estruturas.

O quinto e último reside na adequação específica ao contexto brasileiro. A Constituição de 1988 consagrou uma posição forte para a liberdade de expressão – é direito preferencial que só cede a restrições proporcionais e legítimas. A Constituição, contudo, também enfatiza a proteção da dignidade da pessoa humana, igualdade, pluralismo político e democracia. Qualquer modelo de correção necessita ser compatível com essa combinação específica de proteções constitucionais e capacidades reais do Estado nacional. Deve-se evitar tanto a adesão a um modelo de proteção extremada (alemão) quanto a adoção acrítica de regimes de ampla imunidade (norte-americano). Deve-se, pois, recorrer a um meio termo (regime híbrido) adequado à complexa realidade brasileira.





Mesmo o modelo de correção aqui proposto tem seus riscos. Há, por exemplo, o risco de captura regulatória, ou seja, de que órgãos de supervisão independentes sejam gradualmente dominados por interesses políticos específicos (*lobby*), deixando de atuar com independência e imparcialidade. Há também o risco de fragilidade de implementação, dada a carência estrutural de recursos técnicos e humanos das agências estatais brasileiras, bem como a instabilidade de prioridades políticas entre diferentes governos presidencialistas. Há o risco de que plataformas menores enfrentem custos desproporcionais para cumprir regras complexas, enquanto gigantes globais absorvem tais custos como simples custo de fazer negócio. Há, ainda, o risco de que a regulação se torne pretexto para que governos silenciem críticas políticas, utilizando órgãos supervisores como ferramentas de controle de fala.

Tais riscos, porém, não justificam a inação ou a adesão a alternativas que historicamente se mostram piores. Ao contrário, servem de alerta para que a regulação seja pensada com cuidado, critérios, transparência, *accountability*, regras de composição de órgãos que evitem concentração de poder em um único grupo ou ideologia, mandatos fixos e longos para dirigentes, implementação gradual a começar pelas plataformas de maior impacto sistêmico *etc.*

Ante o exposto, a análise desenvolvida ao longo deste estudo confirma a hipótese 3 (H3) frente ao problema de pesquisa proposto. Ao demonstrar que a correção supera tanto as insuficiências da autorregulação privada quanto os riscos da regulação estatal concentrada, infere-se que tal modelo oferece (i) maior legitimidade democrática, (ii) maior proteção aos direitos e garantias fundamentais e (iii) maior capacidade de adaptação às dinâmicas tecnológicas e comunicacionais.

Oferece maior legitimidade democrática porque transfere a definição e aplicação de regras de moderação de um polo decisório único – seja estatal, seja privado – para um arranjo plural, no qual objetivos e diretrizes são estabelecidos por instâncias democraticamente legitimadas, enquanto sua implementação e fiscalização contam com a participação estruturada da sociedade civil, academia científica e órgãos independentes, reduzindo *déficits* de representação e riscos de arbítrio.

Assegura maior proteção aos direitos e garantias fundamentais ao incorporar, de forma institucionalizada, exigências de transparência, devido processo, proporcionalidade e revisão independente, o que limita decisões opacas, assegura contraditório efetivo e previne tanto censura privada indiscriminada quanto intervenções estatais desmedidas.





Fornece maior capacidade de adaptação aos desafios tecnológicos e comunicacionais porque permite que a aplicação de suas regras seja reajustada ao longo do tempo, possibilitando que eventuais mudanças nos algoritmos das plataformas, no modo como as pessoas usam as redes sociais ou no surgimento de novas formas de desinformação e de atuação coordenada sejam acompanhadas e corrigidas gradualmente, sem depender sempre de novas leis ou de reformas legislativas morosas e complexas. Assim, evita-se a rigidez dos modelos exclusivamente legislativos e torna-se possível responder com mais rapidez e adequação às transformações do ambiente digital.

Dessa forma, a correção se apresenta como a resposta constitucional mais adequada ao desafio de regular a moderação de conteúdo on-line em plataformas digitais sem comprometer a liberdade de expressão, ao mesmo tempo em que possibilita o enfrentamento efetivo do discurso de ódio, da desinformação em massa e de ataques coordenados às instituições democráticas.

6 CONSIDERAÇÕES FINAIS

A presente investigação enfrentou um dos dilemas centrais do constitucionalismo digital contemporâneo: como regular a moderação de conteúdo em plataformas privadas preservando a liberdade de expressão — direito preferencial consagrado pela Constituição de 1988 — sem desconsiderar os riscos concretos de discurso de ódio, desinformação e ataques coordenados às instituições democráticas.

A análise desenvolvida demonstrou que a solução para esse problema não reside nem na autorregulação privada irrestrita, que concentra poder normativo em empresas desprovidas de legitimidade democrática adequada, nem na regulação estatal mandatária rígida, que tende a produzir remoção excessiva de conteúdo e a transferir decisões juridicamente complexas para atores privados sob ameaça de sanções desproporcionais. Ambos os modelos, conforme evidenciado, apresentam deficiências estruturais que os tornam inadequados ao contexto constitucional brasileiro.

O exame crítico do modelo brasileiro vigente, especialmente à luz do artigo 19 do Marco Civil da Internet e de sua recente reinterpretação pelo Supremo Tribunal Federal nos temas de repercussão geral 987 e 533, revelou cenário marcado por





insegurança jurídica, fragmentação normativa e judicialização excessiva. Embora as decisões do Supremo Tribunal Federal tenham desempenhado papel relevante na correção de omissões legislativas e na proteção imediata de direitos fundamentais, constatou-se que a transferência da função regulatória primária para o Judiciário gera déficit de legitimidade democrática, tensiona o princípio da separação de poderes e não oferece respostas estruturais e estáveis para a governança da moderação de conteúdo no longo prazo. A regulação por via jurisdicional, conquanto necessária em situações emergenciais, mostra-se inadequada como solução definitiva para problemática que demanda escolhas políticas complexas, avaliações de impacto multidimensional e capacidade de adaptação contínua às transformações tecnológicas.

A análise comparada dos modelos norte-americano, europeu e alemão corroborou essa conclusão ao demonstrar que não há solução universalmente transplantável. Cada modelo reflete escolhas constitucionais, históricas e institucionais específicas, com ponderações próprias entre liberdade, responsabilização e controle estatal. O modelo norte-americano, fundado na Primeira Emenda e na imunidade da Seção 230 do Communications Decency Act, privilegia radicalmente a liberdade de expressão e a autorregulação privada, contexto incompatível com a tradição constitucional brasileira de maior intervencionismo estatal na proteção de direitos fundamentais. O modelo europeu, consubstanciado no Digital Services Act, estabelece regime sofisticado de obrigações de transparência, due diligence e responsabilização graduada, mas pressupõe capacidade institucional e recursos administrativos ainda incipientes no Brasil. O modelo alemão da NetzDG, embora tenha inspirado debates legislativos nacionais, demonstrou gerar remoção excessiva de conteúdo mediante transferência de decisões penais complexas para moderadores privados sob pressão de sanções administrativas severas. A simples importação de tais arranjos para o Brasil, sem adequada adaptação ao contexto constitucional e institucional doméstico, mostrou-se inadequada e potencialmente contraproducente.

Nesse cenário, o modelo de coregulação apresentou-se como alternativa mais compatível com o ordenamento jurídico brasileiro e suas especificidades institucionais. Ao distribuir responsabilidades entre Estado, plataformas e sociedade civil, mediante mecanismos de transparência, devido processo, proporcionalidade e supervisão independente, a coregulação oferece três vantagens fundamentais.





Primeiro, assegura maior legitimidade democrática, porquanto evita concentrações decisórias unilaterais seja no Estado, seja nas plataformas, distribuindo competências segundo princípio de subsidiariedade que reserva ao legislador as definições normativas estruturantes, às plataformas a implementação operacional segundo suas especificidades técnicas, e à sociedade civil a participação fiscalizatória. Segundo, proporciona maior proteção aos direitos e garantias fundamentais mediante institucionalização de mecanismos de controle, revisão e fundamentação das decisões moderadoras, submetendo tanto o poder público quanto o poder privado a obrigações de accountability compatíveis com o impacto de suas decisões sobre direitos constitucionais. Terceiro, demonstra superior capacidade de adaptação às transformações tecnológicas e comunicacionais, uma vez que combina estabilidade normativa — mediante estabelecimento legislativo de princípios e procedimentos gerais — com flexibilidade operacional — permitindo ajustes técnicos às especificidades de cada plataforma e à evolução tecnológica, sem necessidade de sucessivas reformas legislativas. Esses achados confirmam a hipótese H3 formulada inicialmente, segundo a qual a correção constitui o modelo regulatório mais adequado ao contexto brasileiro.

A correção proposta fundamenta-se em cinco pilares estruturantes. Primeiro, transparência algorítmica, exigindo que plataformas divulguem critérios substantivos de moderação, métricas de aplicação e funcionamento geral de sistemas automatizados, sem necessariamente revelar detalhes proprietários que comprometam a segurança ou viabilidade comercial do serviço. Segundo, devido processo legal aplicado à moderação, garantindo notificação fundamentada, oportunidade de defesa e possibilidade de recurso independente para decisões que resultem em remoção de conteúdo ou suspensão de usuário. Terceiro, proporcionalidade e subsidiariedade, privilegiando medidas menos restritivas como rotulagem, desmonetização ou redução de alcance quando suficientes para mitigar riscos, reservando a remoção e suspensão para hipóteses de ilegalidade manifesta ou dano iminente. Quarto, supervisão independente mediante autoridade reguladora dotada de autonomia funcional, capacidade técnica e legitimidade democrática para fiscalizar o cumprimento de obrigações, receber denúncias, conduzir auditorias e aplicar sanções proporcionais. Quinto, participação social institucionalizada em processos de elaboração de códigos de conduta, avaliação de impacto e revisão





periódica de políticas de moderação, assegurando que a regulação reflita pluralidade de perspectivas e não apenas interesses corporativos ou estatais.

Reconhece-se que a correção não está isenta de riscos, notadamente captura regulatória por grupos de interesse, fragilidades de implementação decorrentes de insuficiência de recursos ou capacidade técnica, e instrumentalização política mediante uso seletivo de mecanismos fiscalizatórios para intimidação de adversários. Tais desafios, contudo, não justificam a inação nem a manutenção de modelos comprovadamente insuficientes. Ao contrário, impõem a necessidade de desenho institucional cuidadoso que incorpore salvaguardas contra esses riscos, mediante garantias de independência da autoridade reguladora, participação ampla de stakeholders nos processos decisórios, controle judicial de decisões administrativas e avaliações periódicas de impacto regulatório.

A efetivação da correção demanda, ademais, reformas legislativas estruturantes que superem a fragmentação normativa atual e estabeleçam regime coerente e sistemático. O Projeto de Lei 2630/2020, conhecido como PL das Fake News, constitui oportunidade para incorporação desses princípios, desde que emendado para evitar os riscos de remoção excessiva identificados em versões anteriores e para fortalecer garantias procedimentais aos usuários. A legislação deve estabelecer não apenas obrigações substantivas de moderação, mas principalmente requisitos procedimentais e institucionais que assegurem legitimidade, previsibilidade e controlabilidade das decisões, públicas e privadas, que afetem a liberdade de expressão no ambiente digital.

Em perspectiva mais ampla, a governança da moderação de conteúdo on-line insere-se no desafio maior de adaptação do constitucionalismo às transformações da sociedade informacional. A concentração de poder comunicacional em plataformas privadas transnacionais, a mediação algorítmica da esfera pública, a dataficação das relações sociais e a automação de decisões outrora exclusivamente humanas exigem reconfiguração dos instrumentos tradicionais de proteção de direitos fundamentais e controle do poder. A correção, nesse contexto, representa não apenas modelo técnico de gestão de riscos, mas arranjo institucional que reconhece a pluralidade de atores relevantes na governança digital e busca compatibilizar suas competências segundo princípios de legitimidade democrática, efetividade na proteção de direitos e capacidade adaptativa.



A construção de uma governança democrática sobre a moderação de conteúdo constitui, portanto, imperativo tanto jurídico quanto político. Jurídico, porque a eficácia dos direitos fundamentais no ambiente digital exige mais que sua proclamação formal, demandando estruturas institucionais que os tornem efetivos contra ameaças tanto estatais quanto privadas. Político, porque a vitalidade da democracia contemporânea depende crescentemente da qualidade do debate público digital, cuja integridade não pode ser assegurada nem pela autorregulação corporativa irrestrita nem pela censura estatal, mas apenas mediante governança compartilhada, transparente e accountable que articule Estado, mercado e sociedade civil na proteção do espaço público digital como bem comum essencial à autodeterminação coletiva.

REFERÊNCIAS

ALMEIDA, Clara Leitão de. *Regulação da transparência em plataformas digitais e legitimidade na moderação de conteúdo*. 132 f. Dissertação (Mestrado em Direito da Regulação) – Escola de Direito do Rio de Janeiro, Fundação Getúlio Vargas – FGV Direito Rio, Rio de Janeiro, jan. 2022. Disponível em: <https://repositorio.fgv.br/server/api/core/bitstreams/bd8ce78f-82d5-46bc-b596-5eefa7588c0b/content>. Acesso em: 6 jan. 2026.

ANDRADE, André Gustavo Corrêa de. Liberdade de expressão e discurso de ódio. *Revista da EMERJ*, Rio de Janeiro, v. 23, n. 1, p. 9-34, jan. 2023. Disponível em: <https://ojs.emerj.com.br/index.php/revistadaemerj/article/view/493>. Acesso em: 6 jan. 2026.

BARROSO, Luís Roberto. *Curso de direito constitucional contemporâneo*. 13. ed. Rio de Janeiro: SRV, 2025.

BOBBIO, Norberto. *A era dos direitos*. Rio de Janeiro: Elsevier, 2004.

BRASIL. Câmara dos Deputados. *Projeto de Lei n. 2630/2020*. Disponível em: <https://www.camara.leg.br/proposicoesWeb/fichadetramitacao?idProposicao=2256735>. Acesso em: 6 jan. 2026.

_____. *Constituição da República Federativa do Brasil*. Brasília, DF, 1988. Disponível em: https://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm. Acesso em: 6 jan. 2026.

_____. Lei n. 12.965, de 23 de abril de 2014. *Marco Civil da Internet*. Brasília, DF, 2014. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2014/lei/l12965.htm. Acesso em: 6 jan. 2026.





BRUGGER, Winfried. Liberdade de expressão e proibição do discurso do ódio: estudo comparativo entre o direito constitucional brasileiro e o norte-americano. *Revista de Direito Público*, Brasília, v. 4, n. 15, fev. 2010. Disponível em: <https://www.portaldeperiodicos.idp.edu.br/direitopublico/article/view/1418>. Acesso em: 6 jan. 2026.

BRUNETTA, Cíntia; LEITÃO, Andre Studart; DIAS, Eduardo Rocha. Da pessoa e para a pessoa: a regulação jurídica dos algoritmos. *Revista de Informação Legislativa: RIL*, Brasília, DF, v. 59, n. 233, p. 163-178, jan./mar. 2022. Disponível em: https://www12.senado.leg.br/ril/edicoes/59/233/ril_v59_n233_p163

CASTRO, Edgard Silva de. *Constitucionalismo digital e moderação de conteúdo: a tensão entre a regulação privada e a regulação pública e o papel da jurisdição constitucional*. 2023. Dissertação (Mestrado em Direito Constitucional) – Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa, Brasília, 2023. Disponível em: <https://repositorio.idp.edu.br/handle/123456789/5063>. Acesso em: 6 jan. 2026.

CELESTE, Edoardo; SANTARÉM, Paulo Rená da Silva. Constitucionalismo digital: mapeando a resposta constitucional aos desafios da tecnologia digital. *Revista Brasileira de Direitos Fundamentais & Justiça*, v. 15, n. 45, p. 63-91, jul./dez. 2021. Disponível em: <https://dfj.emnuvens.com.br/dfj/article/view/1219>. Acesso em: 6 jan. 2026.

CRUZ, Francisco Brito; MENDES, Laura Schertel. STF e a responsabilidade das plataformas: um novo equilíbrio. *JOTA*, 25 jun. 2025. Disponível em: <https://www.jota.info/opiniao-e-analise/artigos/stf-e-a-responsabilidade-das-plataformas-um-novo-equilibrio>. Acesso em: 6 jan. 2026.

ESPOSITO, Elena. Transparência versus explicação: o papel da ambiguidade na IA jurídica. *Revista Direito Mackenzie*, São Paulo, v. 20, n. 2, nov. 2022. Disponível em: <https://editorarevistas.mackenzie.br/index.php/rmd/article/view/15705>. Acesso em: 6 jan. 2026.

FALCÃO, Paula; SOUZA, Edivânio Duarte. Pandemia de desinformação: as fake news no contexto da Covid-19 no Brasil. *RECIIS: Revista Eletrônica de Comunicação, Informação & Inovação em Saúde*, Rio de Janeiro, v. 15, n. 1, p. 55-71, jan./mar. 2021. Disponível em: <https://www.reciis.icict.fiocruz.br/index.php/reciis/article/view/2219/2416>. Acesso em: 6 jan. 2026.

GOMES, Suldblano Oliveira. *Liberdade de expressão e governança da internet: remédios alternativos da moderação de conteúdo nas plataformas digitais*. 2022. Dissertação (Mestrado em Direito Constitucional) – Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa, Brasília, 2022. Disponível em: <https://repositorio.idp.edu.br/handle/123456789/5405>. Acesso em: 6 jan. 2026.

KELLER, Clara Iglesias; MENDES, Laura Schertel; FERNANDES, Victor. Moderação de conteúdo em plataformas digitais: caminhos para a regulação no Brasil. In: *Cadernos Adenauer*, Rio de Janeiro, v. 24, n. 1, p. 63-87, 2023. Disponível em:





<https://www.kas.de/documents/265553/19294631/Cadernos+1-2023+-+cap+4.pdf/efdd5b2d-e5c5-6c5c-8a0c-315cef9cb65b>. Acesso em: 6 jan. 2026.

LIBMAN, Juliana. *Moderação de conteúdo em redes sociais: por uma regulação que promova a liberdade de expressão*. 137 f. Dissertação (Mestrado Profissional em Direito Civil Contemporâneo e Prática Jurídica) — Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Direito, Rio de Janeiro, 2023. Disponível em: <https://www.maxwell.vrac.puc-rio.br/62807/62807.PDF>. Acesso em: 6 jan. 2026.

MACHADO, Joana de Souza. *Política togada no Brasil: ativismo judicial entre promessas, armadilhas e estratégias*. 176 f. Tese (Doutorado em Direito) — Departamento de Direito, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2019. Disponível em: <https://www.maxwell.vrac.puc-rio.br/70424/70424.PDF>. Acesso em: 6 jan. 2026.

MACHADO, Joana; TOLEDO, Larissa Fonseca. Plataformas digitais e discurso de ódio entre liberalismo e constitucionalismo digital: uma análise do modelo regulatório do PL 2630/2020. *International Journal of Digital Law*, Belo Horizonte, v. 5, n. 3, p. 67-93, set./dez. 2024. Disponível em: <https://journal.nuped.com.br/index.php/revista/article/view/1286>. Acesso em: 6 jan. 2026.

MEDEIROS, Jackson da Silva. Considerações sobre a esfera pública: redes sociais na internet e participação política. *Transinformação*, Campinas, v. 25, n. 1, p. 27-33, jan./abr. 2013. Disponível em: <https://www.scielo.br/j/tinf/a/55V8ZbVsCHtdP6CCRrXgbWB/?lang=pt>. Acesso em: 6 jan. 2026.

MENDES, Gilmar Ferreira; FERNANDES, Victor Oliveira. Constitucionalismo digital e jurisdição constitucional: uma agenda de pesquisa para o caso brasileiro. *Revista Brasileira de Direito*, Passo Fundo, v. 16, n. 1, p. 1-33, jan./abr. 2020. Disponível em: <https://seer.atitus.edu.br/index.php/revistadedireito/article/view/4103/2571>. Acesso em: 6 jan. 2026.

NETTESHEIM, Martin. A regulação de grandes plataformas da internet no Direito da União Europeia: as propostas da Comissão para um *Digital Markets Act* e um *Digital Services Act*. *Revista Brasileira de Direitos Fundamentais & Justiça*, Belo Horizonte, v. 16, n. 47, p. 21-45, jul./dez. 2022. Disponível em: <https://dfj.emnuvens.com.br/dfj/article/download/1355/1085/5173>. Acesso em: 6 jan. 2026.

PIMENTA, Ricardo M.; REGLY, Tainá; BELFORT, Hugo; SILVA, Anna Karla S. A mecânica da desinformação eleitoral: fake news e o paralelo com as “trend topics” das redes sociais em 2022. *Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação*, Florianópolis, v. 29, p. 1-26, 2024. Disponível em: <https://www.scielo.br/j/eb/a/pQnFBqBYcKw9XRhPNp3MwSG/>. Acesso em: 6 jan. 2026.

Santos, A. R. R., Leitão, A. S., & Brunetta, C. M. (2025). INTELIGÊNCIA ARTIFICIAL E BPC-LOAS: limites e desafios na avaliação da miserabilidade e no tratamento da





litigiosidade. *Revista CEJ*. Recuperado de
[//revistacej.cjf.jus.br/cej/index.php/revcej/article/view/2936](http://revistacej.cjf.jus.br/cej/index.php/revcej/article/view/2936)

SARLET, Ingo Wolfgang. Liberdade de expressão e o desafio da regulação do discurso do ódio nas mídias sociais: evolução e perspectivas na Alemanha, Brasil e Europa. *Revista EJEJ*, Belo Horizonte, v. 1, n. 1, p. 149-185, 2024. Disponível em: <https://revistaejef.tjmg.jus.br/index.php/revista-ejef/article/view/17>. Acesso em: 6 jan. 2026.

SARLET, Ingo Wolfgang. Liberdade de expressão e o problema da regulação do discurso do ódio nas mídias sociais. *Revista Estudos Institucionais*, v. 5, n. 3, p. 1207-1233, dez. 2019. Disponível em: <https://www.estudosinstitucionais.com/REI/article/view/428/443>. Acesso em: 6 jan. 2026.

SERRA JÚNIOR, Gentil Cutrim; ROCHA, Lourdes de Maria Leitão Nunes. A internet e os novos processos de articulação dos movimentos sociais. *Revista Katálisis*, Florianópolis, v. 16, n. 2, p. 205-213, jul./dez. 2013. Disponível em: <https://www.scielo.br/j/rk/a/5gsbm9b6QGtTCFdpTLs87G/>. Acesso em: 6 jan. 2026.

SILVA, Guilherme Amorim Campos da. A implantação do governo digital no Brasil: o modelo jurídico favorece o desenvolvimento tecnológico, a inclusão da população e o crescimento econômico? *Revista Pensamento Jurídico*. São Paulo, Brasil. V. 19, n1. P. 201-229, 2025. Disponível em <https://ojs.unialfa.com.br/index.php/pensamentojuridico/article/view/1083>. Acesso em 19 de agosto de 2025.

STROPPA, Tatiana; CARVALHO, Letícia Redis; GRINGS, Maria Gabriela; ROXO, Tatiana Bhering; OLIVEIRA, Samuel Rodrigues de. Direito digital, Seção 230 do CDA e artigo 19 do Marco Civil da Internet. *Consultor Jurídico (ConJur)*, 4 maio 2022. Disponível em: <https://www.conjur.com.br/2022-mai-04/direito-digital-secao-230-cda-artigo-19-marco-civil-internet/>. Acesso em: 6 jan. 2026.

SUPREMO TRIBUNAL FEDERAL. STF define parâmetros para responsabilização de plataformas por conteúdos de terceiros. Disponível em: <https://noticias.stf.jus.br/postsnoticias/stf-define-parametros-para-responsabilizacao-de-plataformas-por-conteudos-de-terceiros/>. Acesso em: 6 jan. 2026.

TAVARES, André Ramos. A nova Matrix. Direito (re)programado na civilização plataformizada. São Paulo: Etheria Editora, 2024. 342p.

TAVARES. André Ramos. APP's e plataformas on-line na intermediação econômica no Brasil. In *Revista de Direito Constitucional e Econômico*. Unialfa. N. 1, vol. 1, 2019.

TAVARES, André Ramos. Constituição em rede. *Revista Brasileira de Estudos Constitucionais – RBEC*, Belo Horizonte, ano 16, n.50, p. 53-70, jul/dez. 2022.

TAVARES, André Ramos. *Curso de Direito Constitucional*. 10ª ed. São Paulo: Saraiva, 2012. 1428p.





TAVARES. André Ramos. (coord). Vestígios do Futuro: 100 anos de Isaac Asimov. São Paulo: Etheria, 2020. Edição ebook.

TORRE, Luísa; JERÓNIMO, Pedro. Esfera pública e desinformação em contexto local. *Texto Livre: linguagem e tecnologia*, Belo Horizonte, v. 16, p. e41881, jul. 2023. Disponível em: <https://www.scielo.br/j/tl/a/fDvTzd8KD3vZ6cyvVKSnxNq/?lang=pt>. Acesso em: 6 jan. 2026.

WE ARE SOCIAL. *Digital 2024: 5 billion social media users*. 2024. Disponível em: <https://wearesocial.com/us/blog/2024/01/digital-2024-5-billion-social-media-users/>. Acesso em: 6 jan. 2026.

