

IX ENCONTRO VIRTUAL DO CONPEDI

**DIREITO, GOVERNANÇA E NOVAS TECNOLOGIAS
III**

Todos os direitos reservados e protegidos. Nenhuma parte destes anais poderá ser reproduzida ou transmitida sejam quais forem os meios empregados sem prévia autorização dos editores.

Diretoria - CONPEDI

Presidente - Profa. Dra. Samyra Haydêe Dal Farra Naspolini - FMU - São Paulo

Diretor Executivo - Prof. Dr. Orides Mezzaroba - UFSC - Santa Catarina

Vice-presidente Norte - Prof. Dr. Jean Carlos Dias - Cesupa - Pará

Vice-presidente Centro-Oeste - Prof. Dr. José Querino Tavares Neto - UFG - Goiás

Vice-presidente Sul - Prof. Dr. Leonel Severo Rocha - Unisinos - Rio Grande do Sul

Vice-presidente Sudeste - Profa. Dra. Rosângela Lunardelli Cavallazzi - UFRJ/PUCRio - Rio de Janeiro

Vice-presidente Nordeste - Prof. Dr. Raymundo Juliano Feitosa - UNICAP - Pernambuco

Representante Discente: Prof. Dr. Abner da Silva Jaques - UPM/UNIGRAN - Mato Grosso do Sul

Conselho Fiscal:

Prof. Dr. José Filomeno de Moraes Filho - UFMA - Maranhão

Prof. Dr. Caio Augusto Souza Lara - SKEMA/ESDHC/UFMG - Minas Gerais

Prof. Dr. Valter Moura do Carmo - UFERSA - Rio Grande do Norte

Prof. Dr. Fernando Passos - UNIARA - São Paulo

Prof. Dr. Edinilson Donisete Machado - UNIVEM/UENP - São Paulo

Secretarias

Relações Institucionais:

Prof. Dra. Claudia Maria Barbosa - PUCPR - Paraná

Prof. Dr. Heron José de Santana Gordilho - UFBA - Bahia

Profa. Dra. Daniela Marques de Moraes - UNB - Distrito Federal

Comunicação:

Prof. Dr. Robison Tramontina - UNOESC - Santa Catarina

Prof. Dr. Liton Lanes Pilau Sobrinho - UPF/Univali - Rio Grande do Sul

Prof. Dr. Lucas Gonçalves da Silva - UFS - Sergipe

Relações Internacionais para o Continente Americano:

Prof. Dr. Jerônimo Siqueira Tybusch - UFSM - Rio Grande do Sul

Prof. Dr. Paulo Roberto Barbosa Ramos - UFMA - Maranhão

Prof. Dr. Felipe Chiarello de Souza Pinto - UPM - São Paulo

Relações Internacionais para os demais Continentes:

Profa. Dra. Gina Vidal Marcilio Pompeu - UNIFOR - Ceará

Profa. Dra. Sandra Regina Martini - UNIRITTER / UFRGS - Rio Grande do Sul

Profa. Dra. Maria Claudia da Silva Antunes de Souza - UNIVALI - Santa Catarina

Educação Jurídica

Profa. Dra. Viviane Coêlho de Séllos Knoerr - Unicuritiba - PR

Prof. Dr. Rubens Beçak - USP - SP

Profa. Dra. Livia Gaigher Bosio Campello - UFMS - MS

Eventos:

Prof. Dr. Yuri Nathan da Costa Lannes - FDF - São Paulo

Profa. Dra. Norma Sueli Padilha - UFSC - Santa Catarina

Prof. Dr. Juraci Mourão Lopes Filho - UNICHRISTUS - Ceará

Comissão Especial

Prof. Dr. João Marcelo de Lima Assafim - UFRJ - RJ

Profa. Dra. Maria Creusa De Araújo Borges - UFPB - PB

Prof. Dr. Antônio Carlos Diniz Murta - Fumec - MG

Prof. Dr. Rogério Borba - UNIFACVEST - SC

D597

Direito, governança e novas tecnologias III [Recurso eletrônico on-line] organização CONPEDI

Coordenadores: Luiz Alberto Pereira Ribeiro; Regina Vera Villas Boas; Paulo Campanha Santana – Florianópolis: CONPEDI, 2026.

Inclui bibliografia

ISBN: 978-65-5274-590-3

Modo de acesso: www.conpedi.org.br em publicações

Tema: Constituição, Processo e Justiça Social: o papel da jurisdição constitucional na efetivação de direitos

2. Direito. 3. Governança e novas tecnologias. IX Encontro Virtual do CONPEDI (2: 2026: Florianópolis, Brasil).

CDU: 34



Conselho Nacional de Pesquisa
e Pós-Graduação em Direito Florianópolis
Santa Catarina – Brasil
www.conpedi.org.br

IX ENCONTRO VIRTUAL DO CONPEDI

DIREITO, GOVERNANÇA E NOVAS TECNOLOGIAS III

Apresentação

Texto de apresentação do livro do GT 52 - DIREITO, GOVERNANÇA E NOVAS TECNOLOGIAS III

O “IX Encontro Virtual do CONPEDI – Sociedade Científica do Direito no Brasil” foi realizado de forma virtual entre os dias 22 e 26 de junho de 2026. Este evento exemplar foi um marco de excelência acadêmica e colaboração científica, reunindo pesquisadores e estudiosos de diversas áreas do Direito.

Destacamos especialmente o Grupo de Trabalho – GT 52, intitulado “Direito, Governança e Novas Tecnologias III”, que se destacou pela profundidade e relevância dos temas abordados. Sob a coordenação dos Professores- Drs. Luiz Alberto P. Ribeiro (UEL e PUC/PR), Paulo Campanha (Mackenzie/Brasília) e Regina Vera Villas Bôas (PUC/SP), o GT proporcionou um espaço privilegiado para a discussão de questões fundamentais no campo do direito, governança e das novas tecnologias.

Neste GT foram apresentados trabalhos de elevada qualidade e importância crítica, sob os seguintes títulos:

1 MODERAÇÃO DE CONTEÚDO, GOVERNANÇA DE PLATAFORMAS DIGITAIS E DIREITOS FUNDAMENTAIS: PARÂMETROS CONSTITUCIONAIS PARA O BRASIL.

2 MERCADOS DE CARBONO E O NOVO SISTEMA BRASILEIRO DE COMÉRCIO DE EMISSÕES (SBCE).

6 GOVERNANÇA E REGULAÇÃO DA INFRAESTRUTURA PARA VEÍCULOS ELÉTRICOS EM CONDOMÍNIOS NO BRASIL: RISCOS, RESPONSABILIDADE CIVIL E DESIGUALDADE DE ACESSO

7 ENTRE O VIÉS HUMANO E O ALGORÍTMICO: DESAFIOS DA RACIONALIDADE PROCESSUAL NA ERA DA INTELIGÊNCIA ARTIFICIAL.

8 ENTRE A CELERIDADE E A EFICÁCIA: OS DESAFIOS DA AUTOMAÇÃO NO INSTITUTO NACIONAL DO SEGURO SOCIAL.

9 EMPRESAS TRANSNACIONAIS, RELAÇÕES DE TRABALHO E GOVERNANÇA DIGITAL NAS CONTRATAÇÕES PÚBLICAS: ANÁLISE CRÍTICA À LUZ DA LEI Nº 14.133/2021.

10 ECONOMIA DOS DADOS E GOVERNANÇA ALGORÍTMICA: PERFILAMENTO E VULNERABILIDADE DIGITAL DO CONSUMIDOR NA SOCIEDADE INFORMACIONAL.

11 DEMÔNIO DE LAPLACE, PREVISIBILIDADE E CONTROLE: PROBLEMÁTICAS EPISTEMOLÓGICAS, ÉTICAS E DE JUSTIÇA NAS DECISÕES JURÍDICO ADMINISTRATIVAS BASEADAS EM BIG DATA, ALGORITMOS E INTELIGÊNCIA ARTIFICIAL.

12 DIREITO FUNDAMENTAL À PROTEÇÃO DE DADOS E A TUTELA ESPECIAL NO TRATAMENTO DOS DADOS SENSÍVEIS: DA ESFERA DA PRIVACIDADE À DIGNIDADE DA PESSOA HUMANA E AUTODETERMINAÇÃO INFORMATIVA.

13 DIREITO PROCESSUAL E HOMOTRANSFOBIA ALGORÍTMICA: POR UM DEVER DE CUIDADO JURISDICIONALIZADO.

16 AUTOMAÇÃO DAS DECISÕES JUDICIAIS E OS LIMITES CONSTITUCIONAIS DA INTELIGÊNCIA ARTIFICIAL NO ESTADO DEMOCRÁTICO DE DIREITO.

17 AUTORIA E ORIGINALIDADE NA ERA DA INTELIGÊNCIA ARTIFICIAL: DESAFIOS JURÍDICOS DA PROTEÇÃO AUTORAL DIANTE DAS CRIAÇÕES AUTOMATIZADAS E DA AUTOPUBLICAÇÃO DIGITAL.

18 CAPITALISMO DE VIGILÂNCIA E DIREITOS FUNDAMENTAIS SOCIAIS: OS LIMITES JURÍDICOS DA EXPLORAÇÃO ECONÔMICA DE DADOS PESSOAIS NO BRASIL.

19 CAPITALISMO DE VIGILÂNCIA E HIPERCONSUMO: REFLEXÕES SOBRE A CRISE DA AUTONOMIA NA SOCIEDADE DIGITAL.

20 ADVOCACIA PÚBLICA E GOVERNANÇA INFORMACIONAL NA ADMINISTRAÇÃO PÚBLICA: PROTEÇÃO DE DADOS, POLÍTICAS PÚBLICAS E CONTROLE DE JURIDICIDADE.

A qualidade dos trabalhos apresentados neste GT 52 foi notável, refletindo o compromisso dos participantes com a pesquisa de alto nível e a inovação acadêmica. As contribuições dos estudiosos trouxeram insights significativos e promoveram um debate enriquecedor sobre os desafios contemporâneos e as perspectivas futuras nessas áreas cruciais que envolvem o Direito.

O IX Encontro Virtual do CONPEDI não apenas consolidou seu papel como um canal de referência no cenário acadêmico nacional e internacional, como também reafirmou o compromisso com a qualidade científica e a excelência na produção do conhecimento jurídico.

Os coordenadores do GT 52 deixam aqui abraços fraternos e cumprimentos a todos os participantes do CONPEDI e aos leitores dos presentes textos.

Prof. Dr. Luiz Alberto P. Ribeiro (UEL e PUC/PR)

Prof. Dr. Paulo Campanha (Mackenzie/Brasília)

Profa. Dra. Regina Vera Villas Bôas (PUC/SP)

MODERAÇÃO DE CONTEÚDO, GOVERNANÇA DE PLATAFORMAS DIGITAIS E DIREITOS FUNDAMENTAIS: PARÂMETROS CONSTITUCIONAIS PARA O BRASIL

CONTENT MODERATION, DIGITAL PLATFORM GOVERNANCE AND FUNDAMENTAL RIGHTS: CONSTITUTIONAL PARAMETERS FOR BRAZIL

James Euzébio Pedro Neto

Resumo

O artigo analisa a moderação de conteúdo por plataformas digitais como problema de governança constitucional no Brasil. Parte-se do problema de saber quais parâmetros devem orientar a atuação das plataformas na remoção, restrição, recomendação ou redução de alcance de conteúdos, considerando a tensão entre liberdade de expressão, proteção contra danos, responsabilidade civil e direitos fundamentais. A hipótese sustenta que a moderação somente será constitucionalmente legítima quando estruturada por modelo transparente, proporcional, responsivo, auditável e comprometido simultaneamente com a liberdade de expressão e com a proteção da dignidade, igualdade, privacidade, proteção de dados, direitos de crianças e adolescentes e integridade democrática. A pesquisa adota método dedutivo, abordagem qualitativa e procedimento bibliográfico-documental, com análise da Constituição Federal, Marco Civil da Internet, Lei Geral de Proteção de Dados, legislação recente sobre proteção de crianças e adolescentes em ambientes digitais, jurisprudência constitucional, documentos internacionais e literatura especializada. Conclui-se que a governança de plataformas deve rejeitar tanto a autorregulação opaca quanto a intervenção estatal desproporcional, adotando parâmetros de transparência, devido processo, proporcionalidade, avaliação de riscos, auditabilidade, participação multissetorial e reparação.

Palavras-chave: Governança digital, Liberdade de expressão, Moderação de conteúdo, Plataformas digitais, Responsabilidade civil

Abstract/Resumen/Résumé

This article analyzes content moderation by digital platforms as a problem of constitutional

the Internet, the General Data Protection Law, recent legislation on the protection of children and adolescents in digital environments, constitutional case law, international documents, and specialized literature. It concludes that platform governance must reject both opaque self-regulation and disproportionate state intervention, adopting parameters such as transparency, due process, proportionality, risk assessment, auditability, multistakeholder participation, and effective remedies.

Keywords/Palabras-claves/Mots-clés: Civil liability, Content moderation, Digital governance, Digital platforms, Freedom of expression

1. INTRODUÇÃO

A moderação de conteúdo realizada por plataformas digitais tornou-se uma das questões centrais do constitucionalismo contemporâneo. Redes sociais, mecanismos de busca, aplicativos de compartilhamento de vídeo, serviços de mensagens, marketplaces e ambientes de interação online deixaram de ser simples espaços privados de comunicação para assumir função estrutural na circulação de informações, na formação da opinião pública, no exercício da cidadania, na organização de movimentos sociais e na dinâmica cotidiana das relações econômicas, políticas e culturais. Nesse cenário, decisões privadas sobre remoção, redução de alcance, priorização, impulsionamento, monetização, recomendação ou bloqueio de conteúdos podem afetar diretamente a liberdade de expressão, o direito à informação, a dignidade da pessoa humana, a igualdade, a privacidade, a proteção de crianças e adolescentes e a própria integridade do processo democrático.

O problema não se limita à existência de conteúdos ilícitos ou nocivos na internet. A questão constitucional mais relevante consiste em compreender quem decide, com quais critérios, mediante quais procedimentos e sob quais formas de controle. A moderação de conteúdo não é atividade meramente técnica. Ela envolve escolhas normativas sobre o que deve circular, o que deve ser removido, o que merece maior visibilidade, o que deve ser rotulado como enganoso, o que deve ser restringido por risco à integridade de grupos vulneráveis e como devem ser tratados discursos situados em zonas cinzentas entre crítica legítima, desinformação, sátira, discurso ofensivo e violação de direitos. Por isso, a governança das plataformas digitais passou a ocupar espaço intermediário entre a autonomia privada, a regulação estatal e a proteção constitucional de direitos fundamentais.

A relevância jurídica do tema aumenta diante da assimetria entre usuários e plataformas. As empresas que administram ambientes digitais definem termos de uso, políticas de comunidade, parâmetros algorítmicos de recomendação, mecanismos de denúncia e instâncias internas de revisão. Embora tais regras sejam formalmente privadas, seus efeitos ultrapassam a esfera contratual, pois atingem milhões de pessoas, condicionam a visibilidade de discursos públicos e podem influenciar processos eleitorais, políticas públicas, reputações individuais, proteção de minorias e acesso a informações socialmente relevantes. Como observa Gillespie (2018), as plataformas atuam como curadoras e guardiãs de conteúdos, ainda que frequentemente apresentem suas decisões como neutras ou inevitáveis do ponto de vista técnico.

O problema de pesquisa que orienta este artigo pode ser formulado nos seguintes termos: quais parâmetros constitucionais devem orientar a moderação de conteúdo por plataformas digitais no Brasil, de modo a compatibilizar liberdade de expressão, proteção contra danos, governança tecnológica e direitos fundamentais? A hipótese adotada é a de que a moderação de conteúdo somente será constitucionalmente legítima quando estruturada a partir de um modelo de governança transparente, proporcional, responsivo, auditável e comprometido com a proteção simultânea da liberdade de expressão e dos demais direitos fundamentais. Nem a ausência de regulação, que deixa direitos individuais e coletivos dependentes de decisões empresariais opacas, nem a intervenção estatal desproporcional, que pode converter a proteção contra abusos em mecanismo de censura, oferecem solução adequada ao Estado Democrático de Direito.

O objetivo geral consiste em analisar a moderação de conteúdo como problema de governança constitucional de plataformas digitais. Como objetivos específicos, pretende-se: examinar o papel das plataformas como infraestruturas privadas do discurso público; compreender a moderação de conteúdo como forma de regulação privada; identificar os direitos fundamentais em tensão; analisar o regime jurídico brasileiro, especialmente a Constituição Federal, o Marco Civil da Internet, a Lei Geral de Proteção de Dados, a proteção de crianças e adolescentes em ambientes digitais e a recente atuação do Supremo Tribunal Federal sobre responsabilidade de plataformas; e propor parâmetros constitucionais para uma governança digital compatível com a democracia.

A pesquisa adota método dedutivo, com abordagem qualitativa e procedimento bibliográfico-documental. São analisadas normas constitucionais e infraconstitucionais brasileiras, documentos internacionais sobre empresas e direitos humanos, diretrizes de governança de plataformas digitais, experiência regulatória estrangeira e literatura especializada sobre moderação de conteúdo e poder privado na esfera digital. O artigo não pretende defender um modelo fechado de regulação, mas propor critérios normativos capazes de orientar tanto a atuação das plataformas quanto a formulação de políticas públicas e decisões judiciais no Brasil.

A justificativa do estudo decorre do fato de que a moderação de conteúdo envolve uma tensão permanente. De um lado, a liberdade de expressão exige ambiente aberto, plural e resistente ao controle arbitrário de discursos. De outro, a circulação massiva de conteúdo pode amplificar discursos de ódio, assédio, exploração de crianças e adolescentes, fraudes, manipulações informacionais, ataques coordenados a instituições democráticas e violações à privacidade. O desafio constitucional consiste em evitar dois extremos: a irresponsabilidade

privada, que permite que danos se propaguem sob o argumento de neutralidade técnica, e o controle estatal excessivo, que pode comprometer o espaço público e a crítica democrática.

2. PLATAFORMAS DIGITAIS COMO INFRAESTRUTURAS DO DISCURSO PÚBLICO

As plataformas digitais ocupam posição peculiar no ecossistema comunicacional contemporâneo. Elas não são apenas intermediárias passivas entre emissores e receptores de mensagens. Seu desenho tecnológico influencia o que aparece, o que desaparece, o que se torna viral, o que permanece invisível e o que é convertido em objeto de engajamento. A arquitetura das plataformas organiza relações sociais mediante regras de cadastro, mecanismos de recomendação, métricas de popularidade, filtros automatizados, sistemas de denúncia, políticas de monetização e modelos de publicidade direcionada. O discurso público digital, portanto, não circula em espaço neutro, mas em ambientes governados por decisões técnicas e empresariais.

Essa constatação desloca o debate tradicional sobre liberdade de expressão. A teoria clássica preocupava-se sobretudo com a censura estatal e com a proteção do indivíduo contra restrições impostas pelo poder público. Essa preocupação permanece indispensável, especialmente em sociedades democráticas marcadas por disputas políticas intensas. Contudo, no ambiente digital, parte relevante da capacidade de falar, ser ouvido e alcançar públicos depende também de agentes privados. Balkin (2018) descreve esse fenômeno como característica da sociedade algorítmica, na qual empresas de tecnologia exercem formas de governança privada sobre a circulação de discursos, dados e visibilidade.

O poder das plataformas manifesta-se de diferentes formas. A remoção de conteúdo é apenas a face mais visível da moderação. Há também redução de alcance, desmonetização, suspensão temporária de contas, bloqueio definitivo, rotulagem de informação, restrição etária, limitação de compartilhamento, alteração de ranqueamento, priorização de conteúdos pagos, recomendação automatizada e intervenção em tendências. Muitas dessas medidas não eliminam formalmente a fala, mas afetam sua circulação. Em ambiente marcado pela escassez de atenção, reduzir visibilidade pode equivaler, na prática, a diminuir drasticamente a possibilidade de participação pública.

A linguagem contratual dos termos de uso é insuficiente para explicar essa realidade. Embora o usuário aceite formalmente as condições da plataforma, a relação não se resume a contrato entre iguais. As principais plataformas possuem escala global, poder econômico concentrado, grande volume de dados e capacidade técnica inacessível ao usuário comum. Além disso, a saída da plataforma nem sempre é opção real, pois determinadas redes se tornam

ambientes relevantes de sociabilidade, trabalho, divulgação profissional, mobilização política e acesso à informação. A dependência social de tais ambientes reforça a necessidade de examinar suas regras à luz de parâmetros públicos de legitimidade.

Klonick (2018) identifica as plataformas como novos governantes do discurso online, pois elaboram normas, aplicam sanções, desenvolvem procedimentos internos e decidem sobre conflitos envolvendo expressão, segurança e reputação. Essa governança privada não se confunde com jurisdição estatal, mas produz efeitos semelhantes para os usuários: regras são impostas, denúncias são julgadas, conteúdos são removidos e contas podem ser suspensas. O problema constitucional reside justamente na ausência de garantias equivalentes às que se esperam de decisões que afetam direitos, como transparência, fundamentação, contraditório, proporcionalidade e possibilidade de revisão.

A centralidade das plataformas também decorre do modelo econômico que sustenta a circulação de conteúdos. Grande parte dos serviços digitais opera mediante coleta de dados, publicidade direcionada e maximização de engajamento. A visibilidade de conteúdo não decorre apenas de relevância pública ou qualidade informacional, mas também de métricas que favorecem permanência do usuário, interação, compartilhamento e reação emocional. Esse modelo pode estimular conteúdos polarizantes, sensacionalistas ou desinformativos, pois a lógica de engajamento nem sempre coincide com a proteção do debate democrático. A moderação, nesse contexto, não é atividade periférica destinada a corrigir excessos pontuais, mas componente estrutural de governança do próprio modelo de negócios.

A dimensão pública das plataformas não significa que elas devam ser integralmente equiparadas ao Estado. Empresas privadas possuem liberdade econômica, autonomia organizacional e dever de gerir riscos de seus serviços. Contudo, quando suas decisões afetam direitos fundamentais em larga escala, a autonomia privada não pode ser compreendida como espaço imune a qualquer controle jurídico. O constitucionalismo contemporâneo admite que direitos fundamentais irradiem efeitos sobre relações privadas, sobretudo quando há assimetria de poder e relevância social da atividade desempenhada. A governança de plataformas deve ser examinada a partir dessa perspectiva.

O reconhecimento das plataformas como infraestruturas do discurso público permite superar uma falsa alternativa. Não se trata de escolher entre completa liberdade empresarial e estatização do espaço digital. A questão é construir um arranjo institucional no qual plataformas tenham deveres de transparência, diligência, coerência decisória, proteção de grupos vulneráveis e respeito à liberdade de expressão, ao mesmo tempo em que o Estado seja

impedido de usar a regulação como instrumento de controle político do debate público. A governança adequada deve limitar o poder privado sem criar poder público arbitrário.

3. MODERAÇÃO DE CONTEÚDO E GOVERNANÇA PRIVADA

A moderação de conteúdo pode ser definida como o conjunto de práticas por meio das quais plataformas estabelecem, interpretam e aplicam regras sobre conteúdos, comportamentos e contas em seus ambientes digitais. Ela envolve decisões humanas, automatizadas ou híbridas, baseadas em termos de uso, políticas comunitárias, legislação aplicável, pressões sociais, decisões judiciais, interesses comerciais e avaliações de risco. Seu objetivo declarado costuma ser preservar segurança, integridade e qualidade do serviço, mas seus efeitos alcançam diretamente a esfera de direitos dos usuários.

Gillespie (2018) demonstra que a moderação é inevitável para plataformas de grande escala. Ambientes sem qualquer gestão de conteúdo tendem a ser ocupados por spam, violência, exploração, assédio e manipulação. A promessa de neutralidade absoluta não se sustenta empiricamente. Toda plataforma decide, de algum modo, o que pode permanecer, o que deve ser removido e o que será promovido. A diferença está no grau de explicitação dessas escolhas e na existência de controles sobre sua aplicação. A questão democrática não é eliminar a moderação, mas submetê-la a parâmetros de legitimidade.

As decisões de moderação enfrentam dificuldades próprias. Primeiro, o volume de conteúdo é gigantesco, o que exige automação e priorização. Segundo, a avaliação do contexto é complexa, pois a mesma palavra ou imagem pode assumir sentidos diversos conforme ironia, denúncia, jornalismo, pesquisa, manifestação artística ou violência real. Terceiro, regras globais precisam lidar com culturas jurídicas distintas e sensibilidades locais. Quarto, a pressão por respostas rápidas pode reduzir qualidade decisória. Quinto, a moderação pode ser instrumentalizada por campanhas coordenadas de denúncias abusivas contra adversários políticos, jornalistas, defensores de direitos humanos ou minorias.

A automação amplia esses desafios. Sistemas algorítmicos podem identificar padrões de nudez, violência, spam ou comportamentos coordenados, mas frequentemente possuem dificuldade para compreender contexto, proporcionalidade, linguagem figurada e relevância pública. A inteligência artificial pode apoiar a triagem, mas não elimina a necessidade de governança humana, revisão e responsabilização. Quando decisões automatizadas removem conteúdo ou restringem conta sem explicação adequada, o usuário pode ficar submetido a uma sanção invisível, com baixa capacidade de contestação.

Roberts (2019) chama atenção para outro aspecto normalmente ignorado: a moderação de conteúdo depende de trabalho humano muitas vezes invisível, precarizado e exposto a imagens e narrativas violentas. A ideia de que plataformas funcionam apenas por algoritmos oculta uma cadeia de trabalhadores encarregados de classificar conteúdos traumáticos, aplicar regras ambíguas e decidir sob pressão. Isso demonstra que a governança de plataformas envolve não apenas direitos dos usuários, mas também condições institucionais e laborais de quem executa a moderação. A qualidade da decisão depende de treinamento, tempo, suporte psicológico, supervisão e clareza normativa.

A governança privada das plataformas é marcada por déficit de transparência. Usuários frequentemente não sabem por que determinado conteúdo foi removido, por que uma publicação teve alcance reduzido ou por que uma conta foi suspensa. Mesmo quando há políticas publicadas, sua aplicação concreta pode ser inconsistente. Termos como “conteúdo perigoso”, “desinformação”, “comportamento inautêntico”, “discurso de ódio” ou “violação de integridade” podem ser necessários, mas exigem definição operacional clara. Sem critérios compreensíveis, a moderação se torna imprevisível e compromete tanto a liberdade de expressão quanto a proteção contra danos.

Também há risco de submoderação. A crítica ao excesso de remoções não pode ocultar os danos produzidos pela omissão. Discurso de ódio, perseguição coordenada, ameaça, divulgação não consensual de imagens íntimas, exploração sexual, fraude, incitação à violência e ataques contra instituições democráticas podem gerar danos graves e irreversíveis. Quando uma plataforma conhece riscos evidentes e não adota providências proporcionais, sua inércia pode contribuir para a violação de direitos. A proteção da liberdade de expressão não exige tolerância ilimitada a práticas que silenciam vítimas, intimidam grupos vulneráveis ou distorcem deliberadamente o debate público.

A moderação adequada deve, portanto, lidar com dupla exigência: evitar remoções arbitrárias e evitar omissões negligentes. Essa dupla exigência diferencia uma abordagem constitucional de uma abordagem puramente empresarial. Para a empresa, a decisão pode ser orientada por risco reputacional, custo operacional ou manutenção de engajamento. Para o direito constitucional, a decisão deve ser avaliada a partir de direitos, garantias e impactos sociais. O dever jurídico não é simplesmente remover mais ou remover menos, mas construir procedimentos que permitam decisões proporcionais, justificadas e revisáveis.

Nesse sentido, os Princípios Orientadores das Nações Unidas sobre Empresas e Direitos Humanos oferecem referência relevante. O documento estabelece que empresas devem respeitar direitos humanos, prevenir impactos negativos, realizar diligência devida e

disponibilizar mecanismos de reparação ou reclamação compatíveis com a gravidade dos riscos envolvidos (UNITED NATIONS, 2011). Embora não tenha sido elaborado especificamente para plataformas digitais, seu marco “proteger, respeitar e reparar” permite compreender que empresas com poder social expressivo não podem tratar violações de direitos como externalidades inevitáveis do negócio.

A UNESCO, em suas Diretrizes para a Governança de Plataformas Digitais, também propõe abordagem baseada em direitos humanos, transparência, prestação de contas, participação multissetorial e avaliação de riscos (UNESCO, 2023). A contribuição é importante porque rejeita tanto a autorregulação opaca quanto a regulação estatal arbitrária. A governança de plataformas deve envolver Estado, empresas, sociedade civil, academia, imprensa, comunidades técnicas e usuários, com especial atenção a grupos vulneráveis. A moderação de conteúdo deixa, assim, de ser assunto interno de empresas e passa a ser questão de interesse público.

4. DIREITOS FUNDAMENTAIS EM TENSÃO NA MODERAÇÃO DE CONTEÚDO

A moderação de conteúdo envolve colisões entre direitos fundamentais. A liberdade de expressão ocupa posição central, pois permite crítica, participação política, circulação de ideias, fiscalização do poder e desenvolvimento da personalidade. No Brasil, a Constituição protege a livre manifestação do pensamento, a liberdade de expressão intelectual, artística, científica e de comunicação, bem como o acesso à informação (BRASIL, 1988). Esses direitos são indispensáveis à democracia e devem ser interpretados de modo robusto, especialmente quando se trata de debate público, crítica a agentes estatais e circulação de informações de interesse coletivo.

Contudo, a liberdade de expressão não existe isoladamente. A mesma Constituição protege dignidade da pessoa humana, igualdade, intimidade, vida privada, honra, imagem, proteção de dados, direitos de crianças e adolescentes, pluralismo político e normalidade democrática (BRASIL, 1988). A colisão não autoriza hierarquia automática. Em alguns casos, a expressão deve ser protegida mesmo quando incômoda, ofensiva ou minoritária. Em outros, o conteúdo pode configurar ameaça, discriminação, exploração, fraude, violação de privacidade ou ataque coordenado capaz de restringir a liberdade de outros. O desafio é distinguir crítica legítima de dano constitucionalmente relevante.

A dignidade da pessoa humana funciona como eixo de interpretação. Sarlet (2015) destaca que a dignidade não é fórmula vazia, mas fundamento que exige respeito à pessoa como

sujeito de direitos, impedindo sua redução a objeto de humilhação, discriminação ou instrumentalização. Em plataformas digitais, conteúdos que desumanizam grupos, incentivam violência ou exploram vulnerabilidades podem afetar não apenas indivíduos isolados, mas o reconhecimento social de comunidades inteiras. A moderação de conteúdo deve considerar esses impactos, sem transformar a dignidade em justificativa genérica para censurar críticas legítimas.

A igualdade também é decisiva. Grupos historicamente vulnerabilizados sofrem de modo desproporcional com assédio, discurso de ódio, exposição abusiva, ataques coordenados e campanhas de desinformação. Mulheres, pessoas negras, populações indígenas, pessoas LGBTQIA+, pessoas com deficiência, crianças, adolescentes, jornalistas, pesquisadores, defensores de direitos humanos e autoridades eleitorais podem ser alvos de estratégias destinadas a silenciar participação pública. A ausência de moderação adequada pode produzir efeito inibidor sobre a expressão dessas pessoas, restringindo a pluralidade que a liberdade de expressão pretende proteger.

A privacidade e a proteção de dados também se relacionam com a moderação. Plataformas coletam e processam dados para recomendar conteúdos, identificar comportamentos, personalizar publicidade e detectar violações de regras. A Lei Geral de Proteção de Dados estabelece princípios como finalidade, adequação, necessidade, transparência, segurança, prevenção, responsabilização e prestação de contas (BRASIL, 2018). Esses princípios devem informar sistemas de moderação, especialmente quando decisões são baseadas em perfis comportamentais, sinais algorítmicos ou tratamento de dados sensíveis. A proteção contra conteúdo danoso não pode justificar vigilância desproporcional ou uso opaco de dados pessoais.

A proteção de crianças e adolescentes impõe atenção específica. Ambientes digitais podem expor pessoas em desenvolvimento a exploração sexual, aliciamento, publicidade abusiva, desafios perigosos, cyberbullying, vícios de design, coleta indevida de dados e conteúdos inadequados. A legislação brasileira passou a incorporar preocupação mais direta com a proteção de crianças e adolescentes em ambientes digitais, sinalizando a necessidade de deveres específicos de cuidado (BRASIL, 2025). A moderação de conteúdo, nesse campo, deve observar prioridade absoluta, proteção integral e participação adequada conforme idade e maturidade, sem ignorar o direito de crianças e adolescentes à informação, à expressão e à convivência digital segura.

A democracia representa outro direito ou valor constitucional em tensão. Plataformas digitais podem ampliar participação política, dar voz a grupos antes excluídos e permitir

mobilização social. Ao mesmo tempo, podem facilitar campanhas coordenadas de desinformação, manipulação de audiência, assédio contra instituições, uso de contas inautênticas e microdirecionamento opaco de mensagens políticas. A moderação de conteúdo em contexto eleitoral ou de crise institucional exige prudência reforçada. O combate à manipulação não pode servir para blindar autoridades contra crítica, mas a proteção da crítica não pode justificar tolerância a práticas coordenadas de fraude informacional.

O direito à informação também possui dupla dimensão. Usuários devem ter acesso a informações plurais e confiáveis, mas não cabe ao Estado ou às plataformas fixar uma verdade oficial sobre todo debate público. Em temas científicos, eleitorais, sanitários ou de segurança, a desinformação pode causar danos concretos; ainda assim, respostas devem ser proporcionais e baseadas em critérios verificáveis. Rotulagem, redução de alcance, promoção de fontes confiáveis e contextualização podem, em certos casos, ser alternativas menos restritivas do que remoção. A escolha da medida deve considerar gravidade, alcance, intenção, probabilidade de dano e existência de interesse público.

A colisão de direitos exige procedimento. Em vez de resolver a moderação por fórmulas abstratas, é necessário estabelecer garantias mínimas: regras públicas, notificação ao usuário, fundamentação compreensível, possibilidade de recurso, revisão humana em casos sensíveis, preservação de conteúdo para fins probatórios quando necessário e relatórios periódicos de transparência. A proteção de direitos fundamentais depende tanto do resultado quanto do modo como a decisão é tomada. Uma remoção correta, mas sem explicação ou recurso, ainda pode fragilizar a confiança no sistema.

A perspectiva constitucional também impõe análise de proporcionalidade. A medida de moderação deve ser adequada para reduzir risco, necessária diante das alternativas disponíveis e proporcional em sentido estrito quando comparados os benefícios de proteção e os custos à liberdade de expressão. Remover definitivamente uma conta é medida mais grave do que restringir um conteúdo específico; reduzir alcance pode ser menos severo que exclusão; rotular pode ser adequado quando há controvérsia informacional; preservar conteúdo com restrição de acesso pode ser necessário para investigação. A moderação constitucionalmente orientada deve calibrar respostas.

5. REGIME JURÍDICO BRASILEIRO E EXPERIÊNCIAS DE GOVERNANÇA

O regime brasileiro de plataformas digitais parte da Constituição Federal e de leis infraconstitucionais que precisam ser interpretadas de maneira sistemática. A Constituição

fornece o núcleo valorativo: liberdade de expressão, vedação à censura, dignidade da pessoa humana, igualdade, privacidade, proteção de dados, proteção integral de crianças e adolescentes, acesso à informação e defesa da democracia (BRASIL, 1988). Esses parâmetros impedem tanto a censura estatal quanto a indiferença diante de violações graves de direitos. A governança de plataformas deve ser construída nesse espaço de equilíbrio.

O Marco Civil da Internet representou avanço relevante ao estabelecer princípios, garantias, direitos e deveres para o uso da internet no Brasil (BRASIL, 2014). A lei consagrou a liberdade de expressão como fundamento da disciplina da internet, protegeu a privacidade, regulou guarda de registros e disciplinou a responsabilidade de provedores por conteúdos de terceiros. Seu art. 19 adotou, como regra geral, a responsabilização civil do provedor de aplicações por conteúdo de terceiro após descumprimento de ordem judicial específica. A lógica original buscava evitar remoções privadas excessivas e proteger a liberdade de expressão contra notificações abusivas.

A experiência posterior revelou, contudo, tensões relevantes. A exigência de ordem judicial pode ser importante para evitar censura privada em conteúdos controversos, mas pode ser insuficiente diante de danos evidentes, massivos e de rápida propagação. Conteúdos envolvendo exploração sexual, ataques coordenados, violência contra grupos vulneráveis, fraudes graves ou ameaças à democracia podem demandar respostas céleres. Por outro lado, substituir a ordem judicial por obrigação genérica de remoção imediata pode estimular excesso de retirada, pois plataformas tenderiam a remover preventivamente qualquer conteúdo denunciado para reduzir risco jurídico.

O Supremo Tribunal Federal passou a discutir a constitucionalidade e o alcance do art. 19 do Marco Civil da Internet nos Temas 533 e 987 da repercussão geral. Em 2025, a Corte definiu parâmetros para responsabilização de plataformas por conteúdos de terceiros e reconheceu a necessidade de compatibilizar o regime legal com proteção suficiente de direitos fundamentais e da democracia (SUPREMO TRIBUNAL FEDERAL, 2025). Essa discussão marca ponto decisivo: o modelo brasileiro já não pode ser compreendido apenas como imunidade condicionada à ordem judicial, mas como regime que deve articular liberdade de expressão, deveres de cuidado, devido processo e prevenção de danos.

A Lei Geral de Proteção de Dados acrescenta outra camada normativa. A moderação de conteúdo envolve tratamento de dados pessoais para detectar comportamentos, classificar conteúdos, identificar riscos, verificar contas, prevenir fraudes e aplicar sanções. A LGPD exige bases jurídicas adequadas, transparência, segurança, minimização, prevenção e responsabilização (BRASIL, 2018). Plataformas não podem usar a proteção contra conteúdos

ilícitos como justificativa para coleta ilimitada de dados, perfilamento opaco ou decisões automatizadas sem possibilidade de explicação. A governança de conteúdo e a governança de dados são dimensões interdependentes.

A proteção de crianças e adolescentes em ambientes digitais reforça a necessidade de deveres específicos. A Lei nº 15.211/2025 dispõe sobre essa proteção e confirma que certos grupos demandam desenho regulatório próprio diante de riscos digitais (BRASIL, 2025). No contexto da moderação, isso significa que plataformas devem adotar políticas proporcionais à idade, mecanismos de denúncia acessíveis, respostas rápidas a conteúdos exploratórios e salvaguardas contra designs que amplifiquem riscos. A proteção integral não autoriza vigilância indiscriminada, mas impõe deveres reforçados de prevenção e de cuidado.

A experiência europeia oferece referência comparada relevante. O Digital Services Act instituiu obrigações graduadas para serviços intermediários e plataformas, com ênfase em transparência, deveres de diligência, mecanismos de notificação e ação, fundamentação de decisões de moderação, sistemas internos de reclamação, auditorias e avaliação de riscos sistêmicos para plataformas de maior porte (EUROPEAN UNION, 2022). O modelo europeu não deve ser transplantado automaticamente para o Brasil, mas demonstra que a governança democrática de plataformas pode combinar proteção da liberdade de expressão, deveres procedimentais e responsabilização proporcional ao risco.

As Diretrizes da UNESCO também contribuem para o debate brasileiro ao defender governança multissetorial, proteção de direitos humanos, transparência, acessibilidade de informações, prestação de contas e avaliação de riscos (UNESCO, 2023). Essa abordagem é especialmente útil porque evita a visão exclusivamente punitiva da regulação. Governar plataformas não significa apenas aplicar sanções após danos, mas construir processos que reduzam riscos, ampliem participação social, permitam auditoria independente e fortaleçam a capacidade dos usuários de compreender e contestar decisões.

A comparação internacional evidencia que o Brasil precisa superar a fragmentação normativa. Há Constituição, Marco Civil, LGPD, normas específicas de proteção de crianças e adolescentes, jurisprudência constitucional e projetos legislativos em debate, mas ainda há necessidade de parâmetros mais claros para deveres de transparência, avaliação de riscos, mecanismos de recurso, relatórios auditáveis e procedimentos de resposta a danos graves. A ausência de critérios previsíveis prejudica usuários, empresas, Judiciário e órgãos de controle, pois cada conflito tende a ser resolvido de modo casuístico.

A construção brasileira deve, entretanto, observar cautelas próprias. O país possui histórico de desigualdade, violência política, concentração econômica, judicialização intensa e

riscos de instrumentalização de discursos de proteção para restringir crítica. Por isso, qualquer modelo regulatório deve conter salvaguardas contra censura estatal, captura regulatória, remoções abusivas e discricionariedade empresarial. A governança de plataformas precisa ser constitucionalmente densa: forte o suficiente para proteger direitos, limitada o suficiente para preservar liberdade democrática.

6. PARÂMETROS CONSTITUCIONAIS PARA A GOVERNANÇA DE PLATAFORMAS DIGITAIS

A partir das premissas anteriores, é possível propor parâmetros constitucionais para orientar a moderação de conteúdo no Brasil. O primeiro parâmetro é a centralidade simultânea da liberdade de expressão e da proteção contra danos. A liberdade de expressão não deve ser tratada como obstáculo à governança, mas como seu ponto de partida. Toda intervenção sobre conteúdo deve ser justificada. Ao mesmo tempo, a proteção da expressão não pode ser convertida em autorização para práticas que violem dignidade, igualdade, privacidade, segurança de crianças e adolescentes ou integridade democrática. O modelo adequado deve proteger a liberdade que permite participação plural e conter abusos que silenciam ou manipulam essa participação.

O segundo parâmetro é a transparência substantiva. Não basta publicar termos de uso extensos, genéricos e de difícil compreensão. As plataformas devem disponibilizar regras claras, exemplos de aplicação, critérios de gravidade, tipos de sanção, dados sobre remoções, bases de denúncias, taxas de reversão em recursos e informações sobre uso de automação. A transparência deve permitir que usuários compreendam o que é proibido, por que determinado conteúdo foi afetado e como contestar a decisão. Relatórios agregados devem ser úteis para pesquisadores, autoridades e sociedade civil avaliarem padrões de moderação.

O terceiro parâmetro é o devido processo procedimental. Decisões de remoção, suspensão, desmonetização ou redução significativa de alcance devem ser acompanhadas, sempre que possível, de notificação, fundamentação e via de recurso. Em casos de urgência ou risco grave, a plataforma pode agir imediatamente, mas deve permitir revisão posterior. O contraditório digital não exige reprodução integral do processo judicial, mas demanda garantias mínimas compatíveis com a escala e com a gravidade da sanção. Quanto maior o impacto sobre direitos, maior deve ser a densidade procedimental.

O quarto parâmetro é a proporcionalidade das medidas. A moderação deve operar em escala graduada. Conteúdos ilícitos evidentes e gravemente lesivos podem justificar remoção rápida e preservação de provas. Conteúdos enganosos, mas de interesse público, podem exigir

rotulagem, contextualização ou redução de recomendação, em vez de exclusão imediata. Contas reincidentes em fraude, assédio ou manipulação coordenada podem justificar sanções mais severas. A proporcionalidade evita tanto omissão quanto excesso, permitindo respostas adequadas à natureza do risco.

O quinto parâmetro é a responsabilidade baseada em risco e capacidade. Plataformas de grande porte, com maior poder de amplificação, monetização e coleta de dados, devem assumir deveres mais intensos de avaliação, mitigação e transparência. Serviços pequenos ou comunitários não podem ser submetidos às mesmas obrigações operacionais de grandes plataformas globais, sob pena de reforçar concentração de mercado. A regulação deve ser proporcional à escala, à função social, ao número de usuários, ao modelo de negócios e ao potencial de dano sistêmico.

O sexto parâmetro é a proteção de grupos vulneráveis. Políticas de moderação devem considerar efeitos desproporcionais sobre mulheres, crianças, adolescentes, minorias raciais, povos indígenas, pessoas LGBTQIA+, pessoas com deficiência, jornalistas e defensores de direitos humanos. Isso exige canais acessíveis de denúncia, respostas céleres a ataques coordenados, treinamento contextual de moderadores, consulta a especialistas e cuidado com vieses algorítmicos. A igualdade material demanda reconhecer que a mesma regra pode produzir impactos distintos conforme o grupo atingido.

O sétimo parâmetro é a governança de dados e algoritmos. Sistemas de recomendação, impulsionamento e ranqueamento influenciam a circulação de conteúdos tanto quanto regras de remoção. Assim, a moderação não pode ser separada da arquitetura algorítmica. Plataformas devem avaliar se seus sistemas amplificam conteúdos violentos, discriminatórios, fraudulentos ou manipulativos. A LGPD oferece princípios úteis para essa tarefa, especialmente necessidade, transparência, prevenção, segurança e responsabilização (BRASIL, 2018). A governança constitucional exige observar não apenas o conteúdo isolado, mas o modo como o sistema o distribui.

O oitavo parâmetro é a auditabilidade independente. A sociedade não pode depender exclusivamente da narrativa das próprias plataformas sobre seus riscos e resultados. Pesquisadores qualificados, autoridades competentes e organismos independentes devem ter condições de avaliar políticas, dados agregados, funcionamento de sistemas e impactos sociais, com preservação de segredos comerciais legítimos, segurança e proteção de dados pessoais. Auditoria não é exposição irrestrita do código ou de informações sensíveis, mas mecanismo institucional de confiança e controle.

O nono parâmetro é a preservação contra censura estatal e captura política. A regulação de plataformas não pode autorizar órgãos públicos a ordenar remoções genéricas, controlar narrativas ou punir críticas sob justificativas vagas. Medidas estatais devem observar legalidade, finalidade legítima, necessidade, proporcionalidade, motivação e controle judicial. A proteção contra desinformação e discurso de ódio não pode ser convertida em instrumento de blindagem de governos, autoridades ou grupos majoritários. Esse cuidado é essencial para que a regulação fortaleça, e não enfraqueça, o espaço democrático.

O décimo parâmetro é a participação multissetorial. A elaboração e a revisão de políticas de moderação devem envolver sociedade civil, academia, imprensa, comunidades técnicas, representantes de grupos vulneráveis, autoridades independentes e usuários. A participação amplia legitimidade, melhora diagnóstico de riscos e reduz a chance de soluções tecnocráticas desconectadas da realidade social. A governança digital deve ser aberta a perspectivas diversas, inclusive porque os impactos das plataformas variam conforme território, idioma, condição econômica, raça, gênero, idade e contexto político.

O décimo primeiro parâmetro é a reparação. Usuários afetados por remoções indevidas, suspensões arbitrárias, exposição abusiva ou omissão diante de violações graves devem dispor de caminhos efetivos de reclamação e reparação. A reparação pode envolver restauração de conteúdo, reativação de conta, retificação, explicação, preservação de provas, indenização ou encaminhamento a autoridades competentes. O direito à reparação reforça a ideia de que moderação não é favor empresarial, mas prática que pode gerar consequências jurídicas.

O décimo segundo parâmetro é a educação midiática e informacional. Nem toda resposta a riscos digitais deve ser repressiva. Usuários precisam compreender como funcionam recomendações, publicidade, denúncias, fraudes, manipulações e práticas de segurança. A UNESCO destaca a importância da alfabetização midiática e informacional para reduzir desigualdades de participação e fortalecer autonomia dos usuários (UNESCO, 2023). A governança constitucional deve combinar deveres de plataformas, regulação pública e capacitação social, evitando que todo problema seja tratado apenas por remoção de conteúdo.

Esses parâmetros demonstram que a moderação de conteúdo deve ser pensada como ecossistema. Remoção pontual é apenas um instrumento. A proteção de direitos depende de desenho de produto, política de dados, regras de publicidade, sistemas de recomendação, canais de denúncia, revisão humana, transparência, auditoria, educação e controle jurisdicional. A governança de plataformas digitais exige articulação entre direito constitucional, regulação econômica, proteção de dados, direitos humanos e responsabilidade civil.

A aplicação desses parâmetros permite rejeitar soluções simplistas. Dizer que plataformas devem remover todo conteúdo denunciado transfere poder excessivo a denunciante e incentiva censura privada. Dizer que plataformas só podem agir após decisão judicial ignora a velocidade e a escala dos danos digitais. Dizer que plataformas devem decidir livremente segundo seus termos de uso desconsidera a dimensão pública de sua atuação. A resposta constitucional está em deveres procedimentais, proporcionalidade, risco, transparência e controle democrático.

7. CONSIDERAÇÕES FINAIS

A moderação de conteúdo por plataformas digitais é uma das manifestações mais relevantes do poder privado na esfera pública contemporânea. Plataformas organizam a circulação de informações, estruturam interações sociais, amplificam ou reduzem visibilidade de discursos e aplicam sanções que afetam diretamente a participação dos usuários. Por isso, suas decisões não podem ser tratadas como simples atos internos de administração contratual. Quando atingem liberdade de expressão, dignidade, igualdade, privacidade, proteção de crianças e adolescentes e democracia, tornam-se matéria constitucional.

O artigo partiu do problema de saber quais parâmetros constitucionais devem orientar a moderação de conteúdo no Brasil, de modo a compatibilizar liberdade de expressão, proteção contra danos, governança tecnológica e direitos fundamentais. A hipótese foi confirmada. A moderação de conteúdo somente será constitucionalmente legítima quando estruturada por governança transparente, proporcional, responsiva, auditável e comprometida com a proteção simultânea da liberdade de expressão e dos demais direitos fundamentais. A ausência de regulação e a intervenção estatal excessiva são insuficientes e perigosas.

Verificou-se que plataformas digitais atuam como infraestruturas do discurso público e exercem formas de governança privada. Elas não apenas hospedam conteúdos, mas definem regras, organizam visibilidade, aplicam sanções e influenciam a formação da opinião pública. A moderação, portanto, é inevitável, mas precisa ser submetida a critérios de legitimidade. A questão não é se haverá moderação, mas como ela será realizada, com que grau de transparência, quais garantias serão oferecidas aos usuários e quais controles existirão sobre decisões que afetam direitos.

A análise dos direitos fundamentais demonstrou que a liberdade de expressão deve permanecer no centro do modelo, mas não pode ser absolutizada a ponto de impedir respostas a violações graves. Dignidade, igualdade, privacidade, proteção de dados, proteção integral de crianças e adolescentes e integridade democrática também integram o quadro constitucional. A

moderação adequada exige ponderação contextual, proporcionalidade e procedimentos capazes de evitar tanto remoções arbitrárias quanto omissões negligentes.

No plano normativo, o Brasil dispõe de fundamentos relevantes na Constituição, no Marco Civil da Internet, na Lei Geral de Proteção de Dados, na legislação de proteção de crianças e adolescentes em ambientes digitais e na jurisprudência constitucional sobre responsabilidade de plataformas. Contudo, ainda é necessário consolidar parâmetros mais claros para deveres de transparência, avaliação de riscos, mecanismos de recurso, auditoria independente e proteção de grupos vulneráveis. A experiência europeia do Digital Services Act e as Diretrizes da UNESCO oferecem referências úteis, desde que adaptadas à realidade constitucional brasileira.

Como resposta ao problema proposto, sustentou-se que a governança constitucional da moderação de conteúdo deve observar: centralidade da liberdade de expressão e da proteção contra danos; transparência substantiva; devido processo procedimental; proporcionalidade das medidas; responsabilidade baseada em risco e capacidade; proteção de grupos vulneráveis; governança de dados e algoritmos; auditabilidade independente; salvaguardas contra censura estatal; participação multissetorial; reparação; e educação midiática e informacional. Esses parâmetros não eliminam conflitos, mas oferecem critérios para enfrentá-los de modo democrático.

Conclui-se que a regulação das plataformas digitais não deve buscar controle total do discurso, nem aceitar a completa privatização das decisões sobre a esfera pública. O desafio constitucional brasileiro é construir uma governança que limite o poder das plataformas, impeça abusos estatais, proteja vítimas de danos digitais e preserve a abertura do debate público. A moderação de conteúdo, quando orientada por direitos fundamentais, pode deixar de ser prática opaca de gestão empresarial e tornar-se instrumento de proteção democrática da comunicação digital.

REFERÊNCIAS

BALKIN, Jack M. Free speech in the algorithmic society: big data, private governance, and new school speech regulation. *UC Davis Law Review*, Davis, v. 51, n. 3, p. 1149-1210, 2018. Disponível em: <https://lawreview.law.ucdavis.edu/archives/51/3/free-speech-algorithmic-society-big-data-private-governance-and-new-school-speech>. Acesso em: 27 abr. 2026.

BRASIL. Constituição da República Federativa do Brasil de 1988. Brasília, DF: Presidência da República, 1988. Disponível em:

https://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm. Acesso em: 27 abr. 2026.

BRASIL. Lei nº 12.965, de 23 de abril de 2014. Estabelece princípios, garantias, direitos e deveres para o uso da internet no Brasil. Brasília, DF: Presidência da República, 2014. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2014/lei/112965.htm. Acesso em: 27 abr. 2026.

BRASIL. Lei nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais. Brasília, DF: Presidência da República, 2018. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/113709compilado.htm. Acesso em: 27 abr. 2026.

BRASIL. Lei nº 15.211, de 17 de setembro de 2025. Dispõe sobre a proteção de crianças e adolescentes em ambientes digitais. Brasília, DF: Presidência da República, 2025. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2023-2026/2025/lei/L15211.htm. Acesso em: 27 abr. 2026.

EUROPEAN UNION. Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act). Official Journal of the European Union, Luxembourg, 27 Oct. 2022. Disponível em: <https://eur-lex.europa.eu/legal-content/EN-PT/TXT/?uri=CELEX:32022R2065>. Acesso em: 27 abr. 2026.

GILLESPIE, Tarleton. Custodians of the Internet: platforms, content moderation, and the hidden decisions that shape social media. New Haven: Yale University Press, 2018.

KLONICK, Kate. The new governors: the people, rules, and processes governing online speech. Harvard Law Review, Cambridge, v. 131, n. 6, p. 1598-1670, 2018. Disponível em: <https://harvardlawreview.org/print/vol-131/the-new-governors-the-people-rules-and-processes-governing-online-speech/>. Acesso em: 27 abr. 2026.

ROBERTS, Sarah T. Behind the screen: content moderation in the shadows of social media. New Haven: Yale University Press, 2019.

SARLET, Ingo Wolfgang. Dignidade da pessoa humana e direitos fundamentais na Constituição Federal de 1988. 10. ed. Porto Alegre: Livraria do Advogado, 2015.

SUPREMO TRIBUNAL FEDERAL. STF define parâmetros para responsabilização de plataformas por conteúdos de terceiros. Brasília, DF: STF, 26 jun. 2025. Disponível em: <https://noticias.stf.jus.br/postsnovicias/stf-define-parametros-para-responsabilizacao-de-plataformas-por-conteudos-de-terceiros/>. Acesso em: 27 abr. 2026.

UNITED NATIONS. Guiding Principles on Business and Human Rights: implementing the United Nations “Protect, Respect and Remedy” Framework. New York; Geneva: United Nations, 2011. Disponível em: https://www.ohchr.org/documents/publications/guidingprinciplesbusinesshr_en.pdf. Acesso em: 27 abr. 2026.

UNESCO. Diretrizes para a governança das plataformas digitais: salvaguardar a liberdade de expressão e o acesso à informação com uma abordagem multissetorial. Paris: UNESCO, 2023. Disponível em: <https://unesdoc.unesco.org/ark:/48223/pf0000387560>. Acesso em: 27 abr. 2026.