




**LIBERDADE DE EXPRESSÃO E DISCURSO DE ÓDIO NAS REDES SOCIAIS
SOB A PERSPECTIVA CONSTITUCIONAL**

**FREEDOM OF EXPRESSION AND HATE SPEECH ON SOCIAL MEDIA FROM A
CONSTITUTIONAL PERSPECTIVE**

**LIBERTAD DE EXPRESIÓN Y DISCURSO DE ODIO EN LAS REDES SOCIALES
DESDE UNA PERSPECTIVA CONSTITUCIONAL**

 <https://doi.org/10.56238/levv17n59-010>

Data de submissão: 07/03/2026

Data de publicação: 07/04/2026

Jaqueline Rodrigues
Graduanda em Direito

RESUMO

Este artigo analisa o conflito normativo entre liberdade de expressão e discurso de ódio nas redes sociais sob a perspectiva da Constituição Federal de 1988, investigando os limites constitucionais para a manifestação do pensamento no ambiente digital. Diante da ausência de tipificação específica de condutas odiosas no ordenamento jurídico brasileiro, o estudo examina como o Estado, o Poder Judiciário e as plataformas digitais equilibram direitos fundamentais aparentemente antagônicos, a livre manifestação do pensamento e a dignidade da pessoa humana, mediante a teoria da ponderação de Robert Alexy. A metodologia adotada é qualitativa e descritiva, com revisão bibliográfica crítica de doutrina, jurisprudência do Supremo Tribunal Federal e análise das políticas de moderação das principais plataformas digitais. Constatou-se que, embora a liberdade de expressão seja pilar do Estado Democrático de Direito, jamais foi concebida como direito absoluto, operando dentro de uma lógica de concordância prática com outros bens jurídicos protegidos. A governança privada da expressão pelas redes sociais revela-se ambígua, com critérios opacos de moderação algorítmica que frequentemente removem conteúdos lícitos enquanto mantêm discursos discriminatórios. Conclui-se que a superação desse impasse exige modelo híbrido que combine autorregulação informada das plataformas, supervisão estatal proporcional e educação midiática, sempre com respeito aos cânones do devido processo legal e da proporcionalidade, evitando o falso dilema entre liberdade absoluta e censura estatal.

Palavras-chave: Liberdade de Expressão. Discurso de Ódio. Direitos Fundamentais. Redes Sociais.

ABSTRACT

This article analyzes the normative conflict between freedom of expression and hate speech on social media from the perspective of the 1988 Brazilian Federal Constitution, investigating the constitutional limits for freedom of thought in the digital environment. Given the absence of specific typification of hateful conduct in the Brazilian legal system, the study examines how the State, the Judiciary, and digital platforms balance apparently antagonistic fundamental rights—the free manifestation of thought and human dignity—through Robert Alexy's theory of balancing. The methodology adopted is qualitative and descriptive, employing a critical bibliographic review of legal doctrine, jurisprudence from the Supreme Federal Court (STF), and analysis of the moderation policies of major digital platforms. It was found that, although freedom of expression constitutes a pillar of the Democratic



State of Law, it was never conceived as an absolute right, operating instead within a logic of practical concordance with other constitutionally protected legal goods. The private governance of expression by social networks reveals itself as ambiguous, with opaque criteria for algorithmic moderation that frequently remove lawful content while maintaining discriminatory speech. It is concluded that overcoming this impasse requires a hybrid model combining informed self-regulation by platforms, proportional state supervision, and media literacy, always respecting the canons of due process of law and proportionality, thereby avoiding the false dilemma between absolute freedom and state censorship.

Keywords: Freedom of Expression. Hate Speech. Fundamental Rights. Social Networks.

RESUMEN

Este artículo analiza el conflicto normativo entre la libertad de expresión y el discurso de odio en las redes sociales desde la perspectiva de la Constitución Federal de 1988, investigando los límites constitucionales para la expresión de pensamiento en el entorno digital. Ante la ausencia de clasificaciones específicas de conducta de odio en el ordenamiento jurídico brasileño, el estudio examina cómo el Estado, el Poder Judicial y las plataformas digitales equilibran derechos fundamentales aparentemente antagónicos —la libertad de expresión y la dignidad de la persona humana— utilizando la teoría del equilibrio de Robert Alexy. La metodología adoptada es cualitativa y descriptiva, con una revisión bibliográfica crítica de la doctrina, la jurisprudencia de la Suprema Corte Federal y un análisis de las políticas de moderación de las principales plataformas digitales. Se constató que, si bien la libertad de expresión es un pilar del Estado de Derecho democrático, nunca se ha concebido como un derecho absoluto, operando dentro de una lógica de concordancia práctica con otros bienes jurídicos protegidos. La regulación privada de la expresión por parte de las redes sociales se revela ambigua, con criterios de moderación algorítmica opacos que frecuentemente eliminan contenido legítimo mientras mantienen el discurso discriminatorio. Se concluye que superar este estancamiento requiere un modelo híbrido que combine la autorregulación informada de las plataformas, la supervisión estatal proporcional y la alfabetización mediática, respetando siempre los principios del debido proceso y la proporcionalidad, evitando el falso dilema entre la libertad absoluta y la censura estatal.

Palabras clave: Libertad de Expresión. Discurso de Odio. Derechos Fundamentales. Redes Sociales.



1 INTRODUÇÃO

A liberdade de expressão constitui um dos pilares fundamentais do Estado Democrático de Direito, assegurando ao indivíduo o direito de manifestar livremente suas opiniões, crenças e ideias. Com o avanço tecnológico e o crescimento exponencial das redes sociais, esse direito passou a ser exercido em novas dimensões, possibilitando que milhões de pessoas compartilhem suas visões de mundo de forma instantânea e global (De Freitas, 2025).

Contudo, ainda de acordo com a publicação de De Freitas (2025), essa ampliação do espaço comunicativo também trouxe desafios significativos, sobretudo quanto aos limites éticos e jurídicos da manifestação do pensamento.

A facilidade de acesso à informação e a ausência de filtros imediatos para o conteúdo publicado tornaram-se terreno fértil para a propagação de discursos discriminatórios e ofensivos, suscitando o debate sobre o que configura o legítimo exercício da liberdade de expressão e o que ultrapassa os limites constitucionais, caracterizando o discurso de ódio.

Nesse contexto, o ordenamento jurídico brasileiro enfrenta o desafio de equilibrar a proteção à livre manifestação com a necessidade de salvaguardar outros direitos fundamentais, como a dignidade da pessoa humana, a igualdade e o respeito às minorias. A teoria dos direitos fundamentais de Robert Alexy, ao propor a ponderação entre princípios constitucionais em conflito, oferece uma importante base teórica para compreender e solucionar esses impasses (Geremberg, 2022).

A relevância do tema se intensifica diante das discussões contemporâneas sobre a regulação das plataformas digitais e o combate à desinformação, especialmente após a proposição do Projeto de Lei nº 2.630/2020, conhecido como “Lei Brasileira de Liberdade, Responsabilidade e Transparência na Internet”.

A delimitação do tema centra-se na análise do conflito entre liberdade de expressão e discurso de ódio nas redes sociais digitais, especialmente no contexto brasileiro, sob a ótica da Constituição Federal de 1988 e das decisões do Supremo Tribunal Federal (STF). Assim, o estudo se restringe ao campo constitucional e dos direitos fundamentais, sem adentrar na seara penal, considerando ainda o papel das plataformas digitais na moderação e responsabilização de conteúdos.

Diante desse cenário, surge o seguinte problema de pesquisa: como o ordenamento jurídico brasileiro trata o conflito entre a liberdade de expressão e o discurso de ódio nas redes sociais, e quais são os limites constitucionais para essa manifestação no ambiente digital?

Para dar uma resposta satisfatória a esse problema, o objetivo geral deste trabalho é analisar, sob a perspectiva da Constituição de 1988, os limites da liberdade de expressão nas redes sociais quando colide com o discurso de ódio, buscando compreender o papel do Estado, do Poder Judiciário e das plataformas digitais na proteção dos direitos fundamentais.

Como objetivos específicos, pretende-se: identificar os fundamentos constitucionais da liberdade de expressão e seus limites jurídicos; examinar decisões recentes do STF sobre o tema; e investigar a atuação das plataformas digitais na moderação de conteúdos ofensivos e discriminatórios.

A metodologia adotada é de natureza qualitativa e descritiva, utilizando-se o método de revisão bibliográfica. Onde são analisados livros, artigos científicos, dissertações e decisões judiciais publicadas nos últimos dez anos. A análise baseia-se na interpretação crítica da doutrina e jurisprudência, buscando compreender as principais abordagens teóricas e práticas sobre o tema.

A escolha do tema justifica-se por sua relevância científica, ao contribuir para o avanço dos estudos sobre direitos fundamentais, especialmente no campo da liberdade de expressão e suas limitações constitucionais diante das novas tecnologias de comunicação. Apresenta também relevância social, pois a discussão envolve diretamente a convivência democrática, a proteção das minorias e o combate a discursos de ódio e intolerância que ameaçam os valores constitucionais. Por fim, o tema é viável, considerando a ampla disponibilidade de materiais teóricos e jurisprudenciais atualizados, o que possibilita uma análise consistente e fundamentada sobre o equilíbrio entre a liberdade de expressão e a proteção contra o discurso de ódio no ambiente digital.

2 LIBERDADE DE EXPRESSÃO NO ESTADO DEMOCRÁTICO DE DIREITO

A liberdade de expressão consagrada pela Constituição Federal de 1988 representa um dos pilares do Estado democrático de direito brasileiro, funcionando simultaneamente como instrumento de controle social e condição para a formação da opinião pública.

Seus fundamentos encontram assento nos artigos 5º, incisos IV e IX, e 220 da Carta Magna, que garantem a livre manifestação do pensamento e da atividade intelectual, artística, científica e de comunicação, vedando expressamente qualquer forma de censura prévia (Brasil, 1988).

Contudo, como bem observa Teles (2023), tal liberdade jamais foi concebida como absoluta, mas sim como direito relativo que deve coexistir harmonicamente com outros bens jurídicos igualmente protegidos pela ordem constitucional.

A doutrina contemporânea reconhece que os direitos fundamentais operam dentro de uma lógica de concordância prática, na qual nenhum princípio goza de primazia absoluta frente aos demais. Nesse sentido, a liberdade de expressão encontra seus limites naturais quando colide com a dignidade da pessoa humana, a honra, a imagem e a privacidade alheias, valores estes também erigidos à categoria de direitos fundamentais.

Como destaca De Andrade e Andrade (2022), a vedação ao anonimato prevista no artigo 5º, IV, da Constituição não representa mero obstáculo formal, mas mecanismo essencial para viabilizar a responsabilização do manifestante, transformando a liberdade em exercício consciente e não em licença para abusos.

O advento da internet e das redes sociais potencializou exponencialmente o alcance da expressão individual, rompendo barreiras geográficas e democratizando o acesso aos meios de comunicação. Contudo, essa mesma revolução tecnológica trouxe desafios normativos sem precedentes, evidenciando as limitações do modelo tradicional de regulação estatal.

O Marco Civil da Internet (Lei nº 12.965/2014) surgiu como resposta legislativa a essa nova realidade, estabelecendo princípios fundamentais como a neutralidade de rede e a responsabilidade subsidiária dos provedores de acesso à Internet (Brasil, 2014).

Entretanto, como apontam Almeida et al. (2022), a lei revelou-se insuficiente para lidar com a velocidade e complexidade dos abusos digitais, especialmente diante da dificuldade em identificar autores e da morosidade dos mecanismos judiciais de remoção de conteúdos ilícitos.

A privatização da moderação de conteúdo pelas grandes plataformas digitais criou um paradoxo contemporâneo: enquanto o Estado se vê limitado por garantias constitucionais, empresas privadas assumem funções quase estatais na definição do que é ou não permitido no espaço digital.

Essa governança privada da expressão, como analisa Teles (2023), opera com critérios frequentemente opacos e inconsistentes, removendo conteúdos lícitos enquanto mantém discursos de ódio e desinformação, evidenciando a necessidade de maior transparência e prestação de contas das redes sociais. Diante desse cenário, impõe-se repensar os mecanismos de responsabilização na era digital sem cair no falso dilema entre liberdade absoluta e censura estatal.

A jurisprudência brasileira tem avançado no sentido de aplicar a teoria da ponderação aos casos concretos, reconhecendo que a liberdade de expressão não pode servir de escudo para a prática de crimes contra a honra ou para a disseminação deliberada de desinformação com potencial lesivo à ordem democrática. Como bem sintetiza Coelho (2025, 28), “o equilíbrio exige não apenas aperfeiçoamento legislativo, mas também educação midiática e maior responsabilização das plataformas no combate à desinformação”.

Portanto, os fundamentos constitucionais da liberdade de expressão permanecem válidos na era digital, mas exigem interpretação dinâmica capaz de harmonizar o direito à manifestação com a proteção de outros bens jurídicos essenciais.

A liberdade não pode significar impunidade, assim como a responsabilização não pode configurar censura. O desafio contemporâneo reside precisamente em traçar essa fronteira tênue com critérios racionais, transparentes e compatíveis com os valores democráticos que a própria Constituição de 1988 consagrou como norte do nosso ordenamento jurídico.



3 DISCURSO DE ÓDIO E SUAS IMPLICAÇÕES JURÍDICAS NAS REDES SOCIAIS

A ascensão das redes sociais transformou radicalmente os contornos da liberdade de expressão, convertendo-a em um direito exercido com inédita amplitude, mas também com responsabilidades ainda mal delineadas pelo ordenamento jurídico.

Enquanto a Constituição Federal de 1988 consagra nos artigos 5º, incisos IV e IX, e 220 a livre manifestação do pensamento como pilar do Estado democrático, essa garantia jamais foi concebida como absoluta, devendo coexistir harmonicamente com outros direitos fundamentais igualmente protegidos, notadamente a dignidade da pessoa humana e a igualdade (Brasil, 1988).

O discurso de ódio, entendido como manifestação que incita discriminação, hostilidade ou violência contra grupos vulneráveis com base em características identitárias, situa-se justamente nesse ponto de tensão normativa.

Conforme destacam Almeida e Santos (2021), embora a Convenção Americana sobre Direitos Humanos determine no artigo 13.5 a proibição de apologia ao ódio nacional, racial ou religioso, o ordenamento brasileiro ainda apresenta lacunas significativas na tipificação específica dessas condutas no ambiente digital, gerando insegurança jurídica tanto para vítimas quanto para operadores do direito.

A jurisprudência do Supremo Tribunal Federal tem buscado equilibrar esses valores em conflito, reconhecendo que a liberdade de expressão não abrange manifestações que objetivem a desumanização de grupos historicamente marginalizados. Nesse sentido, o Inquérito 4.781/DF, instaurado em 2019 sob relatoria do Ministro Alexandre de Moraes, revela-se marco paradigmático na compreensão jurisprudencial contemporânea. O procedimento investiga "atos antidemocráticos que atentam contra a honra e a segurança de autoridades dos Três Poderes, com a finalidade de subverter a ordem constitucional e o Estado Democrático de Direito, mediante a utilização das redes sociais" (Brasil, STF, Inquérito 4.781/DF).

Laudos periciais apontaram a existência de um grupo organizado que utilizava robôs e estratégias de "efeito manada" para amplificar conteúdos de ódio, chegando a impulsionar simultaneamente dez perfis para difundir a *hashtag* #ImpeachmentGilmarMendes aos *trending topics* de forma artificial (Coelho, 2025, p. 21).

Contudo, como observa Pereira Filho (2022), o STF ainda não sedimentou parâmetros objetivos e uniformes para distinguir criticamente o discurso legítimo da manifestação odiosa, gerando discrepâncias na aplicação jurisprudencial.

A ementa do Agravo Regimental na Petição 10.391/DF sintetiza o entendimento consolidado:

[...] a liberdade de expressão é consagrada constitucionalmente e balizada pelo binômio LIBERDADE E RESPONSABILIDADE, ou seja, o exercício desse direito não pode ser utilizado como verdadeiro escudo protetivo para a prática de atividades ilícitas. Não se confunde liberdade de expressão com impunidade para agressão [...] (Brasil, STF, Petição 10.391/DF, 2022).

As plataformas digitais, por sua vez, assumiram papel ambíguo nesse cenário, funcionando simultaneamente como espaços privilegiados de exercício da cidadania, onde vozes marginalizadas encontram visibilidade e mobilização coletiva, e como agentes não estatais de moderação privada de conteúdo, cujas decisões algorítmicas e políticas de uso definem, muitas vezes de forma opaca, os limites do discurso público aceitável nas esferas digitais contemporâneas.

Segundo Nitrini (2021), essa governança algorítmica opera com critérios frequentemente opacos, removendo conteúdos lícitos enquanto mantém discursos discriminatórios, evidenciando a necessidade de maior transparência e accountability das redes sociais. A moderação automatizada, embora necessária diante do volume exponencial de postagens, carece de sensibilidade contextual para distinguir ironia legítima de incitação ao ódio.

A problemática se agrava quando se considera que o anonimato relativo proporcionado pelo ambiente digital encoraja manifestações preconceituosas, enquanto a fragmentação da esfera pública em "bolhas informacionais", como bem analisa Serafim (2025), potencializa a radicalização de discursos extremistas.

Nesse contexto, a vedação constitucional ao anonimato revela-se insuficiente sem mecanismos eficazes de identificação de autores e responsabilização proporcional, especialmente diante da resistência inicial das plataformas quanto à jurisdição brasileira, como ocorreu quando o *Facebook* argumentou que a determinação de bloqueio global "extrapolava a jurisdição nacional" (Serafim, 2025, 30).

Diante desse quadro, impõe-se repensar os mecanismos de responsabilização na era digital sem cair no falso dilema entre liberdade absoluta e censura estatal. A jurisprudência brasileira tem avançado no sentido de aplicar a teoria da ponderação aos casos concretos, reconhecendo que a liberdade de expressão não pode servir de escudo para a prática de crimes contra a honra ou para a disseminação deliberada de desinformação com potencial lesivo à ordem democrática.

Como bem sintetizam Almeida e Santos (2021), a edificação social mediada pelas redes exige compreensão dos impactos das interações digitais, demandando regulamentação que equilibre inovação tecnológica e proteção dos direitos humanos.

Assim, os fundamentos constitucionais da liberdade de expressão permanecem válidos na era digital, mas exigem interpretação dinâmica capaz de harmonizar o direito à manifestação com a proteção de outros bens jurídicos essenciais.

O discurso de ódio, ao violar princípios constitucionais da igualdade e dignidade, configura uma ameaça concreta à coesão social e ao Estado Democrático de Direito, devendo ser enfrentado por meio de ações articuladas entre Estado, sociedade e instituições jurídicas, sempre com respeito aos cânones do devido processo legal e da proporcionalidade.



3.1 JURISPRUDÊNCIA E ATUAÇÃO DO SUPREMO TRIBUNAL FEDERAL

A atuação do Supremo Tribunal Federal na delimitação dos contornos da liberdade de expressão na era digital revela um Judiciário em constante tensão entre sua função de guardião da Constituição e os riscos inerentes à judicialização excessiva de debates políticos e sociais.

Desde a emblemática ADPF 130/DF, que declarou a incompatibilidade da Lei de Imprensa com a ordem constitucional de 1988, a Corte consolidou entendimento no sentido de que eventuais abusos na manifestação do pensamento devem ser enfrentados por meios repressivos a posteriori, jamais mediante censura prévia, reafirmando assim o caráter inegociável desse princípio democrático (Brasil, STF, ADPF 130/DF, 2009).

Contudo, essa mesma jurisprudência revela nuances significativas quando confrontada com discursos que atentam contra a dignidade humana coletiva. No histórico Habeas Corpus 82.424/RS, conhecido como caso Ellwanger, o STF reconheceu que manifestações negacionistas do Holocausto extrapolam os limites constitucionais da liberdade de expressão, configurando apologia ao racismo e, portanto, conduta passível de punição penal.

A decisão assentou que o direito fundamental à livre manifestação não protege discursos que objetivem a desumanização de grupos étnicos, estabelecendo parâmetro crucial para a interpretação do artigo 5º, inciso XLII, da Constituição (Brasil, STF, HC 82.424/RS, 2003).

A Corte também enfrentou dilemas complexos no equilíbrio entre direitos em conflito, como evidenciado na ADI 4.451/DF, que tratou das biografias não autorizadas. Ao julgar inconstitucional a exigência de autorização prévia para publicação de obras biográficas, o STF priorizou o direito à informação e à memória histórica sobre eventuais restrições ao direito à imagem, argumentando que a censura prévia representaria ameaça inaceitável ao pluralismo informativo essencial à democracia (Brasil, STF, ADI 4.451/DF, 2018).

Essa decisão ilustra a aplicação pragmática da teoria da ponderação, técnica hermenêutica que busca harmonizar direitos fundamentais aparentemente antagônicos. Mais recentemente, a atuação do STF no Inquérito 4.781/DF e na Petição 12.404/DF evidenciou uma postura judicial mais intervencionista diante dos desafios da desinformação coordenada. A determinação de bloqueio de perfis e, em 2024, a suspensão temporária da plataforma X (antigo Twitter) no território nacional geraram intenso debate doutrinário.

Enquanto uns veem nessas medidas resposta legítima à ameaça à ordem democrática, outros, como Pereira Filho (2022), alertam para os riscos de discricionariedade excessiva, destacando que "é conferido ao poder judiciário um elevado grau de discricionariedade para definir tais limites", o que pode comprometer a previsibilidade jurídica e a segurança das manifestações legítimas.

Essa ambiguidade reflete-se na escassa jurisprudência direta sobre discursos de ódio nas redes sociais. Conforme apontam Santos e Medina (2023), pesquisa realizada entre 2017 e 2019 revelou que

o STF permaneceu praticamente silente sobre manifestações discriminatórias publicadas em plataformas digitais, enquanto os Tribunais Regionais Federais julgaram apenas 17 processos com essa temática.

Tal lacuna jurisprudencial contrasta com a intensidade do fenômeno social, evidenciando a dificuldade do Judiciário em acompanhar a velocidade das transformações tecnológicas. A moderação privada de conteúdo pelas plataformas acentua essa complexidade.

Como bem observa Nitrini (2021), as redes sociais assumem funções quase estatais na definição do que é ou não permitido no espaço digital, operando com critérios frequentemente opacos. Diante disso, o STF oscila entre exigir das plataformas maior responsabilização, como na Petição 10.391/DF, que determinou o bloqueio de perfis do PCO por disseminação de discursos de ódio, e reconhecer os limites da jurisdição nacional frente à arquitetura global da internet.

Deste modo, a jurisprudência do Supremo Tribunal Federal sobre liberdade de expressão digital revela um Judiciário consciente de seu papel constitucional, porém ainda em busca de parâmetros estáveis para enfrentar os desafios contemporâneos.

A necessidade de proteger a democracia contra ataques coordenados não pode justificar a erosão dos próprios pilares democráticos, notadamente a liberdade de manifestação. O equilíbrio exige não apenas decisões pontuais, mas a construção de uma doutrina jurisprudencial sólida, transparente e capaz de resistir às pressões políticas do momento, preservando assim a credibilidade institucional da Corte como última guardiã da Constituição.

4 AS PLATAFORMAS DIGITAIS E A MODERAÇÃO DE CONTEÚDO

As plataformas digitais assumiram papel paradoxal na arquitetura contemporânea da comunicação, funcionam simultaneamente como praças públicas digitais vibrantes e como espaços privados regidos por lógicas comerciais.

Conforme observa Klonick, citada por Barroso (2022), essas empresas tornaram-se os "novos governantes" do discurso online, posicionando-se entre o Estado e os usuários como árbitros privados do que pode ou não circular no ambiente digital. Essa governança algorítmica, contudo, opera com critérios frequentemente opacos e desprovidos das garantias processuais que balizam a atuação estatal.

A moderação de conteúdo em escala massiva revela-se desafio técnico e ético sem precedentes. Diante do bilhão de publicações diárias, as plataformas recorrem a sistemas automatizados que classificam, filtram e removem conteúdos com base em termos de uso elaborados unilateralmente.

Como bem analisam Santos et al. (2023), as seis principais plataformas estudadas, Facebook, Instagram, Twitter/X, YouTube, TikTok e Twitch, apresentam níveis distintos de detalhamento em suas diretrizes sobre discurso de ódio, mas compartilham problemas estruturais: regras vagas que não

definem claramente o que constitui violação e aplicação inconsistente que impacta desproporcionalmente minorias vulneráveis.

Nesse cenário, emerge a tensão entre eficiência operacional e respeito aos direitos fundamentais. A moderação automatizada, embora necessária para lidar com o volume exponencial de postagens, carece de sensibilidade contextual para distinguir ironia legítima de incitação ao ódio ou crítica política de discurso de hostilidade.

Poletto e Moraes (2022) destacam que a anotação, ocultação e remoção de conteúdos ocorrem muitas vezes sem análise humana prévia, gerando falsos positivos que silenciam manifestações legítimas sob o manto da segurança da comunidade.

A ausência de garantias processuais mínimas agrava esse quadro. Silvestre (2023) argumenta que, mesmo em esfera privada, a moderação de conteúdo que restringe direitos fundamentais deveria observar princípios basilares do devido processo legal, notificação prévia fundamentada, possibilidade de contestação e revisão por agente humano imparcial.

Contudo, a maioria dos atuais mecanismos de recurso oferecidos pelas plataformas frequentemente se revelam ineficazes, com respostas padronizadas e tempos de análise lentos que inviabilizam a reparação tempestiva de equívocos.

Valle (2025) propõe superar a visão reducionista que analisa apenas o texto da comunicação, defendendo a consideração do ambiente comunicacional como um todo, incluindo os recursos funcionais e as características arquitetônicas das interfaces das plataformas, algoritmos de amplificação e dinâmicas de bolhas informacionais.

Essa perspectiva revela que a moderação não se limita à remoção de postagens, mas envolve decisões sutis sobre visibilidade, alcance e recomendação de conteúdos, muitas vezes mais impactantes que a simples exclusão.

Diante desses desafios, ganham relevância iniciativas como os Santa Clara Principles, que estabelecem padrões mínimos de transparência e accountability, e o PL 2.630/2020 (Lei Brasileira de Liberdade, Responsabilidade e Transparência na Internet), que busca impor às plataformas obrigações de clareza sobre políticas de moderação e mecanismos efetivos de contestação.

Contudo, como alerta Gillespie (2018), a mera transparência de números agregados não resolve o problema estrutural da falta de legitimidade democrática na governança privada do discurso público. Assim, a moderação de conteúdo pelas plataformas digitais representa um dos maiores desafios jurídicos contemporâneos, exigindo equilíbrio delicado entre proteção contra discursos lesivos e preservação do pluralismo democrático.

A resposta não reside na demonização das plataformas nem na sua isenção de responsabilidades, mas na construção de um modelo híbrido que combine autorregulação informada,

supervisão estatal proporcional e participação da sociedade civil, sempre com respeito aos cânones do Estado Democrático de Direito.

5 CONSIDERAÇÕES FINAIS

A análise desenvolvida ao longo deste trabalho revelou a complexidade intrínseca ao conflito entre liberdade de expressão e discurso de ódio no ambiente digital, demonstrando que a superação desse impasse normativo exige interpretação dinâmica da Constituição Federal de 1988, capaz de harmonizar direitos fundamentais aparentemente antagônicos sem sucumbir ao falso dilema entre liberdade absoluta e censura estatal.

Constatou-se que a liberdade de expressão, embora consagrada como pilar do Estado Democrático de Direito nos artigos 5º, incisos IV e IX, e 220 da Carta Magna, jamais foi concebida como direito absoluto, operando necessariamente dentro de uma lógica de concordância prática com outros bens jurídicos igualmente protegidos, notadamente a dignidade da pessoa humana, a igualdade e o respeito às minorias historicamente vulnerabilizadas.

Diante do problema de pesquisa formulado verificou-se que o arcabouço normativo pátrio apresenta lacunas significativas na tipificação específica de condutas odiosas no ambiente digital, gerando insegurança jurídica tanto para vítimas quanto para operadores do direito.

A Convenção Americana sobre Direitos Humanos (art. 13.5) determina a proibição de apologia ao ódio nacional, racial ou religioso, mas sua internalização no ordenamento brasileiro carece de regulamentação mais precisa para o contexto das redes sociais, onde a velocidade de disseminação e o alcance global dos discursos amplificam exponencialmente seus potenciais danos à coesão social.

A jurisprudência do Supremo Tribunal Federal revelou-se ambivalente nesse cenário. Por um lado, decisões paradigmáticas estabeleceram parâmetros importantes ao reconhecer que manifestações que objetivem a desumanização de grupos étnicos extrapolam os limites constitucionais da liberdade de expressão.

Por outro, a escassez de julgamentos diretos sobre discursos de ódio nas plataformas digitais evidencia a dificuldade do Judiciário em acompanhar a velocidade das transformações tecnológicas. Ademais, medidas mais intervencionistas, como as adotadas no Inquérito 4.781/DF e na suspensão temporária da plataforma X em 2024, geraram intenso debate doutrinário acerca dos riscos de discricionariedade excessiva e erosão da previsibilidade jurídica.

Destaca-se ainda o paradoxo contemporâneo da governança privada da expressão: enquanto o Estado se vê limitado por garantias constitucionais como a vedação à censura prévia (ADPF 130/DF), empresas privadas assumem funções quase estatais na definição do que é ou não permitido no espaço digital.



A moderação algorítmica, embora necessária diante do volume exponencial de postagens, opera com critérios frequentemente opacos, removendo conteúdos lícitos enquanto mantém discursos discriminatórios. A ausência de garantias processuais mínimas, notificação prévia fundamentada, possibilidade de contestação efetiva e revisão por agente humano imparcial, agrava esse quadro, configurando verdadeira "justiça privada" desprovida de legitimidade democrática.

À luz da teoria da ponderação de Robert Alexy, utilizada como fundamento teórico deste estudo, conclui-se que o equilíbrio entre liberdade de expressão e proteção contra o discurso de ódio não pode ser alcançado por meio de soluções simplistas ou dicotômicas.

Por fim, este estudo contribui para o avanço dos debates sobre direitos fundamentais no contexto das novas tecnologias, reafirmando que a proteção contra o discurso de ódio não representa ameaça à liberdade de expressão, mas sim condição para sua plena realização em uma sociedade democrática e plural

Afinal, só em um ambiente onde todos possam se manifestar com segurança e dignidade é que a liberdade de expressão cumpre seu verdadeiro papel: ser instrumento de construção coletiva da cidadania e não ferramenta de opressão de grupos vulneráveis. Cabe ao intérprete constitucional, portanto, traçar com rigor hermenêutico e sensibilidade social os limites tênues entre manifestação legítima e discurso odioso, sempre com respeito aos cânones do devido processo legal e da proporcionalidade que informam o Estado Democrático de Direito brasileiro.



REFERÊNCIAS

- ALMEIDA, Fernando Martins Xavier de; SANTOS, Lucas Octávio Noya dos. Discurso de ódio e liberdade de expressão: o conteúdo jurídico da norma proibitiva e suas implicações sobre a ética do discurso. In: ASSOCIAÇÃO BRASILEIRA DE SOCIOLOGIA DO DIREITO. **Anais do XII Congresso Internacional da ABraSD**. Porto Alegre, 2021. p. 79-85.
- BOTELHO, Gilson Carlos Xavier. O equilíbrio entre a liberdade de expressão e os limites constitucionais nas mídias sociais. **Revista Processus Multidisciplinar**, v. 5, n. 9, p. e091116-e091116, 2024.
- BRASIL. **Supremo Tribunal Federal. Agravo Regimental na Petição nº 10.391/DF**. Relator: Ministro Alexandre de Moraes. Julgamento: 14 nov. 2022.
- BRASIL. **Supremo Tribunal Federal. Inquérito nº 4.781/DF**. Relator: Ministro Alexandre de Moraes. Brasília, DF, 2019.
- BRASIL. **Supremo Tribunal Federal. Ação Declaratória de Inconstitucionalidade por Omissão nº 4.451/DF**. Relator: Ministro Alexandre de Moraes. Brasília, DF, 2018.
- BRASIL. **Supremo Tribunal Federal. Habeas Corpus nº 82.424/RS**. Relator: Ministro Moreira Alves. Brasília, DF, 2003.
- BRASIL. Supremo Tribunal Federal. Inquérito nº 4.781/DF. Relator: Ministro Alexandre de Moraes. Brasília, DF, 2019.
- BRASIL. **Constituição da República Federativa do Brasil de 1988**. Promulgada em 5 de outubro de 1988.
- BRASIL. **Lei nº 12.965, de 23 de abril de 2014**. Institui princípios, garantias, direitos e deveres para o uso da Internet no Brasil. Diário Oficial da União, Brasília, DF, 24 abr. 2014. Seção 1, p. 1.
- COELHO, Graco Chaves Xavier. **Limites jurídicos da liberdade de expressão: fake news, crimes de honra e responsabilidade na internet**. 2025. Trabalho de Conclusão de Curso (Bacharelado em Direito) - Faculdade de Ciências Jurídicas e Sociais, Centro Universitário de Brasília, Brasília, 2025.
- DE ALMEIDA, Rosana Santos et al. A liberdade de expressão e seus limites: uma análise crítica do marco civil da internet. **Research, Society and Development**, v. 11, n. 2, p. e39111225445-e39111225445, 2022.
- DE ANDRADE, Raissa Nacer Oliveira; ANDRADE, Diogo de Calasans Melo. A liberdade de expressão na internet: uma análise acerca dos limites impostos pelo ordenamento jurídico brasileiro. **Revista Contemporânea**, v. 2, n. 1, p. 547-568, 2022.
- DE FREITAS, Isa Omena Machado. Liberdade de Expressão Vs. Discurso de Ódio nas Redes Sociais e seus Limites Legais. **Reflexões sobre Direito e Sociedade: Fundamentos e Práticas - Vol. 14** - Capítulo 32. DOI: 10.47573/aya.5379.3.1.32 - 2025.
- DE LUNETTA, Avaetê; GUERRA, Rodrigues. Metodologia da pesquisa científica e acadêmica. **Revista OWL (OWL Journal)-Revista Interdisciplinar de Ensino e Educação**, v. 1, n. 2, p. 149-159, 2023.



GEREMBERG, Alice Leal Wolf. **A teoria compreensiva de Robert Alexy: a Proposta do Trialismo**. Editora Dialética, 2022.

NITRINI, Rodrigo Vidal. **Liberdade de Expressão nas Redes Sociais: o problema jurídico da remoção de conteúdo pelas plataformas**. Editora Dialética, 2021.

PEREIRA FILHO, Rainel Batista. **Redes sociais e limites à liberdade de expressão: novos desafios para a democracia na era da informação**. Dissertação apresentada ao programa de PósGraduação em Direito (PPGD) do Centro de Ciências Sociais Aplicadas (CCSA) da Universidade Federal do Rio Grande do Norte (UFRN) - 2021.

RUZYK, Carlos Eduardo Pianovski. Desafios da liberdade de expressão nas redes sociais e o papel da responsabilidade civil no direito brasileiro frente à tese da posição preferencial. **Revista IBERC**, v. 6, n. 1, p. 1-20, 2023.

SANTOS, Livia Zanholo; MEDINA, Patricia. Discursos de ódio em redes sociais: uma análise da jurisprudência dos tribunais superiores brasileiros. **Ciências Sociais Aplicadas em Revista**, v. 26, n. 46, p. 277-299, 2023.

SERAFIM, Karine Rodrigues. **Liberdade de expressão, discurso de ódio e redes sociais: desafios legais e sociais no combate aos crimes de ódio no Brasil**. Goiânia: PUC Goiás, 2025.

TELES, Lara Rezende. **Limites à liberdade de expressão: uma análise sob a perspectiva das redes sociais**. Pontifícia Universidade Católica (PUC) de Goiás - Repositório Acadêmico da Graduação (RAG). 2023. Disponível em:
<<https://repositorio.pucgoias.edu.br/jspui/handle/123456789/7193>> - Acesso: 28/12/25.

TREVISAN, Leonardo Simchen. Os direitos fundamentais sociais na teoria de Robert Alexy. **Cadernos do Programa de Pós-Graduação em Direito/UFRGS**, v. 10, n. 1, p. 199-244, 2015. DOI: <https://doi.org/10.22456/2317-8558.54583> - Acesso em: 10/10/2025.