

## **O ciberespaço como vetor criminógeno: a influência de comunidades virtuais de ódio na radicalização juvenil e a responsabilidade das plataformas digitais à luz do marco civil da internet (lei 12.965/14).**

Cyberspace as a criminogenic vector: the influence of virtual hate communities on youth radicalization and the responsibility of digital platforms in light of the internet's civil framework (law 12,965/14).

Rodrigo Silva Braga de Salles<sup>1</sup>

Orientadora: Profa. Rosana Reis de Melo Silva<sup>2</sup>

### **RESUMO**

Este estudo investiga o ciberespaço como vetor criminógeno, analisando o impacto de comunidades virtuais de ódio na radicalização de jovens e a suficiência do Marco Civil da Internet (Lei nº 12.965/14). O objetivo geral é avaliar se o regime de responsabilidade civil das plataformas digitais previsto no artigo 19 contribui para a propagação desses discursos devido à sua morosidade. Sob a ótica da Teoria da Associação Diferencial de Edwin Sutherland, associada aos conceitos de câmaras de eco de Cass Sunstein e filtros invisíveis de Eli Pariser, a fundamentação teórica demonstra que os algoritmos potencializam o aprendizado de condutas desviantes e a polarização extrema. A metodologia consistiu em uma pesquisa bibliográfica e documental de natureza qualitativa. Os resultados revelam uma correlação direta entre o ambiente digital desregulado e o aumento de ataques violentos às escolas no Brasil, evidenciando que o modelo de notice-and-takedown<sup>3</sup> judicializado mostra-se obsoleto frente à dinâmica algorítmica. Conclui-se que o combate ao discurso de ódio online possui sólida fundamentação constitucional na dignidade da pessoa humana, conforme a jurisprudência do

<sup>1</sup> Graduando do curso de Direito, no Centro Universitário FAMETRO. Manaus, Amazonas, Brasil. E-mail: rodrigobsalles@hotmail.com. ID ORCID nº 0009-0008-6527-5394.

<sup>2</sup> Prof.<sup>a</sup> Orientadora e Coordenadora do TCC II, no Centro Universitário FAMETRO. Manaus, Amazonas, Brasil. E-mail: rosanareismello@gmail.com.

<sup>3</sup> Processo operado por provedores de hospedagem online em resposta a ordens judiciais ou alegações de que o conteúdo é ilegal, com o conteúdo sendo removido pelo provedor após a notificação.

Caso Ellwanger, justificando a necessidade de reforma legislativa para instituir um dever de cuidado proativo às plataformas, alinhado ao Digital Services Act<sup>4</sup> europeu.

**Palavras-chave:** Criminologia. Crime de Ódio. Redes Sociais. Marco Civil da Internet.

## **ABSTRACT**

This study investigates cyberspace as a criminogenic vector, analyzing the impact of virtual hate communities on youth radicalization and the adequacy of the Brazilian Internet Bill of Rights (Law No. 12,965/14). The general objective is to evaluate whether the civil liability regime for digital platforms provided for in Article 19 contributes to the spread of such speeches due to its procedural delays. Under the perspective of Edwin Sutherland's Differential Association Theory, combined with Cass Sunstein's concepts of echo chambers and Eli Pariser's invisible filters, the theoretical framework demonstrates that algorithms enhance the learning of deviant behaviors and extreme polarization. The methodology consisted of a qualitative bibliographical and documentary research. The results reveal a direct correlation between the unregulated digital environment and the increase in extreme violence attacks on schools in Brazil, showing that the judicialized 'notice and take down' model is obsolete against algorithmic dynamics. It is concluded that combating online hate speech has a solid constitutional basis in human dignity, as established in the Ellwanger Case, justifying the need for legislative reform to impose a proactive duty of care on platforms, aligned with the European Digital Services Act.

**Keywords:** Criminology. Hate Crime. Online Communities. Brazilian Internet Bill of Rights.

## **1. INTRODUÇÃO**

Este trabalho investiga o papel do ciberespaço como um vetor criminógeno, analisando especificamente como comunidades virtuais de ódio influenciam o processo de radicalização de jovens e adolescentes. O estudo parte da premissa de que o ambiente digital não é apenas um repositório de dados, mas um espaço de interação social e formação identitária que pode acelerar a adoção de ideologias extremistas. Diante da crescente ocorrência de atos violentos no mundo físico motivados por discursos de intolerância online, a pesquisa busca examinar os limites da responsabilidade civil das plataformas digitais à luz do Marco Civil da Internet (Lei nº 12.965/2014).

A escolha do problema de pesquisa fundamenta-se na observação de que a omissão das plataformas na moderação de conteúdos ilícitos pode favorecer a criação de "câmaras de eco", onde visões extremistas são reforçadas e a motivação para crimes de ódio é catalisada. O

---

<sup>4</sup> Marco regulatório abrangente da União Europeia para intermediários e plataformas online, com principal objetivo de prevenir atividades ilegais e prejudiciais online, ao mesmo tempo em que combate a disseminação de desinformação.

problema central questiona em que medida o atual regime de responsabilidade previsto no artigo 19 da Lei 12.965/14 contribui para essa radicalização juvenil e para o consequente aumento de crimes motivados por intolerância contra grupos minoritários no Brasil. A hipótese levantada é que o regime de responsabilidade civil subjetiva é insuficiente e inadequado, pois a exigência de prévia ordem judicial para a remoção de conteúdos permite que discursos de ódio permaneçam online tempo suficiente para impulsionar jovens à violência real.

O objetivo geral deste artigo é analisar a correlação entre a inércia regulatória das plataformas e o risco criminógeno de radicalização, avaliando criticamente o arcabouço normativo brasileiro. Entre os objetivos específicos, destacam-se a conceituação da radicalização no ciberespaço, o exame da tipologia penal aplicável aos crimes de ódio no Brasil e a discussão sobre a introdução do conceito de dever de cuidado (Duty of Care<sup>5</sup>) no combate ao extremismo. A pesquisa delimita-se ao campo do Direito Penal e da Criminologia, focando no contexto brasileiro contemporâneo e no comportamento digital de jovens mediado por grupos virtuais.

A justificativa para a elaboração deste trabalho reside na inegável relevância social da temática, uma vez que a incitação à violência e crimes como racismo e xenofobia ameaçam diretamente a segurança pública e os direitos fundamentais da dignidade humana. Academicamente, o trabalho propõe um diálogo necessário entre o Direito e a Criminologia ao aplicar teorias clássicas, como a Teoria da Associação Diferencial (TAD), de Edwin Sutherland (SUTHERLAND, 1939), e as Técnicas de Neutralização, de Gresham Sykes e David Matza (SYKES; MATZA, 1957), ao novo cenário das redes sociais.

Estruturalmente, o artigo está organizado em seções que abordam, inicialmente, os mecanismos criminológicos de socialização desviante no ambiente virtual. Segue-se uma análise jurídica da fundamentação constitucional e tipificação penal dos crimes de ódio. Por fim, discute-se o regime de responsabilidade das plataformas e propostas de moderação preventiva como forma de mitigar a radicalização juvenil no Brasil.

## **2. O CIBERESPAÇO COMO FATOR CRIMINÓGENO NA RADICALIZAÇÃO JUVENIL**

---

<sup>5</sup> Obrigação jurídica e ética das plataformas de adotarem medidas proativas para proteger seus usuários, identificando e mitigando riscos sistêmicos (como desinformação, discurso de ódio e danos à saúde mental) em seus serviços.

## 2.1 O ambiente virtual como vetor de socialização desviante: o fenômeno das “câmaras de eco” e o reforço da ideologia de ódio

A arquitetura das democracias liberais pressupõe a existência de um espaço público compartilhado, onde a pluralidade de ideias e o confronto dialético fundamentam o debate político. Todavia, a transição para a esfera digital subverteu essa lógica através da formação das chamadas “câmaras de eco”. Conceituadas originalmente pelo jurista e professor norte-americano Cass R. Sunstein (SUNSTEIN, 2001), as câmaras de eco são ambientes informacionais onde o indivíduo é exposto predominantemente a discursos que ratificam suas premissas preexistentes. Se, na gênese do conceito, a seletividade era um ato de agência do usuário, a evolução tecnológica contemporânea revela uma mudança estrutural: a filtragem passou a ser operada por sistemas algorítmicos de aprendizado profundo, chamados de deep learning<sup>6</sup>, cuja função primordial reside em aprender o comportamento do usuário e direcioná-lo a conteúdo que maximize seu tempo de navegação na plataforma. As consequências desse sistema de manipulação do comportamento humano, no entanto, transcendem a mera fragmentação social, alcançando o limiar da radicalização comportamental e da violência tangível.

A tese de Sunstein conceitua o "The Daily Me", uma espécie de jornal individual personalizado criado pelo próprio usuário a partir da extensa variedade de notícias, informações e opiniões divergentes a que é apresentado diariamente, prevendo que o cidadão atuaria como editor-chefe de sua própria realidade, filtrando somente o conteúdo que lhe interessava. Curioso notar que, à época em que Sunstein desenvolveu e apresentou sua teoria, as redes sociais conhecidas atualmente, como Facebook<sup>7</sup> e Instagram<sup>8</sup>, sequer existiam. Sunstein estava prevendo o efeito apenas com base em fóruns e sites de notícias personalizados. No entanto, plataformas modernas de conteúdo rápido, como o Tik Tok<sup>9</sup>, exemplificam uma evolução drástica dessa premissa. Nelas, o filtro não é mais uma construção manual baseada em escolhas conscientes de seguimento, mas uma imposição algorítmica de baixíssima fricção.

Através da análise de micro-sinais comportamentais do usuário, como o tempo de retenção e visualização do usuário em determinado conteúdo, a reiteração do consumo de conteúdos

---

<sup>6</sup> Subcategoria da inteligência artificial e do aprendizado de máquina que utiliza redes neurais artificiais com várias camadas para ensinar os computadores a resolverem problemas complexos, processarem dados não estruturados e aprenderem por conta própria, de forma semelhante ao cérebro humano.

<sup>7</sup> Rede social voltada à conexão de pessoas, troca de mensagens e publicações de texto ou mídias.

<sup>8</sup> Plataforma focada no compartilhamento de fotos e vídeos entre os seus usuários.

<sup>9</sup> Serviço de vídeos curtos e entretenimento vertical, centrada em modelo de descoberta passiva de conteúdo orientado por inteligência artificial.

específicos e semelhantes e a reação do usuário a cada novo conteúdo apresentado, optando por seguir visualizando e interagindo até o fim ou selecionando o próximo, o algoritmo antecipa e molda o desejo do usuário antes mesmo de sua manifestação volitiva. Essa filtragem passiva elimina o que Sunstein denominava “encontros casuais com o contraditório”. O resultado é a criação de um isolamento cognitivo hermético, onde a repetição constante de narrativas curtas e emocionalmente carregadas naturaliza ideologias extremas, incapacitando o usuário de perceber que está em uma bolha, pois o conteúdo é entregue de forma contínua e viciante, criando uma realidade paralela onde certas estéticas, gírias ou ideologias parecem ser a norma global, quando são apenas o seu próprio nicho, conferindo-lhes, assim, uma falsa percepção de consenso universal.

Diferente do isolamento estético e ideológico promovido por plataformas como o TikTok, outras comunidades, como o X<sup>10</sup> (outroa chamado Twitter), operam sob a lógica da polarização por conflito como forma de gerar engajamento. Se a câmara de eco de Sunstein sugere a ausência do contraditório, o algoritmo do X frequentemente expõe o usuário ao dissenso, mas o faz de forma estratégica para gerar reatividade. A câmara deixa de ser um lugar de concordância e passa a ser um lugar de ataque coordenado ao outro.

A arquitetura da plataforma privilegia conteúdos que despertam emoções de alta ativação, como raiva e indignação moral. Mecanismos como a citação de postagem e o sistema de Trending Topics<sup>11</sup> frequentemente lançam o usuário em um estado de “vigilância tribal”. Nesse cenário, o agrupamento de indivíduos com tendências semelhantes ocorre em oposição a um inimigo comum. A câmara de eco, portanto, transmuta-se em uma “unidade de combate retórico”, onde a validação interna do grupo é obtida através da hostilização sistemática do grupo externo, exacerbando o que Sunstein identificou como a radicalização decorrente da discussão intragrupo. A indignação gera mais engajamento do que a concordância, o que cria câmaras de eco negativas, onde o grupo se une não apenas pelo que ama, mas pelo ódio comum a um alvo externo, validando a visão de Sunstein de que o debate público se transformaria, portanto, em uma guerra de trincheiras.

A convergência entre a filtragem invisível e a incitação ao conflito cria um ecossistema propício para a radicalização do comportamento. Ao serem privados de contra-argumentos e imersos em um fluxo contínuo de reforço identitário, os indivíduos sofrem um estreitamento da

---

<sup>10</sup> Plataforma de *microblogging* (baseada primariamente em textos curtos) interação em tempo real.

<sup>11</sup> Assuntos ou termos mais comentados e pesquisados em uma rede social ou plataforma digital em um determinado momento, funcionando como um termômetro em tempo real do que está chamando a atenção do público, seja no mundo ou em regiões específicas.

percepção da realidade. A literatura sociológica e psicológica contemporânea indica que esse isolamento é o primeiro passo para o desengajamento moral.

Quando os algoritmos priorizam narrativas desumanizantes em prol do engajamento, a barreira entre a agressão verbal digital e a ação física torna-se extremamente frágil. A radicalização, alimentada pela retroalimentação algorítmica, pode ultrapassar as fronteiras virtuais, culminando em crimes violentos, atentados contra instituições democráticas e atos de terrorismo doméstico. O indivíduo, condicionado pela câmara de eco a acreditar que sua visão de mundo está sob ameaça iminente e que a violência é uma resposta legítima de defesa do grupo, passa a agir como o braço executor de uma ideologia sintetizada e amplificada por processos automatizados de manipulação digital.

## 2.2 A Teoria da Associação Diferencial (TAD) aplicada ao extremismo online

Em sua obra, Sutherland postulava que o comportamento é aprendido em grupos de interação social. Nesse sentido, os princípios do processo pelo qual se desenvolve o comportamento criminoso são os mesmos do processo através do qual se desenvolve o comportamento legal, restando a diferença entre eles caracterizada na associação com pessoas que se empenham no comportamento criminoso sistemático, por meio do processo de aprendizagem onde a conduta criminal é algo que se aprende. Assim, o indivíduo desenvolve seu comportamento individual baseado nos exemplos e influências que possui e em situações de conflitos, tende a agir de forma semelhante aos seus exemplos. Para o autor, o comportamento criminoso apresenta algumas proposições: a conduta criminal se aprende em interação com outras pessoas mediante um processo de comunicação, com influência pedagógica diretamente proporcional à frequência e duração dos contatos. Dessa forma, a parte decisiva do aprendizado ocorre em meio às relações mais íntimas do indivíduo, que se converte em delinquente quando as definições favoráveis à violação da lei superam as desfavoráveis. Nesse sentido, a Teoria do Aprendizado Social, proposta pelo psicólogo canadense Albert Bandura (BANDURA, 1977), parte da hipótese de que as bases da conduta humana têm suas raízes na aprendizagem que a experiência vital diária enseja ao indivíduo. Ou seja, o homem aprende a conduta desviada e associa-se com referência nela, atuando de acordo com as reações que sua própria conduta recebe dos demais, de modo que o comportamento individual é continuamente modelado pelas experiências da vida cotidiana.

Segundo Sutherland, o crime não é hereditário nem se imita ou inventa. Não é algo fortuito ou irracional: o crime se aprende. A capacidade, destreza e motivação necessárias para o delito

se aprendem mediante o contato com valores, atitudes, definições e pautas de condutas criminais no curso de processos normais de comunicação e interação do indivíduo com seus semelhantes. Não basta viver em um meio criminógeno, nem manifestar determinados traços da personalidade ou situações frequentemente associadas ao delito. A parte decisiva do processo de aprendizagem ocorre no seio das relações mais íntimas do indivíduo com seus familiares ou com pessoas do seu meio. Para a teoria da associação diferencial, é no interior de grupos e no contato com criminosos experientes que são transmitidos aos jovens delinquentes informações e ensinamentos sobre como cometer crimes e sobre como justificar esse comportamento.

A transposição da Teoria da Associação Diferencial para o ambiente virtual exige uma releitura de seus conceitos fundamentais, adaptando-os à lógica dos algoritmos e das subculturas digitais. Originalmente, a obra de Sutherland estabelece a premissa de que o comportamento radical ou criminoso não emerge de um vácuo psicológico ou de uma predisposição inata, mas sim de um processo de aprendizado social rigoroso. No século XXI, esse aprendizado foi transposto do contato físico para a interação mediada por algoritmos. A intimidade digital não é física, mas de afinidade. Fóruns de nicho e plataformas de mensagens criptografadas, como Discord<sup>12</sup> e Telegram<sup>13</sup>, oferecem um senso de pertencimento que tendem a substituir os laços tradicionais. Nesses ambientes, sujeitos em busca de pertença identitária são inseridos em subculturas que operam como vetores de socialização desviante. A associação diferencial manifesta-se, então, através da exposição prolongada e frequente a conteúdos que operam sob uma lógica de isolamento axiológico. O algoritmo de recomendação, ao atuar como um curador de associações, intensifica a exposição do indivíduo a definições favoráveis à violação de normas democráticas e de direitos humanos, ao mesmo tempo em que mitiga o contato com discursos divergentes ou institucionais, consolidando as câmaras de eco as quais o usuário é introduzido.

No domínio específico da misoginia e do discurso de ódio direcionado a grupos minoritários, o processo de aprendizagem abrange tanto a dimensão técnica quanto a normativa. O indivíduo incorpora um vocabulário próprio, frequentemente codificado para eludir mecanismos de moderação, e assimila justificativas racionais que buscam legitimar a agressão. Assim originaram-se as comunidades vinculadas à chamada "manosfera", conjunto de comunidades online misóginas e antifeministas que promovem masculinidade tóxica,

---

<sup>12</sup> Plataforma de comunicação que permite aos usuários conversar por texto, voz e vídeo. Inicialmente popular entre comunidades de jogos eletrônicos, é atualmente utilizado por diversos grupos sociais.

<sup>13</sup> Provedor de mensagens instantâneas baseado em nuvem, focado em velocidade, segurança e sincronização, permitindo o envio de textos, mídias e arquivos de qualquer tipo, além de chamadas de voz e vídeo.

agressividade e a falsa narrativa de que a igualdade de gênero prejudica os homens, englobando subculturas como incels<sup>14</sup>, MGTOW<sup>15</sup> e pick-up artists<sup>16</sup>, disseminando ódio e radicalizando seus integrantes. Nesses espaços, a desumanização da mulher é articulada como uma premissa intelectual, permitindo que o sujeito neutralize a culpa social mediante a adoção de uma narrativa de vitimização masculina. Essa inversão de valores constitui o cerne da TAD: a predominância de definições favoráveis à transgressão sobre as definições convencionais de respeito à alteridade. Quando o balanço normativo do sujeito pende para o desvio, a conduta extremista passa a ser percebida como uma norma interna legítima e recompensada pelo prestígio intragrupal.

A aplicação da TAD às comunidades integrantes da manosphere permite compreender como o discurso de vitimização masculina é construído e assimilado como uma verdade normativa. Sob a ótica de Sutherland, a convicção de que o homem contemporâneo é uma vítima sistêmica de uma estrutura social "ginocêntrica", com tendência a colocar a figura feminina, suas necessidades, seus desejos e perspectivas como ponto central da visão de mundo, cultura e análise social, não é um fenômeno orgânico, mas uma construção ideológica aprendida por meio de interações persistentes em grupos de referência digital.

A manosphere opera mediante a inversão da realidade sociológica tradicional: o sujeito, ao ingressar nessas comunidades, é submetido a uma reeducação axiológica onde as conquistas de igualdade de gênero são ressignificadas como mecanismos de opressão masculina. Sob a ótica da TAD, essa inversão dos valores sociais concretos promove o desbalanceamento entre as definições desfavoráveis à violação das normas (neste caso, as normas de equidade entre gêneros e respeito à alteridade) e as definições favoráveis, com a superação destas sobre as primeiras. O discurso de vitimização atua como o motor dessa balança: ao convencer o jovem de que ele é o alvo de uma injustiça histórica, o grupo fornece a justificativa moral necessária para que o comportamento agressivo ou discriminatório seja percebido como uma forma de "legítima defesa" ou "resistência".

A eficácia dessa socialização é potencializada pelas variáveis de intensidade e frequência das associações. Jovens em situação de vulnerabilidade emocional ou isolamento social offline

---

<sup>14</sup> *Involuntary celibate* (celibatários involuntários): subcultura composta principalmente por homens que se sentem incapazes de conseguir parceiras sexuais ou românticas, sendo conhecida pela misoginia, ao culpar mulheres e a sociedade por sua frustração, frequentemente associada a discursos de ódio e violência extremista.

<sup>15</sup> *Men Going Their Own Way* (Homens Seguindo Seu Próprio Caminho): filosofia e movimento masculinista que defende que os homens devem se abster de relacionamentos com mulheres, focando na autossuficiência masculina, evitando a influência feminina e o que consideram um sistema social e jurídico hostil ao homem.

<sup>16</sup> Artistas da sedução: movimento de homens que utilizam técnicas de manipulação psicológica e estratégias de "jogo" para seduzir mulheres, compartilhando métodos muitas vezes vistos como predatórios, focando em aumento de autoconfiança e atração física.

encontram na mansfêra um grupo de apoio que oferece no apenas pertencimento, mas uma explicao totalizante para suas frustraoes pessoais. A retorica da redpill – conceito extraido do filme Matrix (1999), onde  apresentada ao protagonista da obra a escolha entre tomar uma pilula azul e permanecer na ignorncia sobre o real estado do mundo em que se encontra e tomar uma pilula vermelha, que revela ao personagem a realidade oculta aos olhos da maioria, conceito que, posteriormente, foi adaptado ao discurso das comunidades androcntricas, sendo a “pilula vermelha” o despertar dos homens para a realidade social onde o sistema privilegiaria as mulheres e prejudicaria os homens – funciona como a transmisso de tcnicas e direoes de motivos. O aprendizado no se limita ao discurso, mas inclui o domnio de uma gramtica prpria e de justificativas racionais que neutralizam a empatia. A vitimizao , portanto, o catalisador que transforma a frustrao individual em uma ideologia coletiva radicalizada. A associao diferencial nesses espaos consolida, portanto, o que se pode denominar de isolamento normativo.

 medida que o indivduo intensifica seus vnculos com a mansfêra, as definioes de moralidade da sociedade convencional, que condenam a misoginia e o discurso de dio, passam a ser vistas como evidncias da prpria perseguio sofrida pelos homens. Esse enclausuramento cognitivo impede que o sujeito processe contra-argumentos, uma vez que qualquer crtica externa  imediatamente classificada como parte do sistema opressor. Assim, a vitimizao masculina na mansfêra no  apenas um sentimento, mas um comportamento aprendido e reforado que, conforme previsto por Sutherland, orienta a conduta do indivduo para o conflito com os valores democrticos e humanos da sociedade.

### 2.3 O processo de radicalizao juvenil: a transio do discurso de dio virtual para a motivao de violncia real

A transio do discurso de dio virtual para a efetiva materializao da violncia no mundo fsico constitui um dos fenmenos mais crticos da criminologia contempornea, podendo ser elucidada atravs da transposio dos pressupostos da TAD para o ecossistema digital. Sob a tica de Sutherland, sendo a conduta delitiva produto de aprendizagem social em que o indivduo assimila no apenas as tcnicas para a execuo do ato, mas, fundamentalmente, as orientaoes especficas de motivos, impulsos e racionalizaoes, no contexto da radicalizao juvenil em redes sociais e comunidades da mansfêra, esse aprendizado  catalisado por uma infraestrutura tecnolgica que opera como um vetor de socializao desviante de alta

intensidade. O processo de radicalização inicia-se com a inserção do jovem em grupos de referência virtual como comunidades em plataformas de mensagens, fóruns e redes sociais, que passam a exercer a função dos grupos pessoais íntimos descritos por Sutherland. Nesses espaços, a prioridade, a duração e a intensidade das interações digitais superam as influências de instituições tradicionais, como a família e a escola. A transição para a violência real ocorre quando o balanço normativo do indivíduo é alterado por um excesso de definições favoráveis à violação da lei. Na manosphere, a vitimização masculina é articulada como uma injustiça sistêmica que demanda retribuição, transformando o ódio contra mulheres e minorias em uma imperativa moral. A violência deixa de ser vista como um desvio e passa a ser compreendida como uma resposta legítima e necessária a uma suposta agressão social.

A migração do discurso para a ação concreta é viabilizada pelo aprendizado de técnicas de neutralização moral, que permitem ao sujeito dissociar-se da gravidade de seus atos. O aprendizado preconizado por Sutherland, no ambiente online, assume formas de gamificação, onde a execução de atos violentos ou assédios coordenados é incentivada por sistemas de hierarquia e prestígio intragrupal. Estudos recentes sobre ataques em instituições de ensino no Brasil demonstram que a maioria dos agressores apresentava um histórico de engajamento em subculturas extremistas, onde a idolatria a perpetradores de massacres anteriores servia como um modelo de conduta a ser emulado. A internet, portanto, não apenas dissemina o ódio, mas fornece um arcabouço de justificativas e técnicas que encoraja o indivíduo a transpor a fronteira do mundo virtual, buscando no mundo concreto a validação e a notoriedade cultivadas no anonimato das redes. Nas subculturas, o processo de radicalização culmina no que a literatura especializada denomina efeito contágio ou mimetismo delitivo: a tendência de repetição de crimes violentos ou comportamentos desviantes após a sua divulgação intensa pela mídia ou redes sociais. O fenômeno ocorre quando indivíduos, frequentemente com características psicológicas similares ou motivações preexistentes, imitam o método, cenário ou alvo de um agressor original. No contexto das subculturas extremistas sob a ótica da TAD, essa direção é fornecida pela sacralização de perpetradores passados. Autores de massacres icônicos, como os de Columbine (Colorado, 1999), Realengo (Rio de Janeiro, 2011) e Suzano (São Paulo, 2019), são elevados ao status de “saints<sup>17</sup>” dentro dessas comunidades. O jovem radicalizado não apenas consome informações sobre esses ataques, mas estuda os manifestos e táticas dos perpetradores anteriores como se fossem manuais de instrução. Nesse contexto, associação diferencial ocorre

---

<sup>17</sup> Santo: assassinos em massa que são “beatificados” e transformados em ídolos ou mártires por fóruns extremistas.

de forma assíncrona: o indivíduo se associa ao "legado" do criminoso, buscando superá-lo em número de vítimas ou em impacto mediático, em uma espécie de "gamificação" da morte. Nessas subculturas, o sucesso de um ataque é medido por métricas de videogames (high score<sup>18</sup>), o que desumaniza as vítimas e transforma a violência real em uma performance para o público virtual.

O engajamento em ambientes como a mansferra e os chans<sup>19</sup> funciona como o "grupo pessoal íntimo" definido por Sutherland, onde as definições favoráveis à violação da lei tornam-se hegemônicas. Pesquisas do Grupo de Trabalho sobre Violência nas Escolas do Ministério da Educação (BRASIL, 2023) indicam que o ódio contra mulheres é frequentemente o primeiro estágio da radicalização. A frustração afetiva é canalizada para as comunidades incel, onde o jovem aprende a culpar grupos minoritários por sua condição social. No contexto brasileiro, as subculturas desenvolveram léxicos próprios e estéticas visuais, como uso de máscaras de caveira (fazendo referências a grupos neonazistas) que reforçam a identidade do grupo e sinalizam a prontidão para o conflito real. O isolamento em relação à moralidade convencional é completo: o mundo fora das redes é visto como hostil, justificando o ataque como uma forma de purificação ou vingança.

Nesse sentido, a transição para o mundo concreto ocorre quando o suporte ideológico do grupo valida a ação. Casos como o de Aracruz (ES), em 2022, e de Blumenau (SC), em 2023, demonstram que os agressores não agiram de forma isolada, mas estavam imersos em redes de reforço diferencial. Em muitos desses casos, descobriu-se que os agressores compartilharam planos ou intenções em fóruns e grupos de mensagens, recebendo encorajamento e instruções técnicas de outros membros. Aqui, a associação diferencial atua em sua forma mais pura: o prestígio dentro da subcultura torna-se mais valioso para o jovem do que a própria liberdade ou vida.

Em suma, a radicalização juvenil é um processo de aculturação onde a associação diferencial com grupos extremistas redefine a percepção da realidade e da moralidade do jovem. A transição para a violência física é o estágio terminal de uma trajetória iniciada na aceitação de definições agressivas e na erosão sistemática da alteridade, culminando em um estado de prontidão para a ação em que o inimigo construído virtualmente passa a ser alvo de agressões reais. O combate a esse fenômeno exige, portanto, uma intervenção que compreenda as redes sociais não meramente como ferramentas de comunicação, mas como espaços de formação de identidade e conduta

---

<sup>18</sup> Maior pontuação obtida pelo jogador em uma sessão de determinado jogo eletrônico.

<sup>19</sup> Fóruns anônimos, frequentemente usados como espaços para discursos extremistas, vazamento de fotos íntimas e ataques coordenados contra mulheres.

onde o crime, conforme Sutherland observou, é aprendido com a mesma naturalidade que qualquer outra atividade social.

### **3. CRIMES DE ÓDIO E A TIPOLOGIA PENAL BRASILEIRA**

#### **3.1 A fundamentação constitucional do combate ao discurso de ódio**

Para compreender de que forma o ambiente digital opera como vetor criminógeno capaz de impulsionar a delinquência e a violência offline, faz-se indispensável analisar o arcabouço jurídico-constitucional que legitima a intervenção estatal sobre tais condutas. No ecossistema da internet, subculturas extremistas e chans frequentemente utilizam a garantia da liberdade de expressão, proposta no artigo 5º, incisos IV e IX, da Constituição de 1988, como um escudo absoluto para justificar a propagação de ideologias misóginas, racistas, homofóbicas e antidemocráticas. Contudo, o ordenamento jurídico brasileiro adota o paradigma de uma democracia militante, no qual nenhum direito fundamental possui caráter absoluto ou pode ser exercido de forma a aniquilar os direitos alheios.

O limite à manifestação do pensamento encontra-se ancorado, primordialmente, no princípio da dignidade da pessoa humana (CRFB/88, art. 1º, III) e nos objetivos fundamentais de promover o bem de todos, sem preconceitos de origem, raça, sexo, cor ou quaisquer outras formas de discriminação (CRFB/88, art. 3º, IV). Quando o discurso proferido nas redes ultrapassa a barreira da mera opinião crítica e transforma-se em ferramenta de desumanização e incitação à violência contra minorias sociais, configura-se o abuso do direito. Conforme consolidado pelo Supremo Tribunal Federal (STF) no emblemático julgamento do Caso Ellwanger<sup>20</sup>, os direitos fundamentais não podem se transformar em um salvo-conduto para a prática de atos ilícitos. A liberdade de expressão, portanto, encontra barreiras intransponíveis na proteção aos direitos da personalidade, na garantia da inviolabilidade da honra e da imagem das pessoas (CRFB/88, art. 5º, X) e na expressa vedação constitucional ao anonimato (CRFB/88, art. 5º, IV), o qual é amplamente explorado no ambiente digital para assegurar uma falsa sensação de impunidade aos emissores do ódio.

A conexão entre a perspectiva criminológica e a dogmática constitucional ganha contornos práticos ao se analisar o processo de radicalização juvenil mediado por algoritmos e interações em plataformas digitais. Uma vez que o comportamento delitivo, à luz da TAD, é

---

<sup>20</sup> Habeas Corpus nº 82.424/RS. Relator para o Acórdão: Min. Maurício Corrêa. Julgamento: 17/09/2003.

aprendido por meio da convivência íntima com grupos que apresentam definições favoráveis à violação da lei, no contexto do ciberespaço contemporâneo, a exposição contínua dos jovens a esses fóruns extremistas e subcomunidades que disseminam e endossam discursos violentos promove a assimilação de tais discursos e técnicas de neutralização moral. Como reflexo direto dessa socialização nociva, observa-se o avanço fático de fenômenos de extrema gravidade, a exemplo do planejamento coordenado de ataques violentos em massa contra instituições de ensino no Brasil.

Diante dessa transição, onde o ódio cultivado no plano virtual se materializa em tragédias no mundo real, o combate a essas condutas não se configura mais como mera opção política, passando a ter status de mandamento constitucional. A Carta Magna de 1988 estabelece rigorosos mandados de criminalização, determinando que a lei punirá qualquer discriminação atentatória dos direitos e liberdades fundamentais (art. 5º, XLI) e definindo a prática do racismo como crime inafiançável e imprescritível (art. 5º, XLII). Sob esse prisma, a atuação jurisdicional do STF estendeu a proteção contra o racismo às condutas de homofobia e transfobia, além de equiparar a injúria racial ao crime de racismo, oferecendo substrato legal rígido para coibir ofensas a grupos vulnerabilizados na internet. A proteção da juventude contra esses mecanismos de cooptação extremista em plataformas digitais encontra guarida no artigo 227 da Constituição, que prevê: “É dever da família, da sociedade e do Estado assegurar à criança, ao adolescente e ao jovem, com absoluta prioridade, o direito à vida, à saúde, à alimentação, à educação, ao lazer, à profissionalização, à cultura, à dignidade, ao respeito, à liberdade e à convivência familiar e comunitária, além de colocá-los a salvo de toda forma de negligência, discriminação, exploração, violência, crueldade e opressão”. Nota-se no texto o destaque dado pelo legislador constituinte, especialmente pela previsão expressa, ao dever de garantia do direito à vida, à dignidade, ao respeito e à convivência comunitária de forma segura, com proteção de toda forma de negligência, discriminação, exploração, violência e crueldade, desde o microsistema do afeto ao macrossistema das leis e ações estatais. Desse modo, o desmantelamento das redes de radicalização online e a punição de seus articuladores constituem medidas imperativas para a preservação da própria integridade física e psíquica da juventude brasileira.

Uma vez delineada a ilicitude do discurso de ódio sob a ótica dos direitos fundamentais, o debate desloca-se necessariamente para a responsabilidade civil e administrativa dos intermediários digitais (provedores de aplicação) que operam e lucram com a arquitetura dessas redes. A Lei nº 12.965/2014 (Marco Civil da Internet), em seu artigo 19, estabelece: “Com o intuito de assegurar a liberdade de expressão e impedir a censura, o provedor de aplicações de

internet somente poderá ser responsabilizado civilmente por danos decorrentes de conteúdo gerado por terceiros se, após ordem judicial específica, não tomar as providências para, no âmbito e nos limites técnicos do seu serviço e dentro do prazo assinalado, tornar indisponível o conteúdo apontado como infringente, ressalvadas as disposições legais em contrário”. Esse modelo visava proteger a liberdade de expressão e evitar a censura privada prévia pelos provedores. Entretanto, o atual cenário criminógeno verificado na rede exige uma urgente filtragem constitucional e releitura hermenêutica do dispositivo. A imunidade conferida pelo artigo 19 do Marco Civil da Internet não pode se converter em um salvo-conduto para a inércia corporativa diante de redes coordenadas de desinformação, apologia ao crime e discursos que ameaçam diretamente a integridade física de crianças, adolescentes e jovens. Sob a ótica da eficácia horizontal dos direitos fundamentais e do dever de proteção estatal, o direito de propriedade e a livre iniciativa das big techs<sup>21</sup> não podem prevalecer sobre a segurança coletiva e a dignidade humana.

Nesse contexto, o debate travado no Supremo Tribunal Federal no âmbito do Tema 987 da Repercussão Geral sinaliza a necessidade de superação do modelo puramente reativo do artigo 19 do Marco Civil da Internet. Argumenta-se que, diante da sofisticação de algoritmos que impulsionam e monetizam conteúdos com base no engajamento gerado pela indignação e pelo ódio, as plataformas assumem um papel ativo na distribuição da informação. Consequentemente, impõe-se a responsabilização solidária ou a exigência de deveres de cuidado rigorosos (autorregulação regulada), obrigando os provedores de aplicação a removerem de forma célere conteúdos manifestamente criminosos independentemente de provocação judicial, sempre sob as balizas do devido processo legal e da proporcionalidade.

### 3.2 Análise dos tipos penais aplicáveis à incitação à violência: a Lei nº 7.716/89 (Lei do Racismo) e sua adequação ao contexto virtual

A transição das condutas intolerantes dos espaços físicos tradicionais para o ecossistema digital exige do operador do Direito uma análise rigorosa da tipicidade penal e de sua aptidão para salvaguardar bens jurídicos de envergadura coletiva. No ordenamento jurídico brasileiro, o principal instrumento repressivo contra as manifestações sistêmicas de ódio encontra-se na Lei nº 7.716/1989 (Lei do Racismo), com destaque para o artigo 20, que versa: “Praticar, induzir ou

---

<sup>21</sup> Empresas globais do ramo de tecnologia com alto poder de mercado, capital e influência para ditar tendências econômicas e digitais.

incitar a discriminação ou preconceito de raça, cor, etnia, religião ou procedência nacional.”, tipificando, assim, na seara criminal as condutas mais comumente praticadas nessas comunidades. Diferentemente dos crimes contra a honra tipificados no Código Penal (como a difamação ou a injúria em sua matriz clássica), que resguardam a reputação e a integridade moral individualizada de um sujeito determinado, o artigo 20 da Lei do Racismo tutela a dignidade de toda uma coletividade ou grupo social vulnerabilizado. O núcleo do tipo abrange três verbos de ação: praticar (realizar atos materiais de segregação), induzir (incutir de forma direta a ideia discriminatória na mente de outros indivíduos) e incitar (estimular ou provocar publicamente o preconceito). Trata-se de crime de perigo abstrato e de mera conduta, cuja consumação exige a ocorrência de um resultado de efetivo nexo causal, como a efetiva eclosão de um ato violento físico, bastando que a manifestação do pensamento seja apta a vulnerar a paz pública e o sentimento de isonomia social. Ademais, a elasticidade interpretativa desse dispositivo foi consolidada pelo Supremo Tribunal Federal que, por meio de uma interpretação conforme a Constituição, enquadrando as práticas de homofobia e transfobia no conceito social e jurídico de racismo. Com isso, os discursos proferidos na rede que visam desumanizar minorias sexuais ou de gênero passaram a atrair as severas sanções da Lei nº 7.716/1989, firmando o caráter abrangente da norma no combate ao discurso de ódio.

O advento da internet transformou substancialmente o potencial ofensivo do discurso discriminatório. Se no mundo analógico a propagação do ódio limitava-se ao alcance da voz ou à panfletagem restrita, a arquitetura das redes sociais deu ao infrator um poder sem precedentes. Considerando essa nova dinâmica, o legislador ordinário implementou profundas modificações normativas por meio da Lei nº 14.532/2023, que conferiu nova redação ao parágrafo 2º do artigo 20 da Lei nº 7.716/1989. Conforme os termos atuais do dispositivo legal: “se qualquer dos crimes previstos neste artigo for cometido por intermédio dos meios de comunicação social, de publicação em redes sociais, da rede mundial de computadores ou de publicação de qualquer natureza”, a pena abstrata salta do patamar comum, de reclusão de um a três anos, para o patamar qualificado de reclusão, de dois a cinco anos, e multa. A justificativa para o recrudescimento consiste no severo desvalor do meio empregado. Os algoritmos de engajamento das grandes corporações tecnológicas operam mediante a lógica de monetização do conflito e da indignação, o que faz com que publicações contendo ataques odiosos viralizem de forma exponencial, perpetuando-se no ciberespaço e impondo dano perene à dignidade dos grupos afetados. Essa adequação específica ao contexto virtual é de suma importância para dismantelar a atuação de células extremistas operantes na mansfêra, nos chans e em comunidades privadas

de conversação. Nestes redutos, discursos que incitam o menosprezo e a violência contra mulheres e minorias étnicas encontram ressonância imediata, servindo de base ideológica para processos de radicalização de jovens. A incidência da qualificadora do § 2º afasta a possibilidade de institutos despenalizados e sinaliza uma resposta penal rigorosa do Estado face à delinquência cibernética.

Apesar do aperfeiçoamento legislativo operado pela Lei nº 14.532/2023, a aplicação prática do artigo 20, § 2º, da Lei do Racismo esbarra em complexos desafios probatórios no campo do elemento subjetivo do tipo. A jurisprudência exige, de forma pacífica, a demonstração inequívoca do dolo específico, ou seja, a vontade livre e consciente de discriminar, segregar ou inferiorizar um grupo humano em razão de suas características essenciais. Essa exigência técnica frequentemente colide com estratégias de defesa que invocam o *animus jocandi* — a intenção de fazer piada ou sátira — como causa de atipicidade por ausência de dolo. Refletindo essa complexidade hermenêutica, a Sexta Turma do Superior Tribunal Federal (STJ), em julgamento proferido em dezembro de 2025, absolveu réu acusado de racismo qualificado que havia postado em seu perfil do Facebook uma foto portando um simulacro de arma de fogo com a legenda: "Pronto pra matar gays, nordestinos, negros, quilombolas... e as mulheres? elas não! kkkkkkkk"<sup>22</sup>. Na oportunidade, o colegiado entendeu que, apesar do manifesto mau gosto e acidez da postagem, a conduta ocorreu em um contexto isolado de crítica e deboche a posturas conservadoras, pairando dúvida razoável quanto à real intenção de incitar a violência ou a segregação racial, o que culminou na aplicação do princípio do *in dubio pro reo*.

Todavia, sob a perspectiva criminológica moderna, a aceitação indiscriminada do argumento do “humor” na internet oferece sérios riscos sociais. Conforme evidenciado nas dinâmicas de grupos de radicalização online, o uso de “memes”<sup>23</sup> preconceituosos, piadas irônicas e estéticas satíricas constitui uma tática deliberada de dessensibilização moral e neutralização da culpa dos jovens cooptados. O humor hostil serve como camada protetiva para introduzir discursos odiosos no cotidiano juvenil sem disparar os alarmes da censura social. Portanto, quando a manifestação aparentemente jocosa caminha lado a lado com o engajamento em subculturas que planejam atentados contra a vida, violando, também, o artigo 286 do Código Penal (Incitação ao Crime) e desaguando em episódios de violência extrema agora reprimidos com maior rigor penal pela Lei nº 15.159/2025 (Crimes em Escolas), revela-se o dolo de vulnerar os pilares constitucionais da

---

<sup>22</sup> REsp Nº 2015530 - RS (2022/0226582-4).

<sup>23</sup> Imagens, vídeos e áudios que viralizam na internet, sendo primordialmente conhecidos por sua utilização como piadas, mas também podendo ser utilizados em campanhas publicitárias, forma de linguagem e em divulgações de marcas e serviços online.

igualdade e da alteridade. Nesse sentido, cabe ao Poder Judiciário realizar uma minuciosa filtragem contextual do meio de propagação e da reiteração das condutas para impedir que o manto protetor da liberdade de expressão ou do humor seja utilizado como validador de discursos voltados à aniquilação dos direitos da personalidade de grupos vulneráveis.

### 3.3 A motivação de ódio como elemento subjetivo e qualificadora penal

Para além da tipificação objetiva de condutas discriminatórias, a dogmática penal contemporânea enfrenta a necessidade de decodificar a estrutura psíquica que impulsiona o agente criminoso no denominado crime de ódio. Sob a ótica do Direito Penal, o ódio não deve ser compreendido como um mero sentimento irrelevante para a esfera jurídica, mas sim como um elemento subjetivo especial do injusto, tradicionalmente denominado pela doutrina clássica como dolo específico. Trata-se da intenção voltada não apenas à produção do resultado do crime base (como matar, lesionar ou ameaçar), mas sim à hostilização de uma vítima selecionada especificamente por sua pertença a um grupo social, étnico, religioso ou de gênero determinado.

Para a literatura criminológica internacional, o crime de ódio possui uma natureza essencialmente simbólica e instrumental. A vítima imediata funciona como mero representante de uma categoria indesejada pelo agressor, de forma que a violência praticada visa enviar uma mensagem de intimidação e exclusão para toda a coletividade à qual ela pertence. No plano técnico-jurídico, a identificação desse elemento subjetivo especial exige do operador do Direito uma rigorosa análise probatória, uma vez que a motivação discriminatória precisa restar plenamente demonstrada nos autos para afastar a capitulação da conduta como um delito comum.

No ecossistema digital, a exteriorização desse elemento subjetivo especial do injusto é frequentemente facilitada pelos vestígios deixados pelos próprios autores. Em comunidades virtuais fechadas e fóruns radicais, os agentes costumam publicar manifestos textuais, interagir por meio de eixos temáticos preconceituosos e celebrar estéticas extremistas. Esses elementos servem como evidências da fixação ideológica do indivíduo, de forma a ser possível reconstruir o animus discriminatório que precedeu e orientou a execução material do delito.

Apesar da falta de conceituação da forma autônoma do “crime de ódio” no Código Penal brasileiro como uma categoria genérica de agravamento de pena para todos os delitos, o legislador e a jurisprudência buscam consolidar a canalização jurídica do preconceito por meio de circunstâncias qualificadoras específicas, situadas notadamente no crime de homicídio (artigo

121 do Código Penal). Quando a violência é desencadeada por preconceito racial, intolerância religiosa ou homofobia/transfobia, a conduta tende a ser enquadrada como motivo torpe, nos termos do § 2º, inciso I, do artigo 121. A torpeza penal reside justamente no caráter repugnante da motivação, que causa profunda repulsa social ao eleger a negação da alteridade alheia como pretexto para a ceifa de uma vida humana. O Supremo Tribunal Federal, no julgamento da Ação Direta de Inconstitucionalidade por Omissão (ADO) nº 26, chancelou essa exegese ao fixar o entendimento de que a violência motivada por homofobia configura homicídio qualificado por motivo torpe, dada a manifesta incompatibilidade desse agir com os preceitos do Estado Democrático de Direito. Paralelamente, à vertente misógina do discurso de ódio, amplamente cultivada nas subculturas mencionadas anteriormente, encontra repulsa penal específica na tipificação agora autônoma do crime de feminicídio (art. 121-A do Código Penal). Antes qualificadora do crime de homicídio, o feminicídio é configurado quando este é praticado contra a mulher por razões da condição de sexo feminino, o que envolve, nos termos § 1º, inciso II, do mesmo artigo, menosprezo ou discriminação à condição de mulher. Quando jovens radicalizados em redes virtuais passam a professar um ódio sistêmico ao gênero feminino, culpabilizando as mulheres por suas frustrações afetivo-sociais e deliberando sobre atos de agressão violenta, está configurado o elemento subjetivo do menosprezo, transmutando o dolo homicida comum em dolo qualificado pela opressão de gênero.

A compreensão do ódio como qualificadora penal atinge seu ápice analítico ao se observar o fenômeno dos ataques de violência extrema direcionados a instituições de ensino no Brasil. A análise empírica desses eventos revela um nexo de causalidade criminógena inequívoco entre o consumo de discursos odiosos no ciberespaço e a execução fática de massacres no mundo real. Conforme postula a Teoria da Associação Diferencial, o processo de socialização em subculturas desviantes fornece ao indivíduo tanto o aprendizado técnico do crime quanto às justificativas morais necessárias para neutralizar sua culpa. Nesses espaços de sociabilidade, o ódio contra as minorias e o ressentimento contra a comunidade escolar são elevados ao status de dogmas ideológicos. Os agressores, imersos em dinâmicas de radicalização mútua, passam a enxergar as escolas como alvos preferenciais para a exteriorização de seus impulsos destrutivos, buscando notoriedade perante seus pares virtuais. O elemento subjetivo especial, gestado no ambiente virtual, materializa-se no mundo real por meio de planos meticulosos envolvendo o emprego de armas de fogo, armas brancas e táticas de guerrilha urbana.

Em resposta a essa escalada de violência perpetrada por jovens radicalizados, o ordenamento jurídico brasileiro foi severamente recrudescido por meio da aprovação da Lei nº

15.159/2025 (Lei de Crimes em Escolas), que alterou o Código Penal e a Lei de Crimes Hediondos (Lei nº 8.072/1990) para instituir que qualquer crime cometido nas dependências de instituições de ensino passa a figurar como circunstância de extrema gravidade, introduzindo causas de aumento de pena de 1/3 até a metade se o homicídio for praticado em ambiente escolar. A união das qualificadoras subjetivas, como o motivo torpe oriundo do preconceito ou o feminicídio, com as novas majorantes de caráter objetivo estabelecidas pela Lei de Crimes em Escolas confere ao Judiciário um arsenal repressivo robusto. Reconhece-se, assim, que os crimes nascidos das bolhas de ódio virtuais e executados no seio de escolas violam de forma simultânea a segurança pública, a incolumidade da infância e os pilares constitucionais da igualdade, exigindo reprimenda estatal máxima e exemplar.

#### **4. RESPONSABILIDADE DAS PLATAFORMAS E OS LIMITES DO MARCO CIVIL DA INTERNET (LEI 12.965/14)**

4.1 O regime de responsabilidade civil por conteúdo de terceiros: a regra da responsabilidade subjetiva e a exigência de ordem judicial para remoção de conteúdo

A conformação do ciberespaço como vetor criminógeno capaz de viabilizar a cooptação e a radicalização de jovens em células extremistas virtuais impõe uma indispensável análise sobre o regime jurídico da responsabilidade civil aplicável às plataformas de mídias sociais. No ordenamento brasileiro, o eixo normativo regulador da matéria encontra-se no artigo 19 do Marco Civil da Internet, que instituiu, como regra geral, o princípio da responsabilidade civil subjetiva condicionada das plataformas digitais (provedores de aplicação de internet) por danos decorrentes de conteúdos gerados por seus usuários, estabelecendo a necessidade de prévia e específica ordem judicial de exclusão do material para que reste configurado o dever de indenizar em caso de descumprimento.

Historicamente, a opção político-legislativa adotada pelo constituinte e pelo legislador ordinário em 2014 pautou-se na premissa de salvaguardar a liberdade de expressão, o pluralismo de ideias e evitar a instauração de um mecanismo de censura privada prévia. Sob a égide desse modelo puramente reativo (denominado pela doutrina como judicial notice-and-takedown), transferiu-se ao Judiciário o monopólio da análise da ilicitude das manifestações do pensamento na internet. Buscava-se impedir que as corporações tecnológicas, temendo condenações indenizatórias pecuniárias, suprimissem preventivamente discursos legítimos, manifestações

artísticas ou críticas políticas contundentes de seus usuários, convertendo-se em juízas particulares do debate público. Entretanto, a evolução da criminalidade digital e o surgimento de sistemas algorítmicos voltados à maximização do engajamento, os quais, conforme evidenciado por Eli Pariser, geram “filtros invisíveis” e “câmaras de eco isoladas”, fragilizaram a suficiência desse modelo. Chans, subcomunidades misóginas da manosphere e redes coordenadas de desinformação passaram a proliferar de forma livre no plano da surface<sup>24</sup> e da dark web<sup>25</sup>, operando um processo contínuo de dessensibilização moral e incitação a massacres em escolas sem que houvesse uma pronta contenção por parte das plataformas digitais, as quais se amparavam na blindagem jurídica ofertada pela redação literal do artigo 19 do MCI para justificar sua inércia corporativa.

Diante desse cenário de erosão do tecido social e de ameaça direta à integridade de direitos fundamentais básicos, o STF operou uma histórica e profunda virada jurisprudencial ao julgar, de forma conjunta, o Tema 987 da Repercussão Geral<sup>26</sup> e o Tema 533<sup>27</sup>, fixando a tese do reconhecimento da inconstitucionalidade parcial e progressiva do artigo 19 do Marco Civil da Internet, assentando a existência de um estado de omissão parcial do legislador, uma vez que a exigência irrestrita de ordem judicial prévia confere uma proteção deficiente a bens jurídicos de máxima importância constitucional, como a segurança pública, a dignidade humana e a higidez do regime democrático. Com essa histórica guinada hermenêutica, a Suprema Corte brasileira estabeleceu parâmetros claros de modulação que redefiniram o equilíbrio entre liberdade de manifestação e responsabilidade civil corporativa. Nos termos do entendimento fixado pelo Tribunal Pleno, os provedores de aplicação de internet passam a estar diretamente sujeitos à responsabilização civil por atos ilícitos e crimes gerados por terceiros em casos de falha sistêmica, sendo esta configurada quando a plataforma deixa de adotar adequadas e tempestivas medidas de prevenção, moderação ou remoção de conteúdos manifestamente criminosos, como apologia ao nazismo, racismo estrutural, incitação a atentados terroristas ou massacres escolares e deepfakes<sup>28</sup> misóginos, violando frontalmente o seu dever de atuar na rede de forma responsável, transparente e cautelosa. Insta ressaltar, contudo, que a Suprema Corte manteve o regime protetivo estrito do artigo 19 do Marco Civil da Internet especificamente para as

---

<sup>24</sup> Camada da internet e acessível a qualquer pessoa, com sites indexados por motores de busca padrão. Representam a maior parte da internet utilizada no dia-a-dia.

<sup>25</sup> Pequena parte criptografada e oculta da internet. Não pode ser encontrada por buscadores comuns e é conhecida por exigir softwares específicos para acesso, que mascaram o endereço IP e garantem o anonimato.

<sup>26</sup> RE 1.037.396/SP

<sup>27</sup> RE 1.057.258/SP

<sup>28</sup> Técnica baseada em inteligência artificial capaz de sintetizar áudios e vídeos, sobrepondo as feições de uma pessoa ao corpo de outra e/ou manipulando sons para produzir uma experiência humana realística.

hipóteses de crimes contra a honra isolados (como calúnia, difamação e injúria de caráter estritamente individual), nos quais a caracterização da ilicitude ainda remanesce dependente de prévia e específica ordem judicial, ressalvada a faculdade de remoção voluntária mediante notificação extrajudicial do ofendido. Contudo, para o macrofenômeno do discurso de ódio coletivo, das redes estruturadas de desinformação e das contas inautênticas (perfis falsos ou bots<sup>29</sup> voltados à difamação sistemática), a inércia da plataforma após denúncia interna dos usuários atrai a responsabilidade civil solidária pelos danos causados, independentemente de provocação do aparato judicial. Dessa forma, a adequação do artigo 19 do MCI ao contexto virtual contemporâneo afasta a tese da irresponsabilidade absoluta das big techs. O ordenamento jurídico brasileiro, por meio da interpretação constitucional vinculante dada pelo STF, reconhece que o modelo de negócios baseado no lucro gerado pelo engajamento do ódio impõe às plataformas digitais o dever jurídico de tutelar preventivamente o ambiente que gerenciam. A liberdade de expressão permanece garantida como pilar da democracia, mas deixa de servir como salvo-conduto para o financiamento e o impulsionamento da violência extrema offline, estabelecendo-se um novo marco na contenção civil do discurso de ódio no ciberespaço.

#### 4.2 O debate sobre o Duty of Care

O debate contemporâneo sobre a governança da internet gira em torno da superação do mito da neutralidade técnica das redes de mídias sociais. Se no início da consolidação legislativa da internet os provedores de aplicação eram encarados como meros condutores passivos de transmissão de dados, a sofisticação das técnicas de análise de dados, o direcionamento comportamental e o impulsionamento algorítmico transmutaram essas corporações em arquitetas do espaço público digital. Diante dessa transição, a dogmática jurídica internacional introduziu o conceito de Duty of Care (dever de cuidado proativo ou dever de diligência), exigindo que as plataformas assumam uma postura ativa na mitigação de riscos sistêmicos inerentes aos seus próprios modelos de negócios. Inspirado nas diretrizes do Regulamento de Serviços Digitais da União Europeia (Digital Services Act — DSA), o duty of care traduz-se na obrigação jurídica imposta às empresas de mídias sociais de desenharem suas interfaces, termos de uso e sistemas de recomendação de modo a prevenir ativamente a difusão de danos graves à coletividade. No contexto criminógeno do ciberespaço, onde bolhas algorítmicas atuam como catalisadoras das

---

<sup>29</sup> Contas automatizadas criadas para engajar, seguir pessoas, compartilhar conteúdos ou gerar métricas de forma parcial ou totalmente autônoma.

subculturas de ódio e facilitam a associação diferencial de jovens predispostos à violência, a imposição de um dever de cuidado proativo ataca diretamente a inércia lucrativa das plataformas. Sob essa perspectiva, a responsabilidade civil das corporações tecnológicas deixa de ser aferida exclusivamente de forma reativa, isto é, após a ciência de uma ilegalidade específica, passando a abranger uma dimensão preventiva de gestão de riscos coletivos.

No cenário jurídico brasileiro, o julgamento conjunto dos Temas 987 e 533 pelo Plenário do STF institucionalizou formalmente o duty of care como baliza hermenêutica imperativa. Ao reconhecer a inconstitucionalidade parcial e progressiva do artigo 19 do Marco Civil da Internet, a Suprema Corte consolidou o entendimento de que a atuação puramente reativa por parte dos provedores violava os deveres estatais de proteção aos direitos fundamentais de segurança, igualdade e dignidade. Conforme a tese, as plataformas passam a ter a obrigação de agir de forma proativa, célere e diligente na moderação de conteúdos que configurem crimes graves tipificados. O rol taxativo estabelecido pela Corte abarca práticas manifestamente lesivas, como atos terroristas, tentativas de abolição violenta do Estado Democrático de Direito, apologia ao nazismo, racismo, homofobia, feminicídio e, notadamente, a instigação ou auxílio ao suicídio, à automutilação e à violência contra crianças e adolescentes. Nesse sentido, o algoritmo não pode atuar como vetor cego diante de barbáries manifestas, devendo as corporações empregar ferramentas tecnológicas avançadas de detecção preditiva para estancar o fluxo de conteúdos criminosos hediondos antes que estes gerem efeitos materiais irreversíveis. Sob o ponto de vista da responsabilidade civil subjetiva, o descumprimento desse dever de diligência configura a denominada falha sistêmica, caracterizando-se pela ausência de adoção de medidas razoáveis e eficazes de prevenção, monitoramento estrutural ou remoção tempestiva desses materiais gravíssimos, violando o dever jurídico dos provedores de atuar de forma responsável, transparente e cautelosa. Nessa hipótese de deficiência organizacional, a plataforma assume o risco do dano coletivo e responde civil e solidariamente pelos prejuízos decorrentes de conteúdos gerados por terceiros, independentemente da existência de uma ordem judicial de remoção prévia.

Apesar da necessidade imperiosa de conter a proliferação do extremismo online, a imposição do duty of care suscita severas preocupações no campo das liberdades públicas, alimentando uma calorosa controvérsia doutrinária. O cerne do debate reside no risco de transferência do poder de polícia estatal para o arbítrio de corporações transnacionais. Ao serem compelidas a monitorar proativamente a rede sob a ameaça de severas sanções e responsabilização civil solidária, as plataformas digitais tendem a adotar uma postura

exageradamente reativa e ultraconservadora na moderação de conteúdos. Esse fenômeno, denominado *chilling effect*<sup>30</sup>, manifesta-se quando os sistemas automatizados de filtragem e os moderadores privados, operando sob uma lógica de aversão ao risco jurídico, passam a suprimir de forma indiscriminada conteúdos legítimos, críticos ou satíricos. Manifestações artísticas contestadoras, investigações jornalísticas, debates acadêmicos e peças de humor ácido correm o risco de ser sumariamente silenciados por “falsos positivos” gerados por inteligências artificiais incapazes de interpretar contextos sutis, ironias ou metáforas. Para atenuar esses riscos e impedir a privatização da censura, a jurisprudência do STF determinou que a aferição da falha sistêmica deve guiar-se pelos princípios da proporcionalidade e da autorregulação regulada. Impõe-se às plataformas o dever de editar normas internas claras, transparentes e auditáveis, que prevejam obrigatoriamente ferramentas como: sistemas acessíveis de notificação extrajudicial para os usuários, mecanismos de garantia do devido processo tecnológico, assegurado o direito de ampla defesa e contraditório ao usuário cujo conteúdo foi sinalizado ou removido, bem como a publicação de relatórios anuais de transparência detalhando as métricas de remoção, os critérios algorítmicos empregados e o volume de impulsionamentos e anúncios moderados. Dessa maneira, o equilíbrio do debate sobre o *duty of care* repousa na compreensão de que as mídias sociais detêm responsabilidades sociais correlatas ao poder político e econômico que exercem. O dever de cuidado proativo não legitima o monitoramento universal e indiscriminado da intimidade dos cidadãos ou confere às empresas o papel de tribunais particulares da verdade. Trata-se, em última análise, de uma imposição de diligência técnica e corporativa voltada a impedir que a arquitetura das redes continue a ser explorada de forma livre por eixos de radicalização violenta, conciliando a livre circulação de ideias com a preservação da integridade física e moral das instituições e dos cidadãos no mundo real.

#### 4.3 Análise comparada de propostas de regulação setorial e a necessidade do dever de moderação preventiva para mitigar a radicalização juvenil

A insuficiência dos regimes clássicos de irresponsabilidade civil dos provedores de internet diante da proliferação de conteúdos ilícitos e discursos extremistas culminou em uma profunda transição regulatória em escala global. Superando o modelo puramente reativo do *notice-and-takedown*, historicamente consolidado pela Seção 230 do Communications Decency

---

<sup>30</sup> Efeito inibidor, no sentido do resfriamento da liberdade de expressão.

Act<sup>31</sup> norte-americano e pelo artigo 19 do Marco Civil da Internet no Brasil, as principais democracias ocidentais passaram a estruturar marcos legislativos baseados na gestão de riscos estruturais e no dever de cooperação proativa das plataformas digitais.

No cenário europeu, o principal paradigma dessa evolução é o Digital Services Act (DSA), regulamento da União Europeia que entrou em vigor em 2024. O DSA adota uma abordagem assimétrica, impondo obrigações rigorosas e proporcionais ao tamanho e ao impacto das empresas de tecnologia, classificando as maiores corporações como Plataformas Online de Domínio Muito Grande (Very Large Online Platforms ou VLOPs). Nos termos dos artigos 34 e 35 do regulamento europeu, essas plataformas são obrigadas a realizar auditorias e avaliações anuais detalhadas sobre os chamados “riscos sistêmicos” decorrentes do funcionamento e do design de seus serviços. Tais riscos incluem a disseminação de conteúdos ilegais, os efeitos negativos sobre o exercício de direitos fundamentais e a manipulação intencional de sistemas de recomendação que afetem o discurso cívico, a saúde pública e a segurança de menores de idade. A partir dessa identificação, as plataformas devem implementar salvaguardas preventivas obrigatórias, que englobam desde a alteração de algoritmos de curadoria até a adequação de seus termos de uso.

Paralelamente, o Reino Unido editou o Online Safety Act (OSA) em outubro de 2023, estabelecendo um regime de regulação setorial centralizado na figura do órgão fiscalizador Ofcom<sup>32</sup>. Diferenciando-se pelo foco marcadamente voltado à segurança infantojuvenil, o OSA introduziu o conceito legal de duty of care, obrigando as empresas que gerenciam redes sociais e mecanismos de busca a proteger ativamente as crianças contra conteúdos que, embora possam não atingir a tipicidade criminal estrita, sejam severamente prejudiciais ao seu desenvolvimento psíquico, como a apologia à automutilação, o *cyberbullying*<sup>33</sup> e a misoginia violenta. O regulamento britânico exige a adoção de tecnologias de verificação etária robustas e confere ao regulador o poder de impor multas bilionárias e sanções criminais aos executivos das empresas em caso de omissão sistêmica no monitoramento de ameaças.

No contexto brasileiro, a incapacidade do Poder Legislativo em concluir a votação do Projeto de Lei nº 2630/2020, conhecido como PL das Fake News ou Lei Brasileira de Liberdade, Responsabilidade e Transparência na Internet, provocou um deslocamento

---

<sup>31</sup> Lei aprovada em 1996 nos Estados Unidos, famosa por estabelecer as bases jurídicas da internet moderna. Representa a primeira tentativa do congresso norte-americano de regular conteúdo obsceno e indecente online.

<sup>32</sup> *Office of Communications* (Ofcom) é o órgão governamental que atua como autoridade reguladora e de concorrência para as indústrias de radiodifusão, internet, telecomunicações e postais do Reino Unido.

<sup>33</sup> Prática de violência psicológica, intimidação, humilhação ou perseguição cometida contra alguém através da internet ou de plataformas digitais, como redes sociais, aplicativos de mensagens e jogos online.

institucional da regulação para o Judiciário. Assim, o STF, ao fixar as teses vinculantes para os Temas 533 e 987 de Repercussão Geral, promoveu uma releitura constitucional do artigo 19 do Marco Civil da Internet. O tribunal reconheceu que a exigência de prévia ordem judicial para a responsabilização das plataformas por conteúdos de terceiros não pode operar como manto de imunidade absoluta frente a danos evidentes aos direitos humanos e à segurança do Estado. Dessa forma, estabeleceu parâmetros de corresponsabilidade civil solidária das plataformas digitais quando demonstrada a inércia empresarial após a sinalização extrajudicial simplificada de crimes específicos e graves. Configura-se, portanto, um modelo híbrido no Brasil, que constrange as plataformas digitais a abandonarem a passividade e a assumirem um monitoramento proativo de padrões de violência em suas redes. A imposição de um dever de moderação com viés nitidamente preventivo encontra sustentação na análise criminológica do ecossistema digital contemporâneo. O modelo tradicional de remoção de conteúdos baseado unicamente na denúncia individualizada ou na espera por mandados judiciais revela-se estruturalmente ineficiente para interromper as engrenagens de radicalização ideológica que vitimizam jovens e adolescentes. A radicalização não se caracteriza como um ato isolado e repentino, mas sim como um processo de socialização desviante progressivo, cujos vetores operam de forma contínua no ambiente virtual. Sob a perspectiva sociocriminológica, a Teoria da Associação Diferencial de Edwin Sutherland assevera que o comportamento criminoso, as técnicas para a sua execução e as racionalizações e atitudes favoráveis à violação das normas jurídicas são aprendidos por meio de interações sociais no interior de grupos íntimos. No século XXI, o ciberespaço ressignificou a abrangência dessa teoria ao converter-se no principal espaço de socialização juvenil. O diferencial reside no fato de que o contato com os estímulos favoráveis ao desvio não ocorre mais de forma aleatória, mas é ativamente gerenciado por arquiteturas algorítmicas projetadas para prender a atenção do usuário por meio do engajamento emocional baseado em reações extremas, como a indignação e o ressentimento.

Eli Pariser aponta em sua obra que os algoritmos de recomendação geram bolhas informacionais opacas que isolam o indivíduo de perspectivas divergentes, retroalimentando continuamente suas inclinações e vulnerabilidades psicológicas prévias. Nesse sentido, Cass Sunstein adverte que o confinamento de usuários em ambientes virtuais fechados (os chamados “jardins murados” ou “câmaras de eco”) dispara o fenômeno da polarização de grupo, um viés cognitivo pelo qual os membros de uma comunidade que compartilham de uma mesma visão tendem a adotar posturas cada vez mais extremas após debaterem apenas entre si. É no interior dessa dinâmica de isolamento e amplificação algorítmica que se processa a cooptação de jovens

vulneráveis pelas subculturas extremistas da internet comum (a chamada surface web), pulverizadas nas plataformas populares do momento, como o TikTok, o X, o Instagram e os servidores do Discord. Assim, jovens e adolescentes afetados pelo isolamento social, com sentimentos de inadequação, baixa autoestima e depressão tornam-se os alvos preferenciais do recrutamento ideológico promovido pela manosphere, disseminando seus discursos de ódio misóginos e supremacistas. Conforme apontado em relatórios do Center for Countering Digital Hate (CCDH), comunidades integradas na “incelosphere” operam por meio de táticas sofisticadas de manipulação psicológica, acolhendo inicialmente as frustrações dos jovens para, em um segundo momento, direcionar essa angústia contra alvos específicos, elegendo, assim, mulheres, minorias sociais e as próprias instituições escolares como os bodes expiatórios de suas falhas pessoais. O ambiente digital atua, portanto, como um catalisador criminógeno, fornecendo as justificativas morais, o vocabulário de motivos e o suporte comunitário necessários para transmutar o sofrimento individual em violência coletiva letal. A materialização da radicalização online em tragédias tangíveis no mundo real, evidenciada pela escalada de ataques de violência extrema contra instituições de ensino no Brasil, expõe de forma incontestável a necessidade imperiosa de um dever de moderação preventiva. Aguardar a consumação do dano ou a formalização de uma notificação judicial para agir significa tolerar que o ciclo de violência digital atinja o seu ápice operacional no mundo físico. Conforme os estudos e relatórios produzidos pelo Ministério da Educação, aproximadamente 72% dos ataques perpetrados contra escolas brasileiras na última década exibiram vínculos diretos com dinâmicas de radicalização online (BRASIL, 2023). Antes de invadirem as instituições munidos de armas de fogo ou armas brancas, os agressores compartilharam manifestos em perfis públicos, consumiram conteúdos que glorificavam autores de massacres históricos, trocaram instruções táticas sobre a fabricação de artefatos explosivos e receberam incentivos diretos de outros integrantes dessas redes de ódio. Os espaços virtuais assumiram papel de vetores de contágio social, onde a espetacularização da violência e a promessa de conquista de notoriedade póstuma nas redes atuaram como fatores determinantes para a eclosão de novos atentados.

Diante de um quadro em que a proliferação do discurso de ódio antecede causalmente a eliminação de vidas humanas, a imposição do dever de moderação preventiva às plataformas digitais deixa de ser uma mera opção de política pública e assume o status de imperativo constitucional de proteção à vida e à dignidade humana. Esse dever traduz-se na incorporação

obrigatória do princípio do *Safety by Design*<sup>34</sup>, segundo o qual a proteção dos usuários, em especial das crianças e adolescentes, deve integrar a própria infraestrutura técnica e lógica da plataforma, sobrepondo-se às métricas comerciais de engajamento e lucratividade. Praticar a moderação preventiva não implica a instituição de mecanismos de censura prévia ao debate público legítimo, mas sim a exigência de que as grandes corporações tecnológicas adotem medidas sistêmicas de mitigação de riscos claramente identificáveis, tais como: a desotimização algorítmica de conteúdos extremistas (impedir que os sistemas de recomendação direcionem automaticamente hashtags, vídeos ou contas associadas a discursos misóginos, estéticas neonazistas ou apologia a massacres para usuários que demonstraram vulnerabilidade psicológica ou interesse inicial por termos correlatos), a auditoria de espaços privados de cooptação (desenvolver ferramentas capazes de identificar padrões linguísticos e comportamentais típicos de planejamento de ataques em canais restritos, como servidores do Discord ou grupos de Telegram de acesso restrito, notificando imediatamente as autoridades de segurança pública e de inteligência) e mecanismos efetivos de controle de idade e denúncia (implementar sistemas rígidos de verificação etária para o ingresso em fóruns de discussão com temáticas sensíveis e disponibilizar canais acessíveis e céleres para que a sociedade civil, familiares e instituições de ensino possam reportar perfis em processo de radicalização aguda, garantindo uma resposta em tempo hábil para evitar o desencadeamento de ações físicas violentas). Em suma, a transição para um regime de responsabilidade proativa das plataformas digitais reconhece que as empresas de tecnologia não são meros repositórios neutros de dados, mas agentes que estruturam e lucram com a dinâmica do espaço público digital. Mitigar a radicalização juvenil e salvaguardar a integridade das escolas exige que o ordenamento jurídico imponha às plataformas o ônus de gerenciar os riscos decorrentes de seus modelos de negócios, convertendo a moderação preventiva em uma barreira de contenção indispensável contra o avanço da violência extremista.

## **5. CONSIDERAÇÕES FINAIS**

A presente pesquisa permitiu analisar, sob uma perspectiva interdisciplinar que conjuga a Criminologia Crítica, o Direito Constitucional e o Direito Civil, o papel ativo do ciberespaço

---

<sup>34</sup> Abordagem proativa que prioriza a segurança e a mitigação de riscos desde a fase de concepção de um projeto. Em vez de aplicar correções ou medidas de segurança após a ocorrência de problemas, o objetivo é antecipar falhas e eliminar perigos na origem, exigindo, assim, que os provedores implementem ferramentas preventivas para garantir um ambiente saudável em suas plataformas.

como um vetor criminógeno contemporâneo. Distanciando-se da premissa obsoleta de que a internet figura como mero suporte neutro para a interação humana, restou demonstrado que a arquitetura das plataformas digitais e seus modelos de negócios orientados ao engajamento exercem influência direta na proliferação do extremismo e na radicalização juvenil.

Em primeiro lugar, a transposição da Teoria da Associação Diferencial de Edwin Sutherland para o ecossistema digital mostrou-se plenamente fidedigna. Os espaços virtuais marginalizados, como os fóruns anônimos (chans), subcomunidades da “manosfera” e redes de ódio estruturadas, assumiram o papel dos tradicionais “grupos pessoais íntimos” descritos pela criminologia clássica. É nesses ambientes que indivíduos vulneráveis assimilam não apenas as técnicas para a execução de condutas criminosas, mas, fundamentalmente, as definições normativas, os impulsos e as técnicas de neutralização psicológica que justificam a agressão ao outro, sendo essa dinâmica de retroalimentação maximizada pela engenharia algorítmica de recomendação. Conforme mapeado ao longo do trabalho, o confinamento dos usuários em “filtros invisíveis”, na conceituação de Eli Pariser, e o seu isolamento em “câmaras de eco”, conforme a lição de Cass Sunstein, anulam o dissenso democrático e constroem realidades paralelas hiperpolarizadas. O resultado prático desse isolamento é o fenômeno da polarização de grupo, que impulsiona indivíduos previamente intolerantes em direção ao extremismo radical violento. A conversão desse ódio virtual em violência concreta restou empiricamente evidenciada pelo alarmante incremento epidemiológico dos ataques violentos a instituições de ensino no Brasil nos últimos anos, cujos perpetradores operavam imersos nessas dinâmicas subculturais online.

No âmbito do Direito Constitucional, a investigação confirmou a plena legitimidade e a imperatividade do combate ao discurso de ódio nas redes. A ponderação de princípios revela que a liberdade de expressão não se reveste de caráter absoluto no ordenamento jurídico pátrio, encontrando limites intransponíveis nos preceitos fundamentais da dignidade da pessoa humana, da igualdade e da vedação ao racismo e à discriminação. O referencial hermenêutico histórico fixado pelo Supremo Tribunal Federal no julgamento do Caso Ellwanger permanece como vetor central: manifestações que visem à negação da condição humana ou à incitação à violência contra grupos minoritários ou vulneráveis não encontram agasalho na salvaguarda constitucional da livre manifestação do pensamento.

Por fim, o núcleo crítico da pesquisa identificou o esgotamento do modelo regulatório instituído pelo artigo 19 do Marco Civil da Internet (Lei nº 12.965/14). A exigência de uma ordem judicial específica como condição prévia para a responsabilização civil das plataformas

por conteúdos de terceiros (o denominado regime notice-and-takedown judicializado) revelou-se anacrônica, ineficaz e incapaz de responder à velocidade exponencial dos fluxos informacionais algorítmicos. A morosidade inerente ao trâmite judicial atua como um fator de facilitação, permitindo que conteúdos radicalizadores e discursos de ódio permaneçam acessíveis por tempo suficiente para contaminar novos jovens e consolidar infrações penais. Torna-se premente, portanto, a evolução legislativa do ordenamento jurídico brasileiro rumo à superação da mera passividade das plataformas digitais. À luz do direito comparado, com especial destaque para o pioneirismo do Regulamento de Serviços Digitais (Digital Services Act - DSA) da União Europeia e do Online Safety Act de 2023 no Reino Unido, conclui-se pela necessidade de positivação de um “dever de cuidado” (duty of care) proativo. Impõe-se aos provedores de redes sociais a obrigação jurídica de implementar mecanismos transparentes de análise, mitigação de riscos sistêmicos e auditoria de seus próprios algoritmos, sob pena de corresponsabilização civil objetiva pelos danos decorrentes da omissão sistêmica frente às redes de ódio que alimentam a delinquência juvenil no século XXI.

## 6. REFERÊNCIAS

ALCADIPANI, Rafael; SILVA, Diogo. Radicalização Online e a Teoria da Associação Diferencial: Um Estudo sobre Grupos de Extrema-Direita. *Revista Brasileira de Ciências Criminais*, São Paulo, v. 28, n. 165, p. 112-135, out. 2022.

BANDURA, Albert. *Social Learning Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1977.

BRASIL. [Constituição (1988)]. Constituição da República Federativa do Brasil de 1988. Brasília, DF: Presidência da República. Disponível em: [http://www.planalto.gov.br/ccivil\\_03/constituicao/constituicao.htm](http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm). Acesso em: 26 maio 2026.

BRASIL. Decreto-Lei nº 2.848, de 7 de dezembro de 1940. Código Penal. Diário Oficial da União: Rio de Janeiro, 1940. Disponível em: [https://www.planalto.gov.br/ccivil\\_03/decretolei/del2848compilado.htm](https://www.planalto.gov.br/ccivil_03/decretolei/del2848compilado.htm). Acesso em 26 maio 2026.

BRASIL. Lei nº 8.072, de 25 de julho de 1990. Dispõe sobre os crimes hediondos, nos termos do art. 5º, inciso XLIII, da Constituição Federal, e determina outras providências. Diário Oficial da União: Brasília, 1990. Disponível em: [https://www.planalto.gov.br/ccivil\\_03/leis/18072.htm](https://www.planalto.gov.br/ccivil_03/leis/18072.htm). Acesso em 26 maio 2026.

BRASIL. Lei nº 12.965, de 23 de abril de 2014. Institui princípios, garantias, direitos e deveres para o uso da Internet no Brasil (Marco Civil da Internet). Diário Oficial da União: Brasília,

2014. Disponível em: [http://www.planalto.gov.br/ccivil\\_03/\\_ato2011-2014/2014/lei/112965.htm](http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2014/lei/112965.htm). Acesso em: 26 maio 2026.

BRASIL. Lei nº 14.532, de 11 de janeiro de 2023. Altera a Lei nº 7.716, de 5 de janeiro de 1989 (Lei do Crime Racial), e o Decreto-Lei nº 2.848, de 7 de dezembro de 1940 (Código Penal), para tipificar como crime de racismo a injúria racial, prever pena de suspensão de direito em caso de racismo praticado no contexto de atividade esportiva ou artística e prever pena para o racismo religioso e recreativo e para o praticado por funcionário público. Diário Oficial da União: Brasília, 2023. Disponível em: [https://www.planalto.gov.br/ccivil\\_03/\\_ato2023-2026/2023/lei/114532.htm#art1](https://www.planalto.gov.br/ccivil_03/_ato2023-2026/2023/lei/114532.htm#art1). Acesso em 24 maio 2026.

BRASIL. Lei nº 15.159, de 2025. Altera o Código Penal e a Lei de Crimes Hediondos para dispor sobre o aumento de penas para crimes cometidos em estabelecimentos de ensino. Diário Oficial da União: Brasília, DF, 2025.

BRASIL. Ministério da Educação. Relatório do Grupo de Trabalho Especial: Combate à Violência nas Escolas. Brasília, 2023. Disponível em: [www.gov.br/mec/pt-br/aceso-ainformacao/participacao-social/grupos-de-trabalho/prevencao-e-enfrentamento-da-violencianas-escolas/resultados/relatorio-ataque-escolas-brasil.pdf](http://www.gov.br/mec/pt-br/aceso-ainformacao/participacao-social/grupos-de-trabalho/prevencao-e-enfrentamento-da-violencianas-escolas/resultados/relatorio-ataque-escolas-brasil.pdf). Acesso em: 17 abr. 2026.

CENTER FOR COUNTERING DIGITAL HATE (CCDH). The incelosphere: exposing pathways into incel communities and the harms they pose to women and children. Washington, DC: CCDH, 2022.

FARIA, Igor Carvalho Ulhôa. Notice and take down: ponderações sobre o sistema. *Civilistica.com*, Rio de Janeiro, a. 10, n. 1, 2021.

FARIAS, Laís. A radicalização de comunidades online a partir de ideologias de ultradireita: A mansfêra e a cultura incel. *Revista de Iniciação Científica em Relações Internacionais*, [S. l.], v. 11, n. 21, p. 1–14, 2023.

GARCIA, Cléo; VINHA, Telma. Ataques de violência extrema em escolas no Brasil: Causas e caminhos para a prevenção. *Revista de Educação PUC-Campinas*. Campinas, v.30, 2025.

HOLT, Thomas J.; BOSSALER, Adam M.; SEIGFRIED-SPELAR, Kathryn C. *Cybercrime and Digital Forensics: An Introduction*. 3. ed. London: Routledge, 2022. 540 p.

MACHADO, Michele. A Mansfêra e a Radicalização de Jovens: Uma Análise Criminológica do Discurso de Ódio contra Mulheres em Comunidades Virtuais. In: CONGRESSO BRASILEIRO DE CRIMINOLOGIA, 12., 2023, Porto Alegre. Anais... Porto Alegre: EDIPUCRS, 2023. p. 45-60.

SANSEVERINO, Bruna. Discurso de ódio criminalizado e lesividade do bem jurídico. Tese (Doutorado em Direito Penal), Universidade de São Paulo, São Paulo, 2023.

PARISER, Eli. *O filtro invisível: o que a internet está escondendo de você*. Tradução de Diego Alfaro. 1. ed. Rio de Janeiro: Zahar, 2012. 272 p.

REINO UNIDO. Online Safety Act 2023. Elizabeth II. Chapter 50. London: The Stationery Office (TSO), 2023. Disponível em: <https://www.legislation.gov.uk/ukpga/2023/50/contents/enacted>. Acesso em: 26 maio 2026.

SUPERIOR TRIBUNAL DE JUSTIÇA (STJ). Sexta Turma. Recurso em Habeas Corpus / Apelação Criminal (REsp Nº 2015530 - RS (2022/0226582-4). Relator: Ministro Antonio Saldanha Palheiro. Julgamento em: 16/12/2025.

SUPREMO TRIBUNAL FEDERAL (STF). Ação Direta de Inconstitucionalidade por Omissão (ADO) nº 26/DF. Relator: Min. Celso de Mello. Julgamento: 13/06/2019. Órgão Julgador: Tribunal Pleno. Informativo STF nº 944.

SUPREMO TRIBUNAL FEDERAL (STF). Habeas Corpus nº 82.424/RS. Relator para o Acórdão: Min. Maurício Corrêa. Julgamento: 17/09/2003.

SUPREMO TRIBUNAL FEDERAL (STF). Tribunal Pleno. Recurso Extraordinário nº 1.037.396/SP (Tema 987 da Repercussão Geral). Discussão sobre a constitucionalidade do art. 19 da Lei n. 12.965/2014 (Marco Civil da Internet). Relator: Min. Dias Toffoli. Julgamento: 15/05/2023.

SUPREMO TRIBUNAL FEDERAL (STF). Tribunal Pleno. Recurso Extraordinário nº 1.057.258/SP (Tema 533 da Repercussão Geral). Dever de empresa hospedeira de sítio na internet de fiscalizar o conteúdo publicado e de retirá-lo do ar quando considerado ofensivo. Relator: Min. Luiz Fux. Julgamento: 26/06/2025.

SUNSTEIN, Cass. Republic.com. Princeton: Princeton University Press, 2001. 236 p.

SUNSTEIN, Cass. #Republic: divided democracy in the age of social media. Princeton: Princeton University Press, 2017. 328 p.

SUTHERLAND, Edwin. Principles of criminology. 3. ed. Philadelphia: J. B. Lippincott, 1939.

STHEL, Fernanda Gonçalves. Do virtual ao real: as interações dos grupos de radicalização online com os ataques em escolas no Brasil. Dissertação (Mestrado em Sociologia Política), Universidade Estadual do Norte Fluminense Darcy Ribeiro (UENF), Campos dos Goytacazes, 2025.

SYKES, Gresham; MATZA, David. Techniques of neutralization: a theory of delinquency. American Sociological Review, v. 22, n. 6, p. 664-670, dez. 1957.

UNIÃO EUROPEIA. Regulamento (UE) 2022/2065 do Parlamento Europeu e do Conselho de 19 de outubro de 2022 relativo a um mercado único digital para os serviços digitais (Regulamento dos Serviços Digitais - DSA). Diário Oficial da União Europeia, Bruxelas, L 277, 27 out. 2022. Disponível em: <https://eur-lex.europa.eu/legalcontent/PT/TXT/?uri=celex%3A32022R2065>. Acesso em: 26 maio 2026.

VICENTE, João Paulo. Algoritmos e Radicalização: A Associação Diferencial no Século XXI. *Jornal de Sociologia Digital*. Coimbra, v. 4, n. 2, p. 88-105, jan./jun. 2024.