

DISCURSOS DE ÓDIO NEONAZISTAS EM PLATAFORMAS DIGITAIS EM (DES) FAVOR DE ADOLESCENTES: DESAFIOS E IMPLICAÇÕES NO ÂMBITO DAS CIÊNCIAS CRIMINAIS

NEO-NAZI HATE SPEECH ON DIGITAL PLATFORMS (DIS)FAVORING ADOLESCENTS: CHALLENGES AND IMPLICATIONS IN THE FIELD OF CRIMINAL SCIENCES

DISCURSOS DE ODIO NEONAZIS EN PLATAFORMAS DIGITALES EN (DES)FAVOR DE LOS ADOLESCENTES: DESAFÍOS E IMPLICACIONES EN EL ÁMBITO DE LAS CIENCIAS PENALES

Isabela Quartieri da Rosa¹  
Universidade em Santa Maria, Rio Grande do Sul, Brasil 

Ana Carolina Sassi²  
Universidade em Santa Maria, Rio Grande do Sul, Brasil 

Luis Gustavo Durigon³  
Universidade em Santa Maria, Rio Grande do Sul, Brasil 

Recebido em: 2025-04-22
Aceito em: 2025-08-08

Autor correspondente: Luis Gustavo Durigon *E-mail:* durigon.lgustavo@ufrsm.br

SUMÁRIO: 1 *Introdução;* 2 *Os algoritmos e a inteligência artificial na moderação de conteúdos odiosos de cunho neonazista e a proteção de adolescentes em plataformas digitais;* 3 *Potencialidades e limites do direito penal e constitucional na regulação dos discursos de ódio contra adolescentes no ambiente digital em um contexto democrático globalizado;* 4 *Considerações Finais; Referências.*

CONTEXTUALIZAÇÃO: A disseminação de discursos de ódio neonazistas nas redes sociais representa uma grave ameaça aos adolescentes, que se mostram particularmente vulneráveis a tais conteúdos. Diante deste cenário, a pesquisa investiga como as plataformas digitais podem ser reguladas para garantir a proteção de adolescentes contra discursos de ódio neonazistas, considerando os desafios do direito penal e constitucional no contexto das tecnologias emergentes.

¹ Doutoranda e Mestre em Direito pela Universidade Federal de Santa Maria (UFSM). Especialista em Direito Público pela Universidade de Caxias do Sul (UCS) e Bacharel em Direito pela Universidade Franciscana (UFN). Pesquisadora do Núcleo de Direito Informacional (NUDI/UFSM).

² Bolsista CAPES. Doutoranda e Mestre em Direito pela Universidade Federal de Santa Maria (UFSM). Especialista em Direito do Trabalho e Previdenciário pela UniAméricas. Bacharel em Direito pela Universidade Franciscana (UFN). Pesquisadora do Núcleo de Pesquisa em Direito Informacional (NUDI/UFSM).

³ Bolsista CAPES. Doutoranda e Mestre em Direito pela Universidade Federal de Santa Maria (UFSM). Especialista em Direito do Trabalho e Previdenciário pela UniAméricas. Bacharel em Direito pela Universidade Franciscana (UFN). Pesquisadora do Núcleo de Pesquisa em Direito Informacional (NUDI/UFSM).

OBJETIVOS: O estudo tem como objetivo analisar a eficácia e as limitações das medidas tecnológicas e jurídicas atuais, promovendo uma reflexão crítica sobre o papel do Estado e das empresas nesses (in)fluxos políticos.

METODOLOGIA: A pesquisa é conduzida através de uma revisão bibliográfica qualitativa e uma análise teórica crítica das fontes. A abordagem será transdisciplinar, integrando perspectivas da criminologia crítica, direito constitucional e sociologia, com ênfase nos dois primeiros, para construir uma compreensão abrangente do tema sem nenhum cunho totalitarista, eis que calcada na epistemologia da incerteza, marca das investigações científicas contemporâneas.

RESULTADOS: A pesquisa revela que a regulamentação efetiva e ética no ambiente digital demanda um esforço contínuo de adaptação das normas jurídicas, aprimoramento das tecnologias de moderação e desenvolvimento de acordos internacionais. É fundamental garantir a liberdade de expressão, mas também proteger os adolescentes dos danos causados pelos discursos de ódio, especialmente em uma era marcada pela vigilância e pelos algoritmos.

PALAVRAS-CHAVE: Adolescentes; Direito penal; Discursos de ódio; Plataformas digitais; Vigilância.

CONTEXTUALIZATION: The spread of neo-Nazi hate speech on social networks poses a serious threat to adolescents, who are particularly vulnerable to such content. In this scenario, the research investigates how digital platforms can be regulated to ensure the protection of adolescents against neo-Nazi hate speech, considering the challenges of criminal and constitutional law in the context of emerging technologies.

OBJECTIVES: The study aims to analyze the effectiveness and limitations of current technological and legal measures, fostering critical reflection on the role of the State and companies in these political (in)flows.

METHODOLOGY: The research is conducted through a qualitative literature review and a critical theoretical analysis of the sources. The approach is transdisciplinary, integrating perspectives from critical criminology, constitutional law, and sociology, emphasizing the first two, in order to build a comprehensive understanding of the topic, without any totalitarian bias, as it is grounded in the epistemology of uncertainty, a hallmark of contemporary scientific inquiry.

RESULTS: The research reveals that effective and ethical regulation in the digital environment requires ongoing efforts to adapt legal norms, improve moderation technologies, and develop international agreements. It is essential to guarantee freedom of expression while also protecting adolescents from the harms caused by hate speech, especially in an era marked by surveillance and algorithms.

KEYWORDS: Adolescentes; Criminal law; Digital platforms; Hate speech; Surveillance.

CONTEXTUALIZACIÓN: La difusión del discurso de odio neonazi en las redes sociales representa una grave amenaza para los adolescentes, quienes se muestran particularmente vulnerables a este tipo de contenidos. En este escenario, la investigación analiza cómo se pueden regular las plataformas digitales para garantizar la protección de los adolescentes frente al discurso de odio neonazi, considerando los desafíos del derecho penal y constitucional en el contexto de las tecnologías emergentes.

OBJETIVOS: El estudio tiene como objetivo analizar la eficacia y las limitaciones de las medidas tecnológicas y jurídicas actuales, fomentando una reflexión crítica sobre el papel del Estado y de las empresas en estos (in)fluxos políticos.

METODOLOGÍA: La investigación se desarrolla mediante una revisión bibliográfica cualitativa y un análisis teórico-crítico de las fuentes. El enfoque es transdisciplinario, integrando perspectivas de la criminología crítica, el derecho constitucional y la sociología, con énfasis en las dos primeras, para construir una comprensión amplia del tema, sin ningún sesgo totalitario, ya que se fundamenta en la epistemología de la incertidumbre, característica de la investigación científica contemporánea.

RESULTADOS: La investigación revela que una regulación efectiva y ética en el entorno digital exige un esfuerzo continuo de adaptación de las normas jurídicas, el perfeccionamiento de las tecnologías de moderación y el desarrollo de acuerdos internacionales. Es fundamental garantizar la libertad de expresión, pero también proteger a los adolescentes de los daños causados por el discurso de odio, especialmente en una era marcada por la vigilancia y los algoritmos.

PALABRAS CLAVE: Adolescentes; Discursos de odio; Derecho penal; Plataformas digitales; Vigilancia.

INTRODUÇÃO

A ascensão das tecnologias digitais e o crescente protagonismo das redes sociais na vida cotidiana têm transformado profundamente a forma como os indivíduos interagem e se comunicam, desafiando e tensionando o campo do direito, merecendo novos olhares das ciências criminais. Esse fenômeno de digitalização intensiva trouxe, no entanto, uma série de desafios para a proteção de direitos fundamentais, especialmente para grupos vulneráveis, como os adolescentes, que passam a ter sua integridade e segurança expostas a novas ameaças no ambiente digital⁴. Entre essas ameaças, destacam-se os discursos de ódio com motivações neonazistas, que promovem ideologias discriminatórias, disseminam preconceitos e, muitas vezes, incitam à violência, marca das sociedades contemporâneas. A vulnerabilidade dos adolescentes, somada à dificuldade de controle dos conteúdos que circulam nas redes, exige que o Direito Penal e Constitucional enfrente a questão de forma inovadora e adaptada às especificidades deste ambiente, em que a ação humana se confunde com a operação de algoritmos e inteligência artificial (IA).

No Brasil, a problemática de controle de conteúdo odiosos nas plataformas digitais é acentuada por um contexto legislativo e judicial que ainda carece de regulamentações específicas para lidar com a moderação de conteúdos digitais e a proteção de adolescentes contra essas formas de violência. Um exemplo disso foi a suspensão da plataforma X⁵ (antigo Twitter), decretada em razão da disseminação de discursos de ódio e da obstrução de investigações criminais. Da mesma forma, ocorreu a suspensão da plataforma Rumble Inc⁶, devido à massiva divulgação de discursos nazistas, racistas, fascistas, de ódio e antidemocráticos.

270 A utilização de ferramentas de inteligência artificial e algoritmos por essas plataformas para a moderação de conteúdos oferece um caminho promissor, paradoxalmente não seguro, pois continua a desafiar os limites das garantias constitucionais e da liberdade de expressão. Assim, considerando o avanço das tecnologias e seu impacto sobre os sistemas de justiça penal e proteção de direitos fundamentais, levanta-se o problema de pesquisa: como as plataformas digitais, utilizando tecnologias de vigilância e inteligência artificial, podem a partir da regulação garantir a proteção de adolescentes contra discursos de ódio neonazistas, considerando os desafios do direito penal e constitucional no contexto das tecnologias emergentes?

Diante desse cenário, este estudo tem como objetivo geral analisar a eficácia e as limitações das medidas tecnológicas e jurídicas atuais na proteção dos adolescentes contra a disseminação de discursos de

⁴ BARROSO, Luís Roberto. Revolução tecnológica, crise da democracia e constituição: Direito e políticas públicas num mundo em transformação. Belo Horizonte: Fórum, 2020.

⁵ A decisão trata da suspensão do funcionamento da empresa X Brasil Internet Ltda. no território nacional, em razão do descumprimento de ordens judiciais relacionadas à disseminação de discursos de ódio e desinformação, além da não nomeação de representante legal da empresa no Brasil. A medida inclui a imposição de multa diária de R\$ 50.000,00 para aqueles que fraudarem a decisão judicial utilizando subterfúgios tecnológicos, como VPNs. A decisão foi referendada pela Turma do Supremo Tribunal Federal (STF), com ressalvas do Ministro Luiz Fux em relação ao alcance da multa para pessoas não envolvidas diretamente no processo.

⁶ Trata-se de uma plataforma de compartilhamento de vídeo, fundada em 2013 por Chris Pavlovski, um empresário canadense. Seu funcionamento é semelhante ao Youtube, porém com uma moderação mais permissiva que retém um foco apenas em conteúdos considerados ilegais. A decisão do STF em relação à empresa Rumble Inc. destaca a utilização de redes sociais por grupos extremistas para disseminação de discursos de ódio e mensagens antidemocráticas. A Constituição Brasileira de 1988 proíbe a confusão entre liberdade de expressão e discurso de ódio, e exige o cumprimento das ordens judiciais por empresas que operem no Brasil, incluindo a nomeação de representantes legais. A Rumble Inc. foi suspensa no país por desobedecer reiteradamente ordens judiciais, não nomear representante e continuar veiculando conteúdos prejudiciais. A decisão reafirma a obrigação de respeito ao ordenamento jurídico brasileiro e a proteção dos pilares democráticos.

ódio neonazistas nas plataformas digitais, promovendo uma reflexão crítica sobre o papel do direito penal e constitucional em um contexto de tecnologia avançada e globalizada. Especificamente, examinando o papel e os limites dos algoritmos e das ferramentas de inteligência artificial na identificação e mitigação de conteúdos neonazistas e de discursos de ódio direcionados aos adolescentes em redes sociais e plataformas digitais; bem como avaliando as possibilidades e desafios do Direito Penal e Constitucional no controle e na regulação da disseminação de discursos de ódio contra adolescentes no ambiente digital, especialmente em um cenário de Democracia e Estado de Direito globalizado, ainda que uma de suas marcas na contemporaneidade seja o tensionamento de suas matrizes constitutivas.

Assim, a pesquisa fundamenta-se em uma base teórica que une contribuições de diversas áreas do conhecimento jurídico e tecnológico, especialmente nos estudos sobre algoritmos, vigilância e tecnologia no Direito Penal. Obras como *Algoritarismos* de Amaral, Sabariego e Salles⁷ fornecem o arcabouço teórico sobre a aplicação de algoritmos no sistema de justiça digital, enquanto *Vigilância Líquida* de Bauman e Lyon⁸ oferece um olhar crítico sobre a sociedade de vigilância e o controle digital. Além disso, a obra de Silva⁹ sobre *Racismo Algorítmico* e de Mendes e Freitas¹⁰ sobre as relações entre constituição, Direito Penal e novas tecnologias ajudam a entender os riscos de discriminação e o impacto das novas tecnologias no âmbito penal e constitucional.

Metodologicamente, a pesquisa é conduzida através de uma revisão bibliográfica qualitativa e uma análise teórica crítica das fontes. A abordagem será transdisciplinar, integrando perspectivas da criminologia crítica, direito constitucional e sociologia, com ênfase nos dois primeiros, para construir uma compreensão abrangente do tema sem nenhum cunho totalitarista, eis que calcada na epistemologia da incerteza, marca das investigações científicas contemporâneas. Através da análise dos fundamentos jurídicos e das tecnologias de moderação de conteúdo digital, este estudo busca expor os desafios e as limitações das políticas de moderação de discursos de ódio em plataformas digitais e seu impacto sobre adolescentes.

A relevância do estudo reside na urgente necessidade de adaptar os direitos fundamentais e as normas penais para proteger os adolescentes diante do avanço da tecnologia e da disseminação de conteúdos nocivos nas redes sociais. A integração dos algoritmos e da IA no combate aos discursos de ódio neonazistas se apresenta como uma alternativa importante, mas deve ser acompanhada de um esforço crítico e jurídico para equilibrar o uso dessas tecnologias com a proteção dos direitos e garantias constitucionais. Nesse sentido, esta pesquisa visa contribuir para o desenvolvimento do debate sobre diretrizes normativas que promovam um ambiente digital mais seguro e justo para os adolescentes, em consonância com o projeto constitucional e democrático do Estado de Direito no Brasil, cuja concretização permanece inacabada.

⁷ AMARAL, Augusto Jobim do; SABARIEGO, Jesús; SALLES, Eduardo Baldissera Carvalho. *Algoritarismos*. São Paulo-Valencia, Tirant lo Blanch, 2020.

⁸ BAUMAN, Zygmunt. LYON, David. *Vigilância Líquida*. Ciudad Autónoma de Buenos Aires: Paidós, 2013.

⁹ SILVA, Tarcízio. *Racismo algorítmico: inteligência artificial e descriminalização das redes sociais*. São Paulo: Sesc, 2022.

¹⁰ MENDES, Gilmar Ferreira; FREITAS, Matheus Pimenta. *Constituição, Direito Penal e novas tecnologias*. São Paulo: Almedina, 2023.

2 OS ALGORITMOS E A INTELIGÊNCIA ARTIFICIAL NA MODERAÇÃO DE CONTEÚDOS ODIENTOS DE CUNHO NEONAZISTA E A PROTEÇÃO DE ADOLESCENTES EM PLATAFORMAS DIGITAIS

As plataformas digitais¹¹, como redes sociais e fóruns de discussão, tornaram-se espaços essenciais de socialização e expressão, especialmente para adolescentes. No entanto, essas mesmas plataformas também se transformaram em ambientes onde discursos de ódio¹² e conteúdos extremistas, incluindo ideologias neonazistas, podem circular livremente. Nesse cenário, algoritmos e ferramentas de inteligência artificial (IA) são amplamente adotados como uma forma de identificar e mitigar conteúdos potencialmente perigosos, quando, paradoxalmente, não os propagam, sendo que é justamente neste ponto que se encontra o nó da questão investigada neste artigo. Apesar disso, a aplicação dessas tecnologias para moderar discursos de ódio direcionados aos adolescentes enfrentam uma série de desafios técnicos, éticos, com a possibilidade concreta do surgimento de agentes morais/algoritmos artificiais completos¹³ e jurídicos, exigindo uma análise crítica de suas limitações e potenciais impactos.

O Regulamento (UE, 2022) 2022/2065, conhecido como Regulamento sobre os Serviços Digitais (DSA), aborda de forma abrangente a questão ética das plataformas digitais, estabelecendo um conjunto de normas que visam garantir um ambiente online mais seguro e responsável¹⁴. Com o objetivo de proteger os direitos fundamentais dos usuários, exige que as plataformas digitais sejam mais transparentes em relação às práticas de moderação de conteúdo, como a remoção de discursos de ódio, desinformação e outros conteúdos prejudiciais. As plataformas também devem ser responsabilizadas pela proteção de grupos vulneráveis, como menores de idade, e pela implementação de mecanismos que assegurem a segurança e a privacidade dos dados pessoais dos usuários. Além disso, reforça a necessidade de transparência dos algoritmos usados pelas plataformas para a personalização de conteúdo, permitindo que os usuários saibam como suas informações são manipuladas. Em essência, o DSA representa um esforço para equilibrar a liberdade de expressão com a proteção contra abusos online, promovendo um uso mais ético e responsável das plataformas digitais, ao mesmo tempo em que fortalece a responsabilidade das empresas que operam nesse setor.

A atratividade da Inteligência Artificial reside na capacidade de aprender com grandes volumes de dados e gerar padrões de resposta de forma autônoma. Essa característica, baseada no aprendizado de máquina, torna a IA uma ferramenta poderosa para a moderação de conteúdo digital. Os algoritmos e sistemas de IA empregados nessa tarefa, em sua maioria, se fundamentam em técnicas de aprendizado de

¹¹ No contexto brasileiro, dados da pesquisa TIC Kids Online Brasil (CGI.br) revelam que 89,5% dos adolescentes entre 13 e 18 anos utilizam redes sociais, destacando-se plataformas como Instagram, TikTok, YouTube, Discord, X (antigo Twitter) e Facebook. Tais números evidenciam não apenas a expressiva inserção de adolescentes no ambiente digital platoformizado, mas também a urgência de refletir sobre as responsabilidades dessas plataformas na promoção da proteção integral desse grupo etário, em consonância com seu estágio peculiar de desenvolvimento e com os princípios previstos no Estatuto da Criança e do Adolescente.

¹² O discurso de ódio, conforme abordado no Relatório de Recomendações para o Enfrentamento do Discurso de Ódio e o Extremismo no Brasil (MDHC), refere-se a manifestações que incitam, promovem ou legitimam a discriminação, hostilidade ou violência contra indivíduos ou grupos com base em características identitárias como raça, etnia, gênero, orientação sexual, religião, nacionalidade, entre outras. Tais discursos vão além da mera opinião, pois representam uma ameaça concreta à dignidade humana e à convivência democrática, sendo frequentemente instrumentalizados para fins de radicalização, exclusão social e violação de direitos fundamentais.

¹³ SILVEIRA, Paulo Antônio Caliendo Veloso da. Ética e inteligência artificial. Da possibilidade filosófica de agentes morais artificiais. Porto Alegre: Editora Fi, 2021, p.161.

¹⁴ UNIÃO EUROPEIA. Regulamento (UE) 2022/2065. Parlamento Europeu e do Conselho de 19 de outubro de 2022 relativo aos serviços digitais e que altera a Diretiva 2000/31/CE. Jornal Oficial da União Europeia, 2022. Disponível em: <https://eur-lex.europa.eu>. Acesso em: 14 abr. 2025.

máquina e redes neurais, permitindo a identificação e remoção de conteúdos nocivos de forma mais eficiente e escalável.

Essas ferramentas processam grandes volumes de dados para detectar padrões que possam indicar violações das políticas de uso das plataformas. Como observam Amaral, Sabariego e Salles¹⁵, esses algoritmos desempenham um papel essencial ao analisar tanto texto quanto imagens e vídeos, permitindo que as plataformas monitorem interações e conteúdo em uma escala que seria impraticável para uma equipe humana de moderadores. No entanto, a eficácia desses sistemas é limitada por sua dependência de dados de treinamento e pela dificuldade de captar o contexto dos conteúdos.

Em discursos neonazistas, que muitas vezes utilizam códigos linguísticos, símbolos e memes para escapar da detecção, os algoritmos enfrentam dificuldades específicas. Como Silva¹⁶ aponta, a linguagem desses discursos frequentemente envolve elementos culturalmente específicos e nuances que podem passar despercebidas por sistemas automatizados, criando uma lacuna entre o que o algoritmo detecta e o que efetivamente representa um discurso de ódio. Essa limitação faz com que os sistemas automatizados tenham dificuldades para distinguir entre conteúdo realmente prejudicial e discussões legítimas ou irônicas, levando ao fenômeno dos “falsos negativos” (conteúdos nocivos que não são detectados) e “falsos positivos” (conteúdos não nocivos que são removidos).

Além disso, ao observar os impactos das tecnologias na sociedade é preciso discutir também os reflexos deste momento de hiperconexão vividos pela sociedade. Sobre isso, torna imperativo compreender que ao passo que se está cada vez mais conectado, também se está cada vez mais sob vigilância.

A sociedade contemporânea, estimulada pela dominação das *big techs*, adota um modelo econômico que se baseia na coleta, análise e comercialização de dados pessoais. Nesse sistema, as empresas de tecnologia¹⁷ utilizam algoritmos sofisticados para monitorar o comportamento dos indivíduos *online*, construindo perfis detalhados que são utilizados para direcionar publicidade, influenciar decisões e moldar comportamentos.

Tal matrimônio - entre o capitalismo de vigilância e as plataformas digitais - torna possível aquilo que Didier Bigo¹⁸ denomina de banóptico. O autor, em interpretação contemporânea e ampliada do conceito de panóptico, cunhado por Jeremy Bentham¹⁹, versa que o banóptico transcende a arquitetura prisional original e se torna um sistema de vigilância mais difuso e abrangente, especialmente no contexto da globalização e da segurança transnacional.

O discurso não é mais sobre a reforma do indivíduo, mas sobre o controle de grupos específicos que são vistos como uma ameaça ao solo europeu. Para lidar com essas classificações *ban*, as

¹⁵ AMARAL, Augusto Jobim do; SABARIEGO, Jesús; SALLES, Eduardo Baldisserra Carvalho. Algoritarismos. São Paulo-Valencia, Tirant lo Blanch, 2020.

¹⁶ SILVA, Tarcízio. Racismo algorítmico: inteligência artificial e descriminalização das redes sociais. São Paulo: Sesc, 2022.

¹⁷ O escândalo envolvendo a Cambridge Analytica, revelado em 2018, exemplifica de forma emblemática os riscos associados à coleta e ao uso indevido de dados pessoais. A empresa utilizou informações extraídas de perfis no Facebook, muitas vezes sem consentimento explícito, para construir perfis psicográficos detalhados de eleitores e influenciar comportamentos políticos em campanhas como o Brexit e as eleições presidenciais dos Estados Unidos em 2016. O caso evidenciou as consequências da falta de transparéncia e regulação eficaz no uso de algoritmos por grandes corporações tecnológicas.

¹⁸ BIGO, Didier. (In)segurança globalizada: o campo e o ban-opticon. In: AMARAL, Augusto Jobim do et al. (Orgs.). A cidade como máquina biopolítica. Valencia: Tirant lo blanch, 2022, p. 116-154.

¹⁹ Filósofo e jurista inglês do século XVIII, elaborou o conceito de panóptico em sua obra "Panopticon; or, The Inspection House", escrita em 1787 (embora publicada postumamente apenas em 1791). Nessa obra ao autor propõe um modelo arquitetônico para instituições como prisões, hospitais e escolas, cujo principal objetivo era permitir a vigilância constante e sem que estes soubessem se estão ou não sendo observados. A eficácia do panóptico está justamente nessa incerteza, que leva os vigiados a internalizarem a vigilância e a regularem seus próprios comportamentos.

estruturas arquitetônicas não dependem do confinamento; são aplicadas no intenso fluxo de circulação para incidência sobre indivíduos específicos; as leis e medidas administrativas são adaptadas a conceitos genéricos que substituem a ideia de dano pela ideia de risco, associada a uma determinada imagem, aquela capturada pela tecnologia algorítmica de reconhecimento facial²⁰.

Essa vigilância (in)visível é estimulada por um comportamento preditivo, onde o objetivo é antecipar as necessidades e desejos dos indivíduos, moldando a oferta de produtos e serviços de acordo com seus perfis. Pode-se dizer, com isso, que o modelo de vigilância e poder concebido por Jeremy Bentham em 1785, e amplamente problematizado por Foucault ganha um uma nova - e sofisticada - engrenagem.

Foucault²¹ demonstra como o poder disciplinar se consolida a partir da vigilância constante, exemplificada pelo panóptico de Bentham, onde a possibilidade de ser observado gera auto policiamento. No contexto jurídico contemporâneo, essa lógica é transposta para o ambiente digital, em que algoritmos operam uma vigilância silenciosa e contínua, voltada à previsão de condutas e à modelagem de decisões. Trata-se de um poder difuso, que não apenas pune, mas antecipa comportamentos, inserindo-se nas estruturas normativas e nos direitos fundamentais, especialmente no que tange à privacidade, à autodeterminação informativa e à proteção de dados pessoais.

Portanto, embora Jeremy Bentham não tenha falado diretamente do banóptico, sua concepção do panóptico foi fundamental para que pensadores contemporâneos, como Michel Foucault e depois Didier Bigo, elaborassem críticas e ampliações que deram origem a noções como o banóptico — especialmente relevantes no contexto da vigilância digital e controle securitário no século XXI.

Assim como no panóptico de Bentham, a hipervigilância algorítmica cria um ambiente de controle constante, onde os indivíduos são monitorados e analisados em seus mínimos detalhes. No entanto, enquanto o panóptico era um espaço físico e visível, o capitalismo de vigilância instigado por este modelo de organização social opera de forma invisível e pervasiva no mundo digital.

Esse novo modelo econômico e político²² proposto por Zuboff²³ funciona como um sistema que transforma dados comportamentais em mercadorias preditivas, visando lucro por meio da antecipação e influência de condutas. Diferente do controle disciplinar físico descrito por Foucault, o capitalismo de vigilância atua de forma invisível, extraíndo informações sem consentimento explícito e moldando ações futuras com base em algoritmos. Essa lógica amplia a vigilância panóptica, tornando-a difusa, contínua e profundamente inserida nas dinâmicas do mundo digitalizado, onde a liberdade individual é tensionada pela mercantilização da vida cotidiana.

²⁰ TRAVASSOS, Gabriel Saad; AMARAL, Augusto Jobim. O sistema europeu de reconhecimento facial: o dispositivo banóptico e as novas faces do positivismo. Anais do 7º Congresso Internacional de Direito e Contemporaneidade: mídias e direitos da sociedade em rede, 2024, p.9. Disponível em: <https://www.ufsm.br/app/uploads/sites/563/2024/12/2.1.pdf> Acesso em: 19 dez. 2024.

²¹ FOUCALT, Michel. Vigiar e punir: nascimento da prisão. Trad. Raquel Ramalhete. 42 ed. Rio de Janeiro: Vozes, 2021.

²² O termo "capitalismo de vigilância" foi criado por Shoshana Zuboff para descrever uma lógica econômica baseada na extração massiva de dados comportamentais dos indivíduos, muitas vezes sem seu consentimento explícito, com o objetivo de prever e influenciar futuras ações. Esse modelo transforma a experiência humana em matéria-prima gratuita para práticas comerciais que visam lucro por meio da vigilância digital contínua. Trata-se de uma forma inédita de poder, que combina tecnologia, mercado e controle social, deslocando o eixo do poder disciplinar para um poder preditivo e invisível, profundamente enraizado nas estruturas digitais contemporâneas.

²³ ZUBOFF, Shoshana. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York: PublicAffairs, 2019.

Assim, a regulamentação das plataformas digitais²⁴ para minimizar ou até mesmo neutralizar o discurso de ódio extremista, especialmente contra adolescentes, coloca todo o campo do direito, em especial na parcela que toca ao direito penal - sem desconsiderar a complexidade das ciências penais - diante de desafios inéditos. A dinâmica com que os conteúdos são produzidos e disseminados no ambiente digital exigem uma adaptação constante das normas jurídicas, que no mais das vezes - afora a emergência penal de plantão voltada contra os mesmos de sempre - não acompanham a evolução tecnológica.

O Marco Civil da Internet²⁵ constitui o principal marco regulatório do uso da internet no Brasil, estabelecendo princípios, garantias, direitos e deveres para usuários e provedores. Em relação às plataformas digitais, a norma adota uma abordagem de responsabilidade condicionada, prevendo que os provedores de aplicações de internet — como redes sociais e serviços de compartilhamento de conteúdo — só podem ser responsabilizados civilmente por conteúdos de terceiros mediante o descumprimento de ordem judicial específica (art. 19). Essa diretriz busca preservar a liberdade de expressão e evitar a censura privada, ao mesmo tempo em que assegura instrumentos legais para a remoção de conteúdos ilícitos. Excepcionalmente, o art. 21 prevê a possibilidade de responsabilização sem necessidade de ordem judicial para casos de exposição de nudez ou atos sexuais de caráter privado.

Contudo, o Marco Civil²⁶ demonstra limitações frente aos desafios atuais impostos pela atuação das plataformas digitais, especialmente no que se refere à moderação algorítmica de conteúdo, à transparência dos critérios de recomendação e à proliferação de discursos de ódio. Embora traga dispositivos relevantes sobre proteção de dados (art. 7º e 10) e dever de transparência (art. 7º, VIII e IX), a norma não impõe obrigações específicas quanto à governança dos algoritmos ou ao uso de inteligência artificial para fins de vigilância e manipulação de comportamento. Nesse cenário, ganha relevo a necessidade de atualização e complementação do Marco Civil, seja por meio de normas infralegais, seja pela aprovação de projetos como o PL 2630/2020 (PL das Fake News), que busca estabelecer diretrizes mais rigorosas para a atuação das plataformas na esfera pública digital.

O presente projeto legislativo²⁷ emerge de discussões acerca da necessidade de se instituírem mecanismos normativos voltados à responsabilização pela disseminação de informações falsas e engonosas, as quais vêm comprometendo os fundamentos da democracia brasileira. Tal debate foi incorporado à agenda do Poder Legislativo em decorrência de um conjunto de fatores específicos do contexto social

²⁴ No Brasil, a questão vem sendo debatida no Supremo Tribunal Federal, notadamente a partir do Recurso Extraordinário nº 1037396 interposto pelo Facebook Serviços Online do Brasil Ltda., relatado pelo Ministro Dias Toffoli, após decisão proferida pelo Tribunal de Justiça de São Paulo determinando a exclusão de um perfil falso das redes sociais, bem como através do Recurso Extraordinário nº 1057258, de relatoria do Ministro Luiz Fux, onde a Google Brasil Internet S/A contestou uma decisão por não excluir do Orkut uma comunidade criada para ofender uma pessoa, julgando procedente uma indenização por danos morais. Ambos os recursos discutem a possibilidade de responsabilidade civil das plataformas digitais e a insuficiência do marco civil da internet para tratar da matéria, não avançando para o plano da responsabilidade criminal propriamente dita, o que constituiria uma nova manifestação do ativismo judicial, a exemplo da criminalização da homofobia, sendo ideal que o tema também venha a ser discutido no âmbito do Poder Legislativo, na eventualidade de tipificação penal da matéria. De qualquer sorte, até a presente data, ambos os recursos foram retirados de pauta, em decorrência do pedido de vista do Ministro André Mendonça.

²⁵ BRASIL. Lei n. 12.965, de 23 de abril de 2014. Marco Civil da Internet. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2014/lei/l12965.htm. Acesso em: 16 fev. 2025.

²⁶ Ibid.

²⁷ BRASIL. Projeto de Lei nº 2630, de 2020, e seus apensados. Institui a Lei Brasileira de Liberdade, Responsabilidade e Transparéncia na Internet. Parecer Preliminar de Plenário de 25 de mar. de 2023. Relator Deputado Orlando Silva. Brasília: Portal da Câmara dos Deputados, 2023b. Disponível em: https://www.camara.leg.br/proposicoesWeb/prop_mostrarIntegra?codteor=2265334&filename=PRLP+1+%3D%3E+PL+2630/2020. Acesso em: 10 abr. 2025.

brasileiro, marcado por características singulares. Destaca-se, nesse sentido, a intensificação da desinformação durante o processo eleitoral de 2018, bem como, nos anos subsequentes, o aumento expressivo da circulação de conteúdos enganosos relacionados à pandemia de Covid-19.

Dante desse cenário, o Projeto de Lei passou a dispor sobre a regulamentação da atuação das plataformas digitais no que tange à veiculação e propagação de conteúdos potencialmente danosos nesses ambientes virtuais. No que se refere ao seu escopo de aplicação, uma vez aprovado, o projeto incidirá prioritariamente sobre redes sociais, mecanismos de busca, indexadores, plataformas de compartilhamento de vídeos e serviços de mensageria que possuam mais de 10 milhões de usuários²⁸.

Durante sua tramitação legislativa, o texto originalmente protocolado foi objeto de diversas alterações, resultantes da participação de distintos setores da sociedade. Em junho de 2020, o projeto foi aprovado no Senado Federal, sendo posteriormente submetido à Câmara dos Deputados, onde teve seu regime de urgência aprovado em abril de 2023 para deliberação em plenário. Ao longo desse período, o texto foi analisado por um grupo de trabalho constituído na Câmara, que promoveu reuniões técnicas e deliberativas, tanto presenciais quanto remotas, além de audiências públicas com especialistas de diversas áreas.

A versão mais recente do texto, consubstanciada no parecer do relator deputado Orlando Silva, versa, em essência, sobre a promoção da transparência por parte dos provedores quanto às suas atividades junto aos usuários; o estabelecimento de parâmetros para moderação e recomendação de conteúdos pelas plataformas; e a identificação explícita de conteúdos publicitários. Além disso, o Projeto de Lei objetiva fomentar um ambiente digital livre de práticas de assédio e discriminação.

No tocante à responsabilização das plataformas pela disseminação de conteúdos nocivos, o texto prevê a responsabilidade civil e solidária pela reparação de danos exclusivamente nos casos em que tais conteúdos estejam associados a publicidades veiculadas nas plataformas digitais. Importa destacar que essa forma de responsabilização não se aplica aos conteúdos gerados por usuários, salvo nas hipóteses em que for constatado o descumprimento do dever de cuidado por parte da plataforma²⁹.

Se observa que a complexidade tecnológica, a natureza global da *internet* e a necessidade de conciliar diferentes interesses, como a liberdade de expressão e a segurança pública, tornam o desafio de regulamentar ainda mais complexo. A tipificação penal³⁰ de crimes praticados no ambiente digital, especialmente aqueles relacionados ao discurso de ódio, apresenta uma série questões ainda mais desafiadoras que exigem uma constante atualização e adaptação do ordenamento jurídico.

Quanto aos desafios *técnicos* para enfrentamento da temática, tem-se ausência de uma definição legal precisa e consensual sobre o que constitui discurso de ódio, o que dificulta a aplicação da lei de forma uniforme. A subjetividade na interpretação desse conceito pelos tribunais pode levar a interpretações divergentes, a aplicação desigual da lei e o recrudescimento de estereótipos arraigados ao sistema capitalista,

²⁸ Ibid., p.61.

²⁹ Ibid.

³⁰ Isto não significa que se esteja circunscrevendo a tentativa de minimização da problemática dos discursos de ódio ao mero universo da tipificação penal, sobretudo em relação a responsabilidade dos proprietários das plataformas, sob pena de não legitimar as denúncias do expansionismo penal e a postura epistemológica não cartesiana - complexa e calcada na epistemologia da incerteza - adotada nesta investigação. O que se está a dizer é que, diante de tantos tipos penais defasados, de pouca utilidade ou que potencializam a seletividade do sistema de justiça criminal, ainda que não seja a "solução" adequada, a tipificação penal e a consequente persecução penal (naturalmente que assegurados os direitos e garantias fundamentais) pode ser um caminho inicial para que o campo do direito não passe à margem desta discussão, funcionando, em alguma medida, como diques de contenção de natureza normativa e jurisprudencial.

sendo que eventual tipificação deve atender a legalidade estrita, com tipos penais certos, claros e determinados, na medida em que é preciso ter cuidado para não se procurar soluções mágicas no punitivismo reducionista, sobretudo em países da América Latina.

Ademais, a natureza indireta de tais condutas - visto que se valem da utilização de algoritmos e sistemas automatizados para sua propagação – torna ainda mais complexa a identificação da(s) conduta(s) típica(s), desafiando não só o direito penal como também o processo penal, e consequentemente, todo o sistema de justiça criminal, tornando-se fundamental, para tanto, os apontamentos da criminologia crítica.

A colaboração entre diferentes atores, muitas vezes anônimos e distribuídos geograficamente, dificulta a individualização da responsabilidade. Esta ação pode ocorrer de forma remota e indireta, o que dificulta - ainda mais - a atribuição de responsabilidade.

Não bastasse isso, no contexto digital, a formação da prova (substituir por elementos indiciários) também é um desafio técnico complexo. A coleta, preservação e valoração de provas digitais exigem a utilização de ferramentas específicas e um conhecimento prévio acerca do manejo das tecnologias, o que se torna um obstáculo para o usuário médio. Além disso, a volatilidade de tais elementos - sejam eles indiciários ou probatórios - a possibilidade de adulteração e a complexidade dos sistemas informáticos demandam cuidados especiais para garantir a integridade e a autenticidade dos elementos indiciários ou probatórios, sendo fundamental a preservação da cadeia de custódia.

Além dos desafios técnicos, o uso de algoritmos na moderação de conteúdos sensíveis como discursos de ódio neonazistas enfrenta barreiras éticas³¹ significativas. O conceito de “racismo algorítmico”, como descrito por Silva³², destaca a possibilidade de os algoritmos reproduzirem preconceitos presentes nos dados de treinamento. Quando treinados com dados históricos que refletem vieses sociais, esses sistemas podem reforçar estereótipos ou discriminar grupos específicos, levando a uma moderação que prejudica a igualdade de tratamento dos usuários. Isso é particularmente relevante para adolescentes, que são mais vulneráveis ao impacto psicológico de uma moderação injusta ou equivocada, sendo que hoje a sua presença – ainda que virtual – nas redes sociais é tão preocupante quanto a circulação nos grandes centros urbanos cercados por violência física, daí porque, também, a necessidade de refletir sobre a violência em uma perspectiva sistêmica.

A questão da transparência também é um desafio ético central. Como Baumann e Lyon³³ discutem em *Vigilância Líquida*, os sistemas de vigilância digital – incluindo algoritmos de moderação – funcionam, muitas vezes, como uma “caixa-preta” cujos processos internos são desconhecidos para os usuários e até para os próprios operadores. Essa opacidade algorítmica cria barreiras à responsabilização das plataformas e limita a capacidade dos usuários e reguladores de entender e questionar as decisões tomadas pela IA. A falta de transparência torna difícil avaliar se as decisões de moderação foram justas e adequadas, minando a confiança dos usuários na plataforma e, possivelmente, violando direitos fundamentais como a liberdade de expressão.

Além dos aspectos éticos, a aplicação de algoritmos para moderar discursos de ódio neonazistas enfrenta desafios jurídicos que questionam a adequação das tecnologias automáticas ao contexto de direitos

³¹ O Regulamento (UE) 2022/2065 exige transparência na moderação de conteúdo, mas não aborda diretamente os vieses nos algoritmos. A prática de "racismo algorítmico" pode levar à discriminação e reforçar estereótipos, comprometendo a imparcialidade da moderação. Embora o DSA busque proteger grupos vulneráveis, como adolescentes, a falta de uma reflexão crítica sobre os dados de treinamento utilizados nos algoritmos pode resultar em censura implícita e em injustiças na aplicação das regras.

³² SILVA, Tarcízio. Racismo algorítmico: inteligência artificial e descriminalização das redes sociais. São Paulo: Sesc, 2022.

³³ BAUMAN, Zygmunt. LYON, David. Vigilância Líquida. Ciudad Autónoma de Buenos Aires: Paidós, 2013.

fundamentais. Mendes e Freitas³⁴ analisam a interseção entre novas tecnologias e o direito constitucional e penal, destacando que qualquer forma de restrição de conteúdo deve estar alinhada com princípios constitucionais, como a proporcionalidade e a proteção à liberdade de expressão, mas está também não pode ser usada para mascarar discursos de ódio. Quando algoritmos falham ao identificar corretamente conteúdos nocivos, há um risco de censura excessiva ou insuficiente, o que implica em uma violação desses princípios, o que potencializa e complexifica a temática da presente investigação.

Para enfrentar esses desafios, a moderação de conteúdos digitais deve ser acompanhada por sistemas de supervisão humana. Nesse sentido, Giacomelli³⁵ propõe a implementação de um “gerenciamento tecnológico” no sistema de justiça penal, que inclui a revisão das decisões algorítmicas por profissionais humanos como forma de corrigir os possíveis erros e vieses dos sistemas automatizados. Essa supervisão mista, combinando IA com avaliação humana, pode aumentar a precisão e a legitimidade das decisões de moderação, especialmente em casos complexos como o discurso de ódio neonazista.

Algoritmos e inteligência artificial representam um avanço importante na moderação de conteúdo digital e na proteção dos adolescentes contra discursos de ódio, mas sua eficácia é limitada pelas complexidades da linguagem e pelas nuances culturais e contextuais dos discursos neonazistas. As limitações técnicas, somadas aos desafios éticos e jurídicos, indicam que, embora úteis, esses sistemas não podem ser considerados - naturalmente - uma solução completa para a proteção de adolescentes nas plataformas digitais.

O estudo sobre os algoritmos e a IA na moderação de discursos de ódio neonazista destaca a importância de uma abordagem equilibrada e supervisionada que proteja os adolescentes de conteúdos nocivos sem violar seus direitos fundamentais. A constante revisão e aprimoramento desses sistemas, aliado a um monitoramento humano, são passos essenciais para garantir que a tecnologia realmente contribua para a construção de um ambiente digital mais seguro e inclusivo.

3 POTENCIALIDADES E LIMITES DO DIREITO PENAL E CONSTITUCIONAL NA REGULAÇÃO DOS DISCURSOS DE ÓDIO CONTRA ADOLESCENTES NO AMBIENTE DIGITAL EM UM CONTEXTO DEMOCRÁTICO GLOBALIZADO

A disseminação de discursos de ódio no ambiente digital, especialmente os direcionados contra adolescentes, representa um desafio crítico para o Direito Penal³⁶ e o Direito Constitucional, que enfrentam o complexo papel de proteger direitos fundamentais enquanto preservam valores democráticos e a liberdade de expressão. Em um cenário globalizado e digitalizado³⁷, as normas penais e constitucionais

³⁴ MENDES, Gilmar Ferreira; FREITAS, Matheus Pimenta. Constituição, Direito Penal e novas tecnologias. São Paulo: Almedina, 2023.

³⁵ GIACOMELLI, Felipe Mrack. Gerenciamento tecnológico do sistema de justiça penal. As novas tecnologias no âmbito do policiamento, da investigação e da decisão. Madri: Marcial Pons, 2023.

³⁶ Certo que os desafios não se concentram somente nesta área das ciências criminais, aqui recortada por opção metodológica desta pesquisa, o que não significa que a mesma não esteja atenta com as intersecções realizadas com o processo penal por exemplo, notadamente na fase da investigação preliminar a das dificuldades indicíarias para encontrar elementos de autoria e materialidade em ambientes digitais não raras vezes ocultos ao dia a dia da autoridade policial, demandando a necessidade de Delegacias especializadas para tentar minimizar tal problemática advinda da reinvenção da criminalidade.

³⁷ Embora frequentemente usados como sinônimos, os termos "mundo digitalizado" e "mundo virtual" possuem distinções conceituais relevantes. O "mundo digitalizado" refere-se à realidade concreta mediada por tecnologias digitais, na qual dados

devem equilibrar a necessidade de proteção dos indivíduos, sobretudo os mais vulneráveis, com a manutenção de um espaço público de debate e expressão. Esse desafio é ainda mais complicado pela natureza transnacional das redes sociais, que operam sob jurisdições e legislações diversas, exigindo um esforço de harmonização normativa e de cooperação internacional³⁸.

O Direito Penal, tradicionalmente voltado para a repressão de condutas prejudiciais à ordem pública e à integridade dos indivíduos, ao mesmo tempo que deve funcionar como limitação ao poder punitivo, enfrenta dificuldades para atuar no controle dos discursos de ódio disseminados nas plataformas digitais, especialmente quando envolvem ideologias neonazistas e outras formas de extremismo. Conforme Mendes e Freitas³⁹ discutem, a aplicação do Direito Penal nesse contexto demanda uma adaptação normativa que reconheça as particularidades do ambiente digital, no qual a velocidade de propagação de conteúdos e o anonimato dificultam a identificação e responsabilização dos infratores, tal como já destacado ao longo desta investigação. Além disso, a proteção de adolescentes é ainda mais complexa devido à vulnerabilidade inerente a essa faixa etária⁴⁰, que torna os jovens mais suscetíveis ao impacto psicológico e social do discurso de ódio *online*.

A natureza global da internet gera conflitos de jurisdição, uma vez que um crime cometido em um país pode ter efeitos em outros. A aplicação de leis penais a condutas ocorridas em outros países exige a análise de tratados internacionais e a compatibilização com o princípio da territorialidade⁴¹.

Para Giacomolli⁴², as tecnologias digitais também podem ser utilizadas no âmbito do sistema de justiça penal, não apenas para identificar conteúdos de ódio, mas também para aprimorar a investigação e a responsabilização dos agentes. Entretanto, deve-se partir da premissa de que não há neutralidade na rede.

Inobstante muitas vezes invisíveis para a maior parte das pessoas e apresentadas pelo mercado como uma tecnologia complexa, cuja compreensão é desnecessária desde que cumpridas suas finalidades, os algoritmos que integram os sistemas de Inteligência Artificial não são neutros. Eles, inexoravelmente, guardam as intenções de seus programadores ou desenvolvedores. Tal constatação é de grande importância, na medida em que os sistemas de Inteligência Artificial, em sua maioria, não são construídos e dirigidos exclusivamente aos setores públicos⁴³.

O mito da neutralidade das plataformas digitais é fator condicionante da avalanche de desinformação e violência online na atualidade. Encorajado pelas plataformas por meio de ferramentas criadas por elas para manter seus usuários cada vez mais conectados, em um aparente anonimato, a atual dinâmica das mídias digitais permitiu que as bases da democracia, fundada no pluralismo e no respeito às diferenças, se tornassem refém dos algoritmos.

pessoais, comportamentos e relações sociais são convertidos em informações analisáveis por sistemas automatizados. Já o "mundo virtual" diz respeito a ambientes simulados computacionalmente — como realidades aumentadas, metaversos ou jogos imersivos — que nem sempre mantêm vínculo direto com o mundo físico. No contexto da vigilância algorítmica e da proteção de dados, o foco recai sobre o mundo digitalizado, pois é nele que ocorre a coleta massiva e a utilização estratégica de informações reais dos indivíduos.

³⁸ BARROSO, Luís Roberto. Revolução tecnológica, crise da democracia e constituição: Direito e políticas públicas num mundo em transformação. Belo Horizonte: Fórum, 2020.

³⁹ MENDES, Gilmar Ferreira; FREITAS, Matheus Pimenta. Constituição, Direito Penal e novas tecnologias. São Paulo: Almedina, 2023.

⁴⁰ Entre 12 e 18 anos de idade, segundo o artigo 2º da Lei 8.069/90.

⁴¹ O direito penal internacional, decorrido mais de 60 anos do Tratado de Londres, ainda não ganhou a merecida notoriedade e consolidação enquanto um ramo específico do direito penal, estando o Tribunal Penal Internacional limitado aos crimes de genocídio, de lesa humanidade, de guerra e de agressão, como bem leciona Luis Luisi.

⁴² GIACOMOLLI, Felipe Mrack. Gerenciamento tecnológico do sistema de justiça penal. As novas tecnologias no âmbito do policiamento, da investigação e da decisão. Madri: Marcial Pons, 2023.

⁴³ Ibid., p.51.

A aplicação de ferramentas tecnológicas no Direito Penal levanta questões sobre o devido processo legal, o que exige uma abordagem ponderada que respeite os direitos dos investigados enquanto oferece proteção efetiva às vítimas, como os adolescentes. A repressão penal dos discursos de ódio, portanto, deve estar alinhada com os princípios constitucionais de proporcionalidade e legalidade, a fim de evitar excessos que possam resultar em censura ou repressão indevida.

No mesmo sentido, a obra de Javier Valls Prieto⁴⁴ apresenta um marco importante para discussão sobre a interface entre direito penal, a inteligência artificial e as novas tecnologias. O autor reconhece as potencialidades da IA como instrumento de prevenção e repressão criminal, sobretudo no aprimoramento da eficiência investigativa e na antecipação de condutas ilícitas. No entanto, adverte para os riscos significativos que esse avanço tecnológico impõe aos direitos fundamentais, especialmente no que tange à proteção da privacidade, à liberdade de expressão e à segurança jurídica.

O estudo da intersecção entre o direito penal e tecnologias emergentes permeia o debate ao redor do papel da inteligência artificial em sistemas de vigilância e os desafios que a absorção destas tecnologias pelos sistemas de justiça representa para a preservação de direito, especialmente os de grupos vulneráveis. Pietro versa que não se pode ignorar tais desafios jurídicos que surgem com a utilização de novas técnicas de investigação e prevenção do crime de maneira que se deve buscar identificar as melhores formas de regular as plataformas digitais para combater o discurso de ódio neonazista, sem comprometer outros direitos fundamentais.

Neste sentido, urge a necessidade de encontrar um equilíbrio entre a segurança pública e a proteção dos direitos fundamentais no contexto das tecnologias e na sociedade hiperconectada. Ao considerar esses aspectos, é possível construir um marco regulatório que seja eficaz na proteção da sociedade, sem comprometer os direitos fundamentais dos indivíduos.

O Direito Constitucional oferece um arcabouço essencial para a regulação dos discursos de ódio no ambiente digital, uma vez que é responsável por definir e proteger os direitos e garantias fundamentais dos cidadãos, incluindo a liberdade de expressão. Conforme Barroso⁴⁵ aponta, a liberdade de expressão é um dos pilares da democracia, mas seu exercício deve ser compatível com o respeito aos direitos dos demais, especialmente quando há risco de incitação ao ódio ou à violência contra grupos específicos, como os adolescentes. Nesse sentido, a questão da regulação dos discursos de ódio neonazistas nas redes sociais desafia a interpretação constitucional da liberdade de expressão, que precisa ser adaptada para garantir a proteção contra conteúdos ofensivos e potencialmente perigosos, sem comprometer o direito à manifestação de ideias e opiniões.

Na Ação Penal 2493, o Supremo Tribunal Federal (STF), sob relatoria do ministro Alexandre de Moraes, reafirmou que o discurso de ódio, especialmente quando incita violência, ataca a dignidade de grupos sociais ou promove ataques a instituições democráticas, não está protegido pela liberdade de expressão. O Tribunal reconheceu que declarações homofóbicas configuraram crime com base na Lei de Racismo e entendeu que o uso de redes sociais para incitar ataques contra o Congresso, STF ou TSE constitui crime contra o Estado Democrático de Direito. Essa decisão reforça a compreensão de que o combate ao

⁴⁴ PIETRO, Javier Valls. *Problemas jurídico penales asociados a las nuevas técnicas de prevención y persecución del crimen mediante inteligencia artificial*. Madri: Dykinson, 2017.

⁴⁵ BARROSO, Luís Roberto. Revolução tecnológica, crise da democracia e constituição: Direito e políticas públicas num mundo em transformação. Belo Horizonte: Fórum, 2020.

discurso de ódio não se opõe aos princípios constitucionais, mas, ao contrário, os concretiza ao assegurar um ambiente de respeito à diversidade, à dignidade e à democracia⁴⁶.

A regulação constitucional e penal dos discursos de ódio neonazistas enfrenta, porém, um dilema fundamental: como identificar e restringir o discurso prejudicial sem comprometer a essência do espaço público digital? Conforme Silva⁴⁷ observa em *Racismo Algorítmico*, o uso de tecnologias automatizadas na moderação de conteúdos digitais pode introduzir viés e discriminação nos sistemas de censura, atingindo de maneira desigual certos grupos e limitando indevidamente sua participação nas plataformas. Esse cenário alerta para a necessidade de um controle cuidadoso e transparente da aplicação das normas constitucionais, que preserve o equilíbrio entre proteção contra abusos e promoção de um espaço digital inclusivo e democrático, tarefa difícil e complexa, tendo em vista que o direito e o sistema de justiça como um todo não consegue acompanhar a velocidade da reengenharia delitiva.

A natureza global das redes sociais adiciona uma camada de complexidade à regulação penal e constitucional dos discursos de ódio. Muitas das plataformas digitais operam em diversas jurisdições, o que cria uma fragmentação normativa que dificulta a imposição de medidas de proteção e responsabilização uniformes, (re)desafiando, inclusive, o direito penal internacional. Segundo Bauman e Lyon⁴⁸, a “vigilância líquida” característica do ambiente digital contemporâneo transcende fronteiras e desafia os Estados-nação a encontrar formas colaborativas de regulamentação. Esse cenário exige uma cooperação jurídica internacional e uma harmonização das leis entre diferentes países, para que seja possível estabelecer padrões de proteção que possam ser aplicados de forma eficaz em todas as jurisdições onde atuam as plataformas digitais.

Barroso⁴⁹ ressalta que a globalização e a transnacionalidade dos meios digitais exigem uma renovação do pensamento jurídico, que incorpore a perspectiva da governança digital e a busca de soluções normativas universais. A regulamentação dos discursos de ódio, especialmente os direcionados a adolescentes, demanda a criação de normas internacionais e acordos multilaterais que facilitem o intercâmbio de informações entre os países e viabilizem ações conjuntas de monitoramento e responsabilização. No entanto, tal regulamentação deve ser cautelosa para evitar a imposição de restrições generalizadas à liberdade de expressão e garantir que as políticas de moderação estejam em conformidade com os princípios democráticos.

O desafio de regular discursos de ódio contra adolescentes no ambiente digital exige que o Direito Penal e o Direito Constitucional atuem de forma integrada⁵⁰, buscando um equilíbrio entre a repressão de conteúdos nocivos e a proteção da liberdade de expressão. Ao relacionar essas duas áreas do conhecimento,

⁴⁶ BRASIL. Supremo Tribunal Federal. Ação Penal 2493/DF. Relator: Min. Alexandre de Moraes. Julgamento em 23 de novembro de 2023a. Brasília, DF: STF, 2023. Disponível em: <https://redir.stf.jus.br/paginadorpub/paginador.jsp?docTP=TP&docID=785246760>. Acesso em: 14 abr. 2025.

⁴⁷ SILVA, Tarcízio. *Racismo algorítmico: inteligência artificial e descriminalização das redes sociais*. São Paulo: Sesc, 2022.

⁴⁸ BAUMAN, Zygmunt. LYON, David. *Vigilância Líquida*. Ciudad Autónoma de Buenos Aires: Paidós, 2013.

⁴⁹ BARROSO, Luís Roberto. *Revolução tecnológica, crise da democracia e constituição: Direito e políticas públicas num mundo em transformação*. Belo Horizonte: Fórum, 2020.

⁵⁰ É importante situar a questão no plano normativo, para que o campo do direito consiga reagir às condutas lesivas aos adolescentes. Ao mesmo tempo, isso é só a ponta do iceberg e significaria, em certa medida, reapostar na teoria da prevenção geral da penal, o que não é o intuito desta investigação, na medida em que não se pode tratar um problema complexo de maneira cartesiana. Ainda assim, alguma moldura jurídica normativa precisa ser estabelecida, e dada a gravidade das condutas – no momento não típicas – parece ser justificável a intervenção penal, sobretudo diante da proteção integral preconizada pelo Estatuto da Criança e do Adolescente.

é possível construir um arcabouço jurídico sólido e coerente para enfrentar os desafios impostos pelas novas tecnologias.

O direito penal, por sua vez, oferece ferramentas para a tipificação de condutas ilícitas praticadas nas plataformas digitais, como a disseminação de discursos de ódio ou a produção e compartilhamento de conteúdo pedófilo, cabendo apontar que nem mesmo a gravidade das condutas, que muitas vezes pode contribuir com quadro depressivos, instigação de homicídios a até mesmo suicídio, dentre outras formas de difusa violência, devem justificar a inobservância de tipos penais claros e precisos, como forma de evitar, também, o expansionismo punitivista a qualquer preço.

Já o direito constitucional garante os direitos fundamentais dos indivíduos, como a liberdade de expressão, a privacidade e a igualdade, que devem ser ponderados ao se estabelecer qualquer tipo de restrição à liberdade de expressão online.

Em um contexto globalizado e perverso, marcado pela tirania do dinheiro e da informação⁵¹ torna-se indispensável que o Direito Penal se adapte às especificidades das redes digitais e que o Direito Constitucional estabeleça diretrizes claras para a proteção dos direitos fundamentais nas plataformas digitais, considerando as nuances éticas e jurídicas envolvidas. Dessa forma, a análise conjunta do direito penal e do direito constitucional permite identificar os limites da intervenção estatal no ambiente digital, garantindo, ao mesmo tempo, a proteção da sociedade e a promoção de um ambiente virtual seguro e democrático.

A discussão acerca da regulamentação de plataformas digitais não deve implicar em um controle estatal excessivo que comprometa a autonomia das grandes empresas de tecnologia. Ao contrário, deve buscar estabelecer um marco regulatório - para além do marco civil da internet - que promova a isonomia e a legitimidade nas relações digitais, mitigando os desequilíbrios de poder inerentes ao ambiente virtual.

A assimetria informacional presente nas plataformas, caracterizada pela influência desproporcional dos algoritmos sobre as massas, torna-se um terreno fértil para a disseminação de desinformação e propagação de discursos de ódio. Nesse contexto, a responsabilização dos intermediários digitais, cujos algoritmos amplificam e disseminam conteúdos nocivos, emerge como uma necessidade imperativa para garantir a integridade do espaço digital e a proteção dos direitos dos usuários.

Nesse mesmo contexto, destaca-se o julgamento do Recurso Extraordinário nº 1.037.396/SP, em que o Supremo Tribunal Federal enfrentou a responsabilidade das plataformas digitais quanto à remoção de conteúdos ofensivos. A Corte entendeu que, uma vez notificada judicialmente sobre a presença de material ilegal ou odioso, a plataforma tem o dever de agir prontamente para removê-lo, sob pena de responder solidariamente pelos danos causados. Essa decisão reforça a necessidade de regulamentação mais efetiva das plataformas, especialmente para proteger grupos vulneráveis, como adolescentes, diante da disseminação de discursos de ódio online⁵².

Por conseguinte, a abordagem penal e constitucional sobre os discursos de ódio digital deve buscar o desenvolvimento de soluções cooperativas e transnacionais, fundamentadas na proteção dos adolescentes e no respeito aos princípios do Estado de Direito. Considerando que a propagação de ideias voltadas à inferiorização ou eliminação de grupos sociais configura conduta incompatível com os princípios

⁵¹ SANTOS, Milton. Por uma outra globalização. Do pensamento único à consciência universal. 5º ed. Rio de Janeiro/São Paulo: Record, 2001, p.38.

⁵² BRASIL. Supremo Tribunal Federal. Petição nº 12.404. Relator: Min. Alexandre de Moraes. Decisão monocrática referendada. Brasília, DF, 2024. Disponível em: <https://portal.stf.jus.br/processos/detalhe.asp?incidente=5160549>. Acesso em: 7 abr. 2025.

democráticos, a utilização estratégica das redes sociais para amplificar discursos de ódio não apenas intensifica seus efeitos nocivos, como também exige uma resposta firme e coordenada por parte do Estado.

Dessa forma, com a globalização das redes sociais e a expansão dos discursos extremistas, cuja utilização da propaganda é um dos seus *modus operandi*, com a malícia de usar as liberdades democráticas contra elas mesmas⁵³ o Estado de Direito enfrenta uma de suas mais complexas demandas no ambiente digital, que requer inovações jurídicas para garantir a segurança e o bem-estar dos jovens usuários das plataformas, sem abrir mão dos valores democráticos.

4 CONSIDERAÇÕES FINAIS

A presente pesquisa buscou responder ao problema de como o Direito Penal e o Direito Constitucional podem atuar de forma eficaz e proporcional para regular a disseminação de discursos de ódio contra adolescentes nas plataformas digitais, em um contexto perversamente globalizado e democrático.

A partir da análise realizada, é possível concluir que, embora as tecnologias de inteligência artificial ofereçam ferramentas inovadoras para a moderação de conteúdos nocivos, há desafios técnicos, éticos e jurídicos que limitam sua eficácia e exigem complementação com abordagens normativas e princípios jurídicos sólidos, ainda que o plano normativo longe esteja de – por si só – proporcionar a solução mágica de um problema tão complexo a ser enfrentado pelas democracias contemporâneas.

Neste sentido, a regulação passa a ser um “mal necessário”, ou seja, é preciso filtrar – ainda que minimamente – o conteúdo dos discursos de ódio neonazistas denunciando sua autoria e apontando os elementos indiciários de materialidade a autoridade competente, ao mesmo tempo que se precisa cuidar para que o sistema jurídico não fique refém daquilo que as grandes plataformas de mídia - a maioria delas alocadas no Vale do Silício - passem a assumir, a partir de uma perspectiva privatista, os rumos das questões de ordem pública no âmbito das democracias, sobretudo em países propícios ao colonialismo digital.

Neste sentido, o direito penal e o processo penal, com as observações da criminologia crítica, deve adaptar suas normas e processos à realidade do ambiente digital, estabelecendo meios eficientes para a responsabilização dos envolvidos em discursos de ódio, ao mesmo tempo em que preserva garantias fundamentais como a proporcionalidade e a legalidade.

Esse ajuste é essencial para que as medidas punitivas não resultem em censura excessiva, especialmente quando as plataformas utilizam algoritmos que podem falhar na interpretação de conteúdos ambíguos ou simbólicos, como os discursos neonazistas objeto desta investigação. A supervisão humana, portanto, emerge como uma necessidade para assegurar a precisão e justiça nas decisões de moderação, oferecendo uma camada de proteção contra os erros inerentes aos sistemas automatizados.

Além disso, a globalização das plataformas digitais demanda uma resposta jurídica que ultrapasse fronteiras e harmonize legislações, permitindo que os Estados-nação cooperem na regulação de conteúdos prejudiciais. A criação de normas e padrões internacionais, que facilitem o intercâmbio de informações e promovam a responsabilidade compartilhada entre nações, seria um avanço significativo na proteção dos adolescentes, uma vez que a maioria das redes sociais opera de forma transnacional. Dessa forma, a fluidez

⁵³ STANLEY, Jason. Como funciona o fascismo. A política do nós e eles. 2º ed. Porto Alegre: L&PM, 2019, p.44.

das interações digitais, as políticas de monitoramento e controle do discurso de ódio precisam ser flexíveis, mas fundamentadas em um arcabouço legal que respeite os valores democráticos e os direitos dos usuários.

A resposta ao problema de pesquisa aponta, assim, para a necessidade de uma abordagem mista e cooperativa, na qual o direito penal, o direito constitucional e a tecnologia - quando em órbita o processo penal - atuem de forma complementar para combater os discursos de ódio contra adolescentes sem comprometer a liberdade de expressão, como uma tentativa de minimização do problema no âmbito normativo, insuficiente dada a complexidade secular do tema, mas necessário inclusive para o futuro dos Estados Democráticos.

A regulamentação efetiva e ética no contexto digital exige um esforço contínuo de adaptação das normas jurídicas, bem como o aprimoramento das tecnologias de moderação e o desenvolvimento de acordos internacionais que promovam um ambiente seguro e inclusivo para todos, particularmente para os jovens, que se encontram em uma fase de formação e são especialmente vulneráveis ao impacto de conteúdos extremistas, paradoxalmente, marca das sociedades contemporâneas.

Em suma, para que o Estado de Direito possa cumprir sua função de proteção no ambiente digital globalizado, é imperativo que as legislações nacionais e internacionais sejam revistas e que novas formas de cooperação sejam estabelecidas, de forma a equilibrar a repressão dos abusos e o respeito aos direitos fundamentais dos usuários.

REFERÊNCIAS

284

AMARAL, Augusto Jobim do; SABARIEGO, Jesús; SALLES, Eduardo Baldissera Carvalho. **Algoritarismos**. São Paulo-Valecia, Tirant lo Blanch, 2020.

BARROSO, Luís Roberto. **Revolução tecnológica, crise da democracia e constituição: Direito e políticas públicas num mundo em transformação**. Belo Horizonte: Fórum, 2020.

BAUMAN, Zygmunt. LYON, David. **Vigilância Líquida**. Ciudad Autónoma de Buenos Aires: Paidós, 2013.

BENTHAM, Jeremy. **O Panóptico**. 3 ed. Belo Horizonte: Autêntica, 2019.

BRASIL. **Lei nº 8069, de 13 de julho de 1990**. Dispõe sobre o Estatuto da Criança e do Adolescente e dá outras providências. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/l8069.htm. Acesso em: 08 ago. 2024.

BRASIL. **Lei nº 12.965, de 23 de abril de 2014**. Marco Civil da Internet. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2014/lei/l12965.htm. Acesso em: 16 fev. 2025.

BRASIL. **Lei nº 13.709, de 14 de Agosto de 2018**. Dispõe sobre a Lei Geral de Proteção de Dados Pessoais (LGPD). Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm. Acesso em: 10 fev. 2025.

BRASIL. Supremo Tribunal Federal. **Recurso Extraordinário nº 1.037.396/SP**. Recorrente: Facebook Serviços Online do Brasil Ltda. Recorrida: Lourdes Pavioto Corrêa. Relator: Ministro Dias Toffoli. Brasília, DF, julgado em 27 nov. 2024. Disponível em: <https://portal.stf.jus.br/processos/detalhe.asp?incidente=5160549>. Acesso em: 7 abr. 2025.

BRASIL. Supremo Tribunal Federal. **Ação Penal 2493/DF**. Relator: Min. Alexandre de Moraes. Julgamento em 23 de novembro de 2023a. Brasília, DF: STF, 2023. Disponível em: <https://redir.stf.jus.br/paginadorpub/paginador.jsp?docTP=TP&docID=785246760>. Acesso em: 14 abr. 2025.

BRASIL. **Projeto de Lei nº 2630, de 2020, e seus apensados**. Institui a Lei Brasileira de Liberdade, Responsabilidade e Transparéncia na Internet. Parecer Preliminar de Plenário de 25 de mar. de 2023. Relator Deputado Orlando Silva. Brasília: Portal da Câmara dos Deputados, 2023b. Disponível em: https://www.camara.leg.br/proposicoesWeb/prop_mostrarintegra?codteor=2265334&filename=PRLP+1+%3D%3E+PL+2630/2020. Acesso em: 10 abr. 2025.

BRASIL. Supremo Tribunal Federal. **Petição nº 12.404**. Relator: Min. Alexandre de Moraes. Decisão monocrática referendada. Brasília, DF, 2024. Disponível em: <https://portal.stf.jus.br/processos/detalhe.asp?incidente=5160549>. Acesso em: 7 abr. 2025.

BRASIL. Supremo Tribunal Federal. **Petição nº 9.935**. Relator: Min. Alexandre de Moraes. Julgamento em 17 de março de 2025. Brasília, DF, 2025. Disponível em: <https://jurisprudencia.stf.jus.br/pages/search/sjur526333/false>. Acesso em: 7 abr. 2025.

BIGO, Didier. (In)segurança globalizada: o campo e o ban-opticon. In: AMARAL, Augusto Jobim do *et al.* (Orgs.). **A cidade como máquina biopolítica**. Valencia: Tirant lo blanch, 2022, p. 116-154.

CGI.br. **TIC Kids Online Brasil 2024** [índices]. Disponível em: <https://cetic.br/pt/tics/kidsonline/2024/criancas/C7/>. Acesso em: 07 abr. 2025.

285

FOUCALT, Michel. **Vigiar e punir**: nascimento da prisão. Trad. Raquel Ramalhete. 42 ed. Rio de Janeiro: Vozes, 2021.

LUIS, Luis. Os princípios constitucionais penais. 2. ed. Revista e aumentada. Porto Alegre: Sergio Antonio Fabris, 2003.

GIACOMOLLI, Felipe Mrack. **Gerenciamento tecnológico do sistema de justiça penal**. As novas tecnologias no âmbito do policiamento, da investigação e da decisão. Madri: Marcial Pons, 2023.

MENDES, Gilmar Ferreira; FREITAS, Matheus Pimenta. **Constituição, Direito Penal e novas tecnologias**. São Paulo: Almedina, 2023.

MDHC. **Relatório de recomendações para o enfrentamento do discurso de ódio e o extremismo no Brasil**. Christian Ingo Lenz Dunker, Débora Diniz Rodrigues, Esther Solano. et al. / Camilo Onoda Luiz Caldas, Manuela Pinto Vieira d'Ávila, Brenda de Fraga Espindula. et al. (Coord.) - 1. ed. - Brasília: 2023.

PIETRO, Javier Valls. **Problemas jurídico penales asociados a las nuevas técnicas de prevención y persecución del crimen mediante inteligencia artificial**. Madri: Dykinson, 2017.

SANTOS, Milton. **Por uma outra globalização**. Do pensamento único à consciência universal. 5º ed. Rio de Janeiro/São Paulo: Record, 2001.

SILVA, Tarcízio. **Racismo algorítmico**: inteligência artificial e descriminalização das redes sociais. São Paulo: Sesc, 2022.

SILVEIRA, Paulo Antônio Caliendo Veloso da. **Ética e inteligência artificial.** Da possibilidade filosófica de agentes morais artificiais. Porto Alegre: Editora Fi, 2021.

STANLEY, Jason. Como funciona o fascismo. A política do nós e eles. 2º ed. Porto Alegre: L&PM, 2019.

TRAVASSOS, Gabriel Saad; AMARAL, Augusto Jobim. **O sistema europeu de reconhecimento facial: o dispositivo banóptico e as novas faces do positivismo.** Anais do 7º Congresso Internacional de Direito e Contemporaneidade: mídias e direitos da sociedade em rede, 2024. Disponível em: <https://www.ufsm.br/app/uploads/sites/563/2024/12/2.1.pdf> Acesso em: 19 dez. 2024.

UNIÃO EUROPEIA. **Regulamento (UE) 2022/2065.** Parlamento Europeu e do Conselho de 19 de outubro de 2022 relativo aos serviços digitais e que altera a Diretiva 2000/31/CE. Jornal Oficial da União Europeia, 2022. Disponível em: <https://eur-lex.europa.eu>. Acesso em: 14 abr. 2025.

ZUBOFF, Shoshana. **The Age of Surveillance Capitalism:** The Fight for a Human Future at the New Frontier of Power. New York: PublicAffairs, 2019.