DISCURSO DE ÓDIO E O EFEITO BORBOLETA NO COMPORTAMENTO ONLINE: PEQUENOS POSTS, GRANDES CONSEQUÊNCIAS

Ana Carolina Sassi
Doutoranda e Mestre pelo Programa de Pós-Graduação em Direito da Universidade Federal de Santa Maria (PPGD/UFSM) com pesquisa na área da Criança e do Adolescente
Bacharel em Direito pela Universidade Franciscana (UFN)
Membro do Núcleo de Pesquisa em Direito Informacional (NUDI/UFSM)
Membro do Núcleo de Direito Constitucional (NDC/UFSM)
Bolsista CAPES
ORCID: https://orcid.org/0009-0004-6637-9671
e-mail: acsassi@gmail.com

Rosane Leal da Silva
Doutora em Direito pela Universidade Federal de Santa Catarina (UFSC), com pesquisa sobre proteção de adolescentes na sociedade informacional
Professora Associada do Curso de Graduação e Mestrado em Direito da Universidade Federal de Santa Maria e do Curso de Direito da Universidade Franciscana
Líder do Grupo de Pesquisa Núcleo de Direito Informacional (NUDI/UFSM), onde se situa a presente pesquisa
ORCID: https://orcid.org/0000-0002-9636-2705,
e-mail: rolealdasilva@gmail.com

Recebido em: 04/08/2025
Aprovado em:16/09/2025

RESUMO

As dinâmicas complexas e interconectadas das redes sociais têm contribuído para a criação de ambientes nos quais o discurso de ódio e a polarização se expandem rapidamente. Por meio de pequenas interações com conteúdos de ódio, indivíduos acabam se envolvendo em conteúdos radicais, levando a um efeito cascata de comportamentos violentos ou intolerantes. Frente a isso, a polarização nas redes sociais muitas vezes tem início com pequenos desentendimentos ou discordâncias que, amplificados por meio do efeito borboleta e dos processos de comunicação, resultam em conflitos intensos e divisões entre grupos. O presente trabalho questiona: em que medida a moderação de conteúdo contra o discurso de ódio, conseguem conter a propagação de ódio nas redes sociais? Utilizando-se da abordagem interdisciplinar, o trabalho investiga a disseminação de conteúdos de ódio nas redes sociais, analisando o fenômeno sob a perspectiva do efeito borboleta, em articulação com estudos nas áreas de direito. Do estudo, conclui-se que a moderação de conteúdo pelas plataformas digitais é insuficiente para combater o discurso de ódio implícito, o que possibilita sua ampla disseminação pela rede, além de haver uma opacidade quanto aos processos de moderação realizados, o que torna as políticas combativas ineficazes.

Palavras-chave: efeito borboleta; dinâmicas complexas; discursos de ódio; moderação; redes sociais; moderação de conteúdo.

HATE SPEECH AND THE BUTTERFLY EFFECT ON ONLINE BEHAVIOR: SMALL POSTS, BIG CONSEQUENCES

ABSTRACT

The complex and interconnected dynamics of social networks have contributed to the creation of environments where hate speech and polarization spread rapidly. Through small interactions with hateful content, individuals gradually become engaged with radical material, leading to a cascading effect of violent or intolerant behavior. In this context, polarization on social media often begins with minor misunderstanding or disagreements which, amplified by the butterfly effect and communication processes, result in intense conflicts and divisions between groups. This study questions the extent to which content moderation against hate speech are capable of containing the spread of hate on social networks? Using an interdisciplinary approach, the research investigates the dissemination of hateful content on social media, analyzing the phenomenon through the lens of the butterfly effect, in conjunction with studies in the fields of law. From the study, it is concluded that content moderation by digital platforms is insufficient to combat implicit hate speech, allowing its widespread dissemination across the network. Additionally, there is opacity regarding the moderation processes carried out, making the counteractive policies ineffective.

Keywords: butterfly effect; complex dynamics; content moderation; hate speech; moderation; social networks.

1 INTRODUÇÃO

Divulgar imagens, compartilhar vídeos, escutar músicas, publicar textos. Essas atividades são apenas uma parte do que o ambiente digital proporciona à população. O desenvolvimento das tecnologias de informação e comunicação avançou, e construiu espaços públicos onde é possível se relacionar com outras pessoas, dividindo suas crenças e compartilhando valores. Não há dúvidas, que a internet trouxe muitos benefícios para a sociedade, mas também não se questiona os malefícios que viabiliza.

Nesse contexto, pesquisas apontam para o aumento de denúncias de disseminação do ódio nas redes sociais, conforme o Observatório Nacional dos Direitos Humanos (Brasil, 2024) dentre os anos de 2017 a 2022, o crescimento de crimes de ódio praticados na internet chegou a 74 mil casos e 293,2 mil denúncias. Embora o Marco Civil da Internet, legislação que regula o uso da internet no país e a Lei Geral de Proteção de Dados, ferramenta de regulação do tratamento de dados, são legislações que estão em vigência desde 2014 e 2018 respectivamente, não foram suficientes para evitar e controlar o crescimento do ódio online.

Aliado a isso, as dinâmicas complexas e interconectadas das redes sociais parecem favorecer a disseminação do ódio online, seja pelo fornecimento de filtros, das possibilidades multimídias, ou mesmo da combinação de suas ferramentas com a criatividade e desempenho

dos usuários. Acontece que as redes sociais são campos robustos de tratamento de dados, que por trás do discurso de "experiência personalizada" passam a ditar quais conteúdos serão vistos e quais não serão.

Pequenas interações na rede revelam um perfil de consumo e comportamento do usuário, que pode acabar viralizando e expandindo o campo de contato de um conteúdo ideológico, causando um efeito cascata de comportamentos semelhantes, com potencial de elevar polarizações e trazer consequências sociais significativas. Diante disso, questiona-se em que medida a moderação de conteúdo contra o discurso de ódio conseguem conter a propagação de ódio nas redes sociais?

Para responder a esse questionamento, adotou-se uma abordagem interdisciplinar que articula o viés jurídico com conceito da física, a fim de investigar como ocorre a disseminação de conteúdos de ódio nas redes sociais. O foco principal foi analisar esse fenômeno sob a perspectiva do efeito borboleta, inspirado na obra "*Turbulent Mirror: An Illustrated Guide to Chaos Theory and the Science of Wholeness"*, de John Briggs e David Peat. O conceito foi utilizado como metáfora para capturar a natureza caótica e imprevisível das atividades digitais nas redes sociais, articulando-o com estudos nas áreas de comunicação e direito. Além disso, pretendeu-se examinar o papel da moderação das plataformas de redes sociais diante dos discursos de ódio e verificar a existência de políticas combativas. Utilizou-se do método monográfico e da técnica de pesquisa bibliográfica, baseados em obras de autores contemporâneos, ao mesmo tempo que se apoiou na "Pesquisa sobre regulação de plataformas digitais?" realizada pelo Comitê Gestor de Internet (CGI.br, 2023) para o desenvolvimento do trabalho.

Para completar a fundamentação teórica deste trabalho, foram utilizadas as obras de Zygmunt Bauman (2005, 2011), Manuel Castells (2012, 2021), Shoshana Zuboff (2019), Noam Chomsky (2001), Judith Butler (2021), Lawrence Lessig (2006) e Cass Sunstein (2017). Essas referências serviram como base para investigar como as interações digitais, que por mais triviais que sejam, influenciam a comunicação, moldam a sociedade e impactam a economia.

O trabalho está dividido em duas seções. A primeira aborda as interações sociais nas redes e sua relação com o efeito borboleta, enquanto a segunda buscou-se entender o papel das redes sociais na moderação de conteúdos permitidos nas plataformas, com base na análise de suas diretrizes e termos de uso, com foco na disseminação de discursos de ódio online. Além disso, investigou-se a existência de políticas combativas implementadas pelas próprias

plataformas.

De acordo com o Datareportal (2024), no Brasil há 187,9 milhões de usuários conectados na internet, dentre os quais 144 milhões são ativos nas redes sociais. Das redes sociais mais utilizadas pelos brasileiros estão o *Facebook* e o *Instagram*, ambos pertencentes a Meta *Platforms,* e o Tik Tok, pertencente à empresa chinesa Bytedance's. O exorbitante número de conectividade da população em conjunto com o aumento constante de episódios odientos, revela a necessidade de pesquisas que visam a compreensão não só de possíveis efeitos, mas também de como ações singelas podem gerar grandes consequências para a sociedade.

2 REDES SOCIAIS E PROCESSOS COMUNICATIVOS ONLINE: A APLICAÇÃO DO EFEITO BORBOLETA

O avanço das tecnologias de informação e comunicação transformou significativamente a maneira como as pessoas interagem e influenciam umas às outras. Nessas circunstâncias, as redes sociais foram grandes propulsoras, atuando como catalisadoras das mudanças na comunicação social, e fornecendo aos processos comunicativos um ambiente no qual pequenas ações têm a capacidade e o potencial de desencadear grandes efeitos.

As redes sociais são plataformas digitais que oferecem serviços online que permitem a conexão entre usuários e fornecedores de bens, serviços ou informações, por meio da utilização de tecnologias de comunicação digital (CGI.br, 2023). Também denominadas como plataformas transacionais, possuem a finalidade de facilitar transações entre diferentes grupos, sendo o seu elemento principal a conexão entre indivíduos. Essa conexão, "possibilita a coexistência e interdependência de múltiplos atores" em um ecossistema, o qual se desenvolve por meio de uma comunicação baseada em códigos e linguagem (CGI.br, 2023, p. 30).

De acordo com Romanini (2023) a comunicação depende de códigos e linguagens que produzem e comunicam sentido. Trata-se de um sistema complexo que envolve diferentes estados cognitivos, desde a percepção até a argumentação, que são partilhados socialmente. Os processos comunicativos ocorrem sob a coordenação de símbolos e signos que são mobilizados e articulados com o objetivo de gerar sentido à sociedade. Com as redes sociais, novas linguagens e códigos são gerados, em um processo de constante transformação, que influencia as escolhas que cada usuário faz, e permanece pendente de validação dos demais,

por meio das ferramentas de curtidas, comentários e compartilhamentos.

A concepção de que pequenas ações realizadas nas redes sociais podem gerar grandes consequências, especialmente em relação aos processos comunicativos e sociais, deriva do denominado "efeito borboleta", originado da teoria do caos. Para Briggs e Peat (1989) o comportamento caótico dos sistemas dinâmicos podem causar grandes efeitos imprevisíveis, no entanto, o caos não deve ser entendido como desordem, mas sim como padrões ocultos. Na interconectividade, o caos pode ser uma ferramenta para compreender as dinâmicas sociais, políticas e culturais, além de estar intimamente ligada à criatividade. Desse modo, é fundamental compreender que o funcionamento das redes sociais, sua dinamicidade e a rápida disseminação de informações formam um sistema complexo que considera as interações entre os indivíduos e o ambiente como um todo.

Esse conceito descreve a sensibilidade às condições iniciais em sistemas não lineares, em que pequenas alterações podem levar a grandes diferenças nos resultados a longo prazo. A sua definição deriva da ideia de que o bater das asas de uma borboleta em um lugar remoto poderia desencadear uma série de eventos que, eventualmente, influenciariam em uma escala muito maior. Por conseguinte, as suas implicações são significativas: é difícil prever o comportamento a longo prazo de sistemas caóticos, pois pequenas incertezas nas condições iniciais podem crescer exponencialmente, revelando que sistemas aparentemente simples podem ser altamente complexos e imprevisíveis (Briggs; Peat, 1989).

Assim, o novo ecossistema comunicacional, encontra uma ressonância clara com o paradigma do efeito borboleta, vez que a ideia de que um pequeno evento pode desencadear significativas consequências, se encaixa perfeitamente nas interações digitais. Isso porque, a dinâmica das redes sociais permite que ações, aparentemente insignificantes, como curtidas e compartilhamentos de conteúdos, possam gerar problemas sociais que ultrapassam barreiras temporais e geográficas, afetando a estrutura das relações humanas e da comunicação de massa (Sassi, 2025).

Os sistemas dinâmicos complexos, como o das plataformas de redes sociais, alteram as configurações sociais e a produção de sentido de uma comunidade, o que ocorre frente a quebra de hábitos, da introdução fortuita de novidade, da expansão da capacidade de reagir no espaço-tempo e da capacidade de produzir efeitos gerais, e se resume em um processo contínuo de transformação de parâmetros culturais:

> Esses movimentos de longa duração, alimentado pelas interações em tempo real das plataformas de redes sociais, são exemplos de como sistemas complexos podem entrar em deriva e caminhar para catástrofes a partir de

> reverberações nos extratos mais básicos que alteram suas relações com grande hipersensibilidade – um fenômeno popularmente conhecido como "efeito borboleta" (Romanini, 2023, p. 93).

Nesse cenário, a análise das redes sociais sob a ótica dos sistemas complexos pode ser enriquecida pela perspectiva de Zygmunt Baumann, que descreve a sociedade contemporânea como aquela marcada pela liquidez das relações. O autor apresenta uma metáfora sobre o estado contemporâneo das relações humanas e sociais, a liquidez. A partir da concepção de "Vida Líquida", refere que as interações, físicas ou virtuais, tornaram-se fluidas, instáveis e momentâneas, encaixando-se perfeitamente no modelo interativo das redes sociais. A volatilidade das conexões impacta as formas de comunicação, que passam a se desenvolver de forma efêmera, fragmentada e por vezes descomprometida, o que faz com que uma simples postagem ou comentário tenha um grande potencial de reverberação no ambiente digital podendo causar eventos imprevisíveis (Baumann, 2005).

O conceito de liquidez torna o paradigma do efeito borboleta especialmente relevante, já que, conforme os estudos de Baumann (2011), qualquer ação digital, por menor que seja, pode ser amplificada além de seu contexto inicial. Isso é exemplificado pelo fenômeno da viralização, em que conteúdos publicados por usuários comuns rapidamente alcançam uma grande audiência. Trata-se da rápida disseminação de informações, imagens, vídeos ou mensagens que se espalham e atingem um vasto público em um curto período de tempo.

Geralmente, a propagação viral é impulsionada pelo compartilhamento em massa, curtidas e interações. Essa rápida publicização ocorre devido aos seguintes fatores: alcance e velocidade com que as redes sociais possibilitam o compartilhamento instantâneo de conteúdos; o engajamento social; e a natureza interconectada das redes. Para Recuero (2017) o espalhamento de conteúdos está vinculado às interações constituídas nos meios online que tendem a permanecer no tempo, possibilitando o prolongamento de conversações e a sua recuperação em outros momentos, o que permite sua ocorrência em tempos diversos, e propicia a ampliação de possibilidades de manutenção e recuperação de conexões e valores sociais.

As redes sociais intensificam não só os processos de significação, como também a incerteza e a ansiedade, tornando as relações mais instáveis, e ao mesmo tempo, potencializando o impacto das ações humanas. A busca por visibilidade, consequentemente popularidade, tem causado impactos diretos na exposição da vida humana nas redes sociais, vez que o objetivo de promover conexão, identidade e interação afetou a qualidade e veracidade das informações que são partilhadas, contribuindo para o aumento da

desinformação[1].

Gerou-se uma sociedade de tabloide, na qual a criatividade e visibilidade andam juntas e produzem informação concisa e espetacularizada. Isso reflete as relações e interações sociais instáveis e passageiras, mas com capacidade de desencadear efeitos em larga escala, influenciando comportamentos, tendências e até movimentos sociais. A desinformação e o ódio disseminados nas redes sociais decorrem, muitas vezes, da desvalorização do outro, vez que "falar mal do outro é, indiretamente, falar bem de si e da pessoa para qual se retransmite a informação" (Roxo, 2016, p. 07), cooperando para que esse tipo de conteúdo seja ferramenta impulsionadora de popularidade por meio da criação de bolhas de ódio.

Desse modo, o IP.rec (CGI.br, 2023, p. 151) endossa a dualidade inerente ao avanço da Internet e seus impactos nas possibilidades de desenvolvimento social e democratização da comunicação e do conhecimento "se as redes sociais democratizaram o acesso à informação e ampliaram a voz de minorias sociais, também é possível observar que houve a intensificação de problemas, como desinformação, extremismos, discurso de ódio e incitação ao terrorismo".

Para Cass Sunstein (2017) a arquitetura das redes sociais é construída para promover valorização e formação de câmaras de eco, que funcionam exacerbando as divisões sociais e políticas dos usuários, no qual são isolados em bolhas ideológicas. Esse tipo de isolamento afeta diretamente a estrutura democrática, vez que o próprio conceito democrático requer o diálogo de diferentes perspectivas para a formação da sociedade. A amplificação dessas bolhas gera cisões na sociedade, limita o contato com o diferente e reforça a divisão ideológica, moldando a esfera pública digital de maneira prejudicial à democracia e desencadeando crises políticas e sociais, que utilizam o ódio e o diferente como argumento de poder.

A combinação da infodemia e o modelo de negócios das plataformas digitais transformou as redes sociais em um ambiente propício à disseminação de desinformação e conteúdos ilícitos, como discursos extremistas e de ódio. Isso ocorre por meio de sistemas algorítmicos que priorizam engajamento, aumentando a visibilidade de conteúdos nocivos que geram reações intensas. Segundo o CGI.br (2023), essa dinâmica, atrelada à coleta excessiva de dados, amplifica conteúdos extremos para manter usuários conectados, comprometendo os direitos à comunicação e à informação.

Por conseguinte, o ódio é uma estratégia de poder que move sentimentos e práticas

[1] A desinformação se trata da divulgação e compartilhamento de informações falsas ou enganosas com a intenção de enganar, manipular ou prejudicar o público.

negativas, e quando associado às formas de comunicação e às práticas de interação online, faz uso das ferramentas multimídias para salientar repetidamente códigos de comoção, pertencimento e segregação. Ocorre que a incitação a uma ação coletiva que propaga, escala e intensifica a repetição e o contágio, de forma inconsciente, por meio da imitação e reprodução, magnetiza crenças e desejos na rede. Na esteira da visibilidade, poder é significar, pertencer e liderar, enquanto nas redes sociais isso significa mobilizar o máximo de seguidores e interações, retendo uma comunidade na sua bolha influenciadora (Sassi; Rosa, 2024).

Os efeitos da indústria da desinformação são severos e relacionam-se com a violência em suas diversas dimensões, evidenciando o uso de discursos de ódio como estratégia para capturar e manter a atenção dos usuários. Esses discursos envolvem uma progressão de violações que, pautadas em agressividade, hostilidade e opressão, evoluem para extremismos discursivos (Sassi, 2025). Tal processo desumaniza seus alvos e generaliza seus destinatários, configurando uma estratégia de poder que consolida a intolerância e a exclusão de pessoas ou comunidades, exacerbando conflitos sociais e polarizações (Brasil, 2023a).

Os discursos de ódio são um fenômeno de ampla complexidade que variam conforme os contextos culturais, políticos e sociais, e são valorizados pela arquitetura algorítmica das redes sociais. Nesse sentido, a indústria da desinformação, ao empregar discursos de ódio como ferramenta estratégica, não apenas perpetua a violência em suas diversas formas, mas também intensifica a polarização e a desumanização social. Esses discursos evoluem para extremismos que deslegitimam indivíduos e comunidades, utilizando a opressão como um mecanismo de poder e controle, que ressalta a importância de se ter uma transparência na moderação de conteúdos nas plataformas digitais (Sassi, 2025).

Um exemplo dessa dinâmica pode ser observado no estudo etnográfico realizado na plataforma TikTok, que revelou a exposição de adolescentes a conteúdos de ódio de cunho neonazista. Na obra *Mídias Cruzadas e Discursos de Ódio Neonazistas*, Sassi (2025) demonstra que o algoritmo da plataforma rapidamente recomenda conteúdos multimodais a novos usuários, de modo que publicações que difundem a ideologia neonazista funcionam como vitrines de atração e captura de seguidores, direcionando-os posteriormente para outras plataformas, onde encontram mensagens ainda mais radicalizadas e excludentes.

Outrossim, Castells (2021) explora o impacto das transformações comunicativas, enfatizando o papel das redes na reconfiguração das estruturas de poder e nas formas de interação social. Para o autor, na sociedade em rede o poder se distribui de maneira

descentralizada, e as redes sociais são o principal meio em que esse poder pode ser manifestado e explorado. Nessa perspectiva, as redes sociais acabam por ser utilizadas como ferramenta de mobilização política e social, vez que nelas as fronteiras entre o local e o global são borradas, facilitando que pequenas ações possam ser amplificadas e disseminadas em questões de segundos.

Um dos exemplos apresentados pelo sociólogo, sobre como pequenas ações podem iniciar grandes movimentos são observados durante a Primavera Árabe, que inicia com protestos locais, e a partir do auxílio das redes sociais, expande sua causa, provocando transformações políticas em toda a sua região (Castells, 2012). Dessa forma, uma ação aparentemente insignificante no panorama global, pode ser amplificada através dos algoritmos e disseminada rapidamente, atraindo visibilidade e poder àquela demanda, o que demonstra que o ambiente digital é um espaço de interações imprevisíveis em que o pequeno/local pode rapidamente se tornar uma pauta global.

A ideia do efeito borboleta é materializada quando uma ação de protesto local inspira movimentos em diversas partes do mundo, gerando ondas de mudança social e política. Uma resistência local que se transforma em movimento global, demonstra o poder que a conectividade das redes possui. O núcleo desse poder está nas próprias interações dos usuários, que assumem uma postura participativa, escolhendo qual conteúdo, informação ou mídia querem compartilhar com seus amigos virtuais, qual formador de opinião se identificam, e a quem vão dar voz à narrativa.

Pela lógica performativa de Butler (2021), infere-se que a capacidade performativa do sujeito é intensificada pelas redes sociais, onde o ato de compartilhar opinião ou expressar uma ideia pode causar uma reação em cadeia. Por meio de pequenos atos discursivos é possível influenciar emoções e instigar ações de grandes audiências, uma vez que discursos se convertem em ações que produzem efeitos dentro e fora do ambiente digital. Uma mensagem, um vídeo ou uma imagem viral causa repercussões positivas ou negativas nos espectadores, que além de interagir com esse conteúdo também poderão replicá-lo. Dessa forma, uma única expressão performativa pode afetar indivíduos e grupos de maneiras distintas e imprevisíveis.

As redes sociais não são apenas espaços de interação dos usuários, são ecossistemas complexos que ensejam uma análise crítica do seu impacto no comportamento e nas interações digitais dos usuários. Isso porque, as plataformas digitais não são apenas veículo de informações, a sua infraestrutura molda a produção, a circulação e a aceitação de discursos, um exemplo disso é o fomento a discursos de ódio, principalmente, em virtude da polarização

de ideologias políticas (Mercuri; Lima-Lopes, 2020).

Segundo a Safernet (2024), o "discurso de ódio nas redes é usado como uma plataforma política para engajar a audiência, dar notoriedade ao emissor e assim trazer mais votos". Consequentemente, a internet tornou-se campo fértil para disseminação de discursos de ódio, principalmente em períodos eleitorais, em virtude da polarização ideológica. Nesse sentido, Noam Chomsky (2001) denuncia que a relação entre comunicação e manipulação das redes sociais, a qual se utiliza de ferramentas de controle e propaganda política, acaba por impulsionar discursos de ódio.

De acordo com o linguista e filósofo, a propagação de desinformação ou a amplificação de discursos políticos específicos, alerta para o mau uso das redes sociais como ferramenta de controle e manipulação da opinião pública. Esse uso é frequentemente escolhido para compartilhar notícia falsa, tendenciosa ou manipulada, já que seu objetivo é justamente moldar a percepção pública e influenciar comportamentos de massa por meio de narrativas cuidadosamente construídas. Logo, as redes sociais, ao facilitarem a propagação desse conteúdo, permitem que ações singulares levem a consequências políticas, demonstrando mais uma vez a presença do efeito borboleta nas atividades comunicacionais e sociais digitais.

Partindo das diferentes abordagens teóricas, chega-se à compreensão de que as redes sociais não são neutras, mas com base nas atividades interativas dos usuários, escolhem e determinam o tipo de conteúdo que cada usuário receberá. E ainda que justifiquem que estão promovendo a personalização dos serviços, na realidade há outros objetivos velados que, levados a cabo, contribuem para a formação de bolhas e a quebra do diálogo democrático.

Essa constatação aponta para a necessidade de análise crítica da forma como a moderação de conteúdo é realizada, sobretudo pelo seu poder de influenciar o acesso dos usuários a conteúdos e informações. Nessa empreitada acadêmica, deve-se considerar as diretrizes e termos de uso divulgados pelas redes sociais, com olhar mais acurado sobre a propagação dos discursos de ódio, pois é sabido que essa estratégia reforça o modelo de negócio das plataformas, tema que será desenvolvido no próximo item.

3 MODERAÇÃO DO ÓDIO? DIRETRIZES, TERMOS DE USO E POLÍTICAS COMBATIVAS

Ao discutir os processos de interação pela lógica dos sistemas complexos das redes

sociais, deve-se prestar atenção à lógica comportamental na qual as plataformas inserem seus usuários. Basicamente, os laços sociais construídos digitalmente requerem um comportamento ativo com os demais, seja por meio da publicação ou compartilhamento de conteúdos, seja pela validação por meio de curtidas ou reações.

As formas interativas dos usuários geram uma perfilarização de suas práticas, e isso determina que setor de conteúdos será recomendado para aquele indivíduo. Especialmente porque, as redes sociais se tornaram a principal fonte informativa da população, nelas é possível se informar sobre culinária, esportes, política, moda, viagens, dentre outros temas que despertem o consumo dos usuários.

Nessa senda, Shoshana Zuboff (2019) argumenta que as interações online, embora possam parecer triviais, são registradas, processadas e utilizadas para moldar o comportamento dos usuários. A autora denomina de "capitalismo de vigilância" o modo com que são utilizadas ferramentas de controle e exploração, com viés econômico, para moldar a percepção e o comportamento dos indivíduos no ambiente digital. Esse controle sistêmico, que ocorre de maneira invisível, é direcionado e explorado pelas *Big Techs*, por meio de dados gerados por ações triviais, e utilizados para influenciar comportamentos em larga escala.

O *Instagram* e o *Facebook*, ambos pertencentes à *Meta Platforms*, apresentam funcionalidades similares, como a possibilidade de publicar imagens e vídeos no feed, além de compartilhar, curtir e comentar nas postagens de amigos. Ambas as plataformas também permitem a publicação de vídeos curtos (stories), que ficam disponíveis por 24 horas no perfil dos usuários. O *Facebook*, a plataforma mais antiga do grupo, oferece, além dessas funcionalidades, a possibilidade de compartilhar fotos, vídeos, *links* e publicações em um feed de notícias, criar grupos e páginas para empresas ou interesses específicos, além de contar com um sistema de mensagens instantâneas (Meta, 2024a).

O *Instagram* surgiu como uma rede social para o compartilhamento de registros fotográficos, que foi amplamente adotada pelos jovens, e hoje possibilita que seus usuários compartilhem vídeos em modelos diferentes e também possui salas de bate-papo. A plataforma incentiva a interação entre os usuários por meio de curtidas, comentários, mensagens diretas e a opção de seguir perfis de interesse. Ademais, constitui espaço fundamental para influenciadores digitais, marcas e negócios, já que permite a criação de conteúdo patrocinado e anúncios direcionados (Meta, 2024b).

Dentre as plataformas mais utilizadas atualmente, o Tik Tok é a mais recente, gerida

pela empresa chinesa Byte Dance. Essa rede social funciona com a publicação de vídeos curtos, muitas vezes com música, dublagens, desafios e efeitos visuais criativos. Possui um algoritmo altamente personalizado que recomenda vídeos com base nos interesses e comportamento do usuário. Amplamente acessada por crianças e adolescentes, é reconhecida por lançar tendências e conteúdos virais. Tal como o *Instagram*, atualmente é um espaço fundamental para influenciadores, criadores de conteúdo e marcas, vez que podem alcançar grandes audiências de forma orgânica ou por meio de anúncios pagos (Tik Tok, 2024).

Os termos de uso dessas plataformas pouco demonstram como a coleta de dados e interesses dos usuários é explorada para sustentar o modelo de negócio dessas plataformas. A personalização da experiência, junto com a possibilidade de conexão com pessoas e organizações com interesses semelhantes, é o que impulsiona o funcionamento dessas redes. A abordagem amigável com que expõem a coleta e utilização de dados torna essas plataformas atrativas e incentiva a participação online.

Veja que o *Facebook* assim como o *Instagram* vende a ideia de proporcionar uma experiência personalizada por meio do direcionamento de publicações, stories, eventos, anúncios e outros conteúdos com base nos interesses que o usuário demonstra e os dados das conexões, das escolhas e configurações que seleciona e compartilha dentro e fora da plataforma. O foco da publicidade do seu modelo de negócio é conectar e permitir a expressão do que é e com o que é importante para o indivíduo, o que pode ser observado no início dos seus termos de uso:

> **Proporcionar uma experiência personalizada para você:** sua experiência no Facebook não se compara à de mais ninguém. Isso inclui desde as publicações, os stories, os eventos, os anúncios e outros conteúdos que você vê no Feed de Notícias do Facebook ou na nossa plataforma de vídeo até as Páginas do Facebook que você segue e outros recursos que pode usar, como o Facebook Marketplace e a pesquisa. Por exemplo, usamos os dados sobre as conexões que você faz, as escolhas e as configurações que seleciona e o que compartilha e faz dentro e fora dos nossos Produtos para personalizar a sua experiência. **Conectar você com as pessoas e organizações com as quais se importa:** ajudamos você a encontrar e se conectar com pessoas, grupos, empresas, organizações e outras entidades do seu interesse nos Produtos da Meta que você usa. Usamos dados para fazer sugestões para você e outras pessoas - por exemplo, grupos para participar, eventos para participar, Páginas do Facebook para seguir ou enviar uma mensagem, programas para assistir e pessoas de quem você pode querer se tornar amigo. Laços mais fortes criam comunidades melhores, e acreditamos que os nossos serviços são mais úteis quando as pessoas estão conectadas a pessoas, grupos e organizações que sejam relevantes para elas. **Permitir que você se expresse e fale sobre o que é importante para você:** há muitas maneiras de se expressar no Facebook para comunicar aos amigos, familiares e outras pessoas o que é importante para você. Por exemplo, é possível compartilhar atualizações de status, fotos, vídeos e stories nos Produtos da Meta (de forma consistente com as suas configurações), enviar mensagens para um amigo ou diversas pessoas ou fazer ligações de voz ou de vídeo com eles, criar eventos ou

> grupos ou adicionar conteúdo ao perfil e mostrar análises de como as outras pessoas interagem com o conteúdo que você cria. Também desenvolvemos e continuamos explorando novas formas de usar a tecnologia, como a realidade aumentada e o vídeo 360. Assim, as pessoas podem criar e compartilhar conteúdo mais expressivo e envolvente nos Produtos da Meta. Ajudar você a descobrir conteúdo, produtos e serviços que podem ser do seu interesse: Nós mostramos anúncios personalizados, ofertas e outros tipos de conteúdo patrocinado ou comercial para você descobrir, com mais facilidade, conteúdo, produtos e serviços que várias empresas e organizações que usam o Facebook e outros Produtos da Meta oferecem. A seção 2 abaixo explica isso com mais detalhes (Facebook, 2024a).

Já o *Instagram*, de uma forma mais sucinta que o *Facebook*, divulga o oferecimento de oportunidades personalizadas de criar, conectar, comunicar, descobrir e compartilhar na sua plataforma. Ressalta-se que em seus termos de uso consta o reconhecimento à diferença de pessoas, que serve de argumento para oferecer variados tipos de contas de usuários e recursos que ajudariam a ampliar a presença e a comunicação com outras pessoas, dentro e fora do seu ambiente:

> **Oferecer oportunidades personalizadas de criar, conectar, comunicar, descobrir e compartilhar.** As pessoas são diferentes. Por isso, oferecemos diferentes tipos de contas e recursos para ajudar você a criar, compartilhar, ampliar sua presença e comunicar-se com pessoas dentro e fora do Instagram. Queremos também fortalecer seus relacionamentos por meio de experiências compartilhadas realmente importantes para você. Por isso, desenvolvemos sistemas que tentam entender com quem e com o que você e as outras pessoas se importam, e usamos essas informações para ajudar você a criar, encontrar, compartilhar e participar de experiências importantes. Parte do que fazemos é destacar conteúdo, recursos, ofertas e contas que possam ser do seu interesse e oferecer formas para você experimentar o Instagram, com base no que você e as outras pessoas fazem dentro e fora da plataforma. **Conectar você com marcas, produtos e serviços de maneiras importantes para você.** Usamos dados do Instagram e de outros Produtos das Empresas da Meta, bem como de parceiros, para veicular anúncios, ofertas e outros conteúdos patrocinados que acreditamos ser do seu interesse. Além disso, tentamos fazer com que esse conteúdo seja tão relevante quanto todas suas outras experiências no Instagram (Instagram, 2024a).

Esses termos revelam que, para as plataformas, é essencial oferecer diferentes tipos de contas e recursos que incentivem a criação, o compartilhamento e a ampliação da presença online, além de fortalecer os relacionamentos por meio de experiências compartilhadas. O que deixa explícito que o modelo de negócios dessas redes é baseado na coleta e tratamento de dados, que são usados para prever comportamentos e determinar o conteúdo direcionado a cada usuário.

Sob essa ótica, o efeito borboleta se manifesta de maneira ainda mais complexa. Veja-se: à medida que pequenos gestos, como clicar em um anúncio ou seguir uma página, geram dados que alimentam os algoritmos, as *Big Techs* detêm o poder para influenciar ações futuras. Nesse sentido, o *influencer digital* Felca em denúncia ao processo de adultização de

crianças e adolescentes, demonstrou como os algoritmos de recomendação das plataformas atuam com base no interesse em ver do usuário pelo seu comportamento online (Youtube, 2025).

A exposição evidenciou que, em especial nas plataformas de redes sociais baseadas em vídeos verticais, fatores como curtidas, compartilhamentos e tempo de visualização contribuem significativamente para a radicalização e para a exposição dos usuários a conteúdos nocivos. Quanto maior o consumo de determinado tipo de material, mais intensamente o algoritmo o recomendará, aprofundando o ciclo de engajamento e reforçando padrões de consumo prejudiciais. Tal processo ilustra a dinâmica com que as redes sociais potencializam pequenas atividades e as transformam em mercadorias valiosas.

Diante disso, Lawrence Lessig, na obra "*Code and Other Laws of Cyberspace*" (2006), revela que o ambiente digital, regido por códigos e leis, podem influenciar a forma como o efeito borboleta opera nas redes. Isso porque, seus códigos são desenvolvidos por corporações, e atuam como ferramenta regulatória que rege os comportamentos dos usuários. São mecanismos algorítmicos que influenciam o fluxo informacional e o recebimento de recomendações de conteúdo. Esses algoritmos desenvolvem ciclos de retroalimentação que acabam reforçando as bolhas ideológicas e construindo ecossistemas comunicativos personalizados, proporcionando um agrupamento ideológico.

Essas regulações e restrições pelo sistema de códigos impostos pelas plataformas digitais, podem mediar, potencializar ou suprimir ações e conteúdos gerados nesse meio. No entanto, o modelo de negócio baseado em engajamento incentiva que conteúdos polarizadores e odientos sejam recomendados e disseminados pelo ambiente digital. Isso ocorre em virtude do apelo emocional com que esse tipo de conteúdo é elaborado, o que mobiliza vários sentidos dos leitores.

Nessa senda, a moderação de conteúdo deveria garantir um fluxo informacional seguro e saudável online, vez que os sistemas de moderação têm por objetivo controlar o que é permitido nas plataformas, e é construído com base nas diretrizes e políticas internas. Para Zuboff (2019), a moderação de conteúdo, para além de garantir um ambiente seguro e saudável, desempenha papel crucial no gerenciamento do comportamento online.

A moderação de conteúdo serve para identificar conteúdos que violem os termos de uso, como discursos de ódio, e deve observar as especificidades locais, como a legislação e os padrões culturais. De acordo com Ribeiro e Favero (2024, p. 266) "a moderação é um mecanismo de governança que estrutura a participação em uma comunidade para facilitar a

cooperação e prevenir abusos".

E não poderia ser diferente, pois se em uma sociedade existem regras e normas básicas para a convivência social, o ambiente digital, parte indissociável da vida humana atual, também deve estabelecer limites ao exercício de direitos online. Isso porque, o Estado Brasileiro tem como objetivo fundamental a construção de uma sociedade livre, justa e solidária, na qual a participação social é garantida a todos os seus cidadãos (Brasil, 1988). Nesse sentido, pode-se compreender que a moderação de conteúdos pelas redes sociais visa à elaboração de regras para garantir o debate público e a pluralidade social.

De forma sucinta, a moderação nada mais é do que a atividade de controle realizada pela plataforma de rede social, na qual determina o que se pode ver e permanecer no seu ambiente digital (Sassi, 2025). Essas regras variam de acordo com a plataforma social que o usuário participa e, conforme levantamento realizado sobre as redes sociais mais utilizadas por brasileiros, o *Instagram*, o *Facebook* e o *Tik Tok* atualmente constituem as mais acessadas pela população (Dourado, 2024a).

Analisadas os termos de uso das três plataformas, identificou-se que apenas o *Tik Tok* apresenta uma singela disposição na seção sobre acesso a qual refere que os usuários não podem publicar:

> [...] qualquer material deliberadamente criado com o intuito de provocar ou antagonizar as pessoas, especialmente por meio de implicâncias e *bullying*, ou com a intenção de assediar, prejudicar, ferir, assustar, angustiar, constranger ou incomodar as pessoas; qualquer material contendo qualquer tipo de ameaça, inclusive ameaças de violência física; qualquer material racista ou discriminatório, incluindo a discriminação de pessoas por conta de sua raça, religião, idade, sexo, deficiência física ou mental ou sexualidade [grifo das autoras] (Tik Tok, 2024c).

O discurso de ódio é uma manifestação que ataca as características pessoais e inerentes de um grupo, tais como a raça, religião, idade, sexo e nacionalidade. Essas mensagens podem ser definidas como "qualquer tipo de comunicação falada ou escrita ou comportamento que ataque ou use linguagem pejorativa ou discriminatória com referência a uma pessoa ou grupo com base em sua religião, etnia, nacionalidade, raça, cor, descendência, gênero ou outro fator de identidade" (MDHC, 2023, p. 22).

Esse tipo de manifestação visa avaliar negativamente aqueles indivíduos, gerando a sua exclusão social, que deriva de ideais preconceituosos e fomenta atos discriminatórios, que podem incitar práticas violentas para além do ambiente digital[2]. A propagação de discursos de

[2] Um exemplo disso é o crescimento de ataques violentos nas escolas, que conforme indicou o relatório "Ataques de violência extrema em escolas no Brasil", os autores desses atos participavam de comunidade odientas que

ódio objetiva inferiorizar aqueles e aquelas a quem é direcionado, desvalorizando-os como sujeito de direitos na sociedade.

Ademais, a moderação de conteúdo também observa as diretrizes de comunidade das plataformas. Nas diretrizes, as plataformas estabelecem quais conteúdos são permitidos ou proibidos de divulgar. Segundo elas, o seu ambiente é um "reflexo da nossa comunidade de culturas, idades e crenças diversificadas" (Instagram, 2024b), por isso é importante "garantir que todas as vozes sejam valorizadas" por meio da criação de padrões "que incluem diferentes pontos de vista e crenças" (Facebook, 2024b), proporcionando ao usuário "criar uma experiência acolhedora, segura e divertida" (Tik Tok, 2024b). Tais fragmentos evidenciam que a moderação de conteúdos tem pontos em comum em uma e outra plataforma.

Tal posicionamento tenta evidenciar para o público o compromisso com a sociedade e com os usuários, comprometendo-se a promover um ambiente seguro, saudável e inclusivo a todas as pessoas. Nesse ambiente, segundo anunciado, a disseminação de valores, crenças e culturas deve respeitar o direito do outro, não causando danos a sua existência ou discriminando-o por suas características e escolhas. Segundo os posicionamentos assumidos publicamente pelas plataformas, não necessariamente refletidos em suas práticas, o debate entre liberdade de expressão *versus* discurso de ódio deve ser limitado a partir do princípio da dignidade humana.

A análise dos documentos evidencia um alinhamento com o disposto na Constituição Federal de 1988, pois a liberdade de expressão não constitui um direito absoluto, vez que o ordenamento jurídico brasileiro estabelece o princípio da dignidade da pessoa humana como delimitador da liberdade de expressão. A dignidade da pessoa humana visa a sua garantia como membro da sociedade, e é inerente a todos os indivíduos da espécie humana, apenas por serem pessoa sem distinção ou discriminação (Sarmento, 2016).

O discurso de ódio ataca diretamente a dignidade da pessoa humana, uma vez que é um "ataque direto a pessoas, e não a conceitos e instituições, baseado no que chamamos de características protegidas: raça, etnia, nacionalidade, deficiência, religião, casta, orientação sexual, sexo, identidade de gênero e doença grave" (Meta, 2024c). Ou como define o Tik Tok (2024d):

> O discurso e comportamento de ódio incluem atacar, ameaçar, desumanizar ou degradar um indivíduo ou grupo com base em seus atributos protegidos. Atributos protegidos significam características pessoais com as quais você nasceu, que são imutáveis ou que causariam danos psicológicos graves se você fosse forçado a

disseminavam o neonazismo como uma cultura, um modo de vida (Vinha, 2023).

> mudá-las ou atacado por causa delas. Isso inclui raça, etnia, nacionalidade, religião, tribo, classe social, orientação sexual, sexo, gênero, identidade de gênero, doença grave, deficiência e status de imigração.

São discursos e comportamentos incompatíveis com a convivência saudável e harmoniosa[3] de uma sociedade. Em decorrência do seu caráter nocivo, as mensagens de ódio são proibidas (pelo menos em seus termos de uso) pelas plataformas. No entanto, o blog alemão *Belltower News* (2023) denuncia que os padrões estabelecidos pelas plataformas digitais não são suficientes para coibir o discurso de ódio online. O blog argumenta que, embora as plataformas ofereçam conteúdos e ferramentas multimídia projetados para proporcionar uma experiência acolhedora, segura, divertida e personalizada, essas mesmas ferramentas também facilitam a divulgação do ódio de forma implícita e velada entre os usuários:

> Utiliza imagens comprometedoras, emoções manipuladoras e todas as técnicas de desinformação que existem: com relatos falsos de locais, mentiras completas, imagens e vídeos antigos, material fora de contexto, sequências de jogos de computador e animações de IA que passam por realidade, conclusões falsas e interpretações equivocadas.[4] [tradução das autoras] (Belltower News, 2023).

Dessa forma, as diretrizes das plataformas visam o combate ao ódio explícito, apesar de atualmente estar-se-á diante de um ódio onipresente que se manifesta por meio de expressões de humor, da utilização da emoção, e de contexto históricos e culturais distorcidos, além de contar com a desinformação para trazer um tom de veracidade a fatos e dados odientos, normalizando-os entre os usuários online.

Apesar de afirmarem adotar medidas para combater os discursos de ódio em seus ambientes, não há evidências concretas de que essas ações estejam sendo efetivas. Sabe-se que é mais fácil lidar com discursos explícitos, mas existem estratégias que manipulam as mensagens para "camuflar" o ódio, dificultando sua identificação e combate pelas plataformas, tais como o emprego de uma comunicação multimodalizada, na qual o algoritmo de moderação se torna ineficaz frente a utilização de gifs e códigos empregados para driblar

[3] Nós, representantes do povo brasileiro, reunidos em Assembleia Nacional Constituinte para instituir um Estado Democrático, destinado a assegurar o exercício dos direitos sociais e individuais, a liberdade, a segurança, o bem-estar, o desenvolvimento, a igualdade e a justiça como valores supremos de uma sociedade fraterna, pluralista e sem preconceitos, fundada na harmonia social e comprometida, na ordem interna e internacional, com a solução pacífica das controvérsias, promulgamos, sob a proteção de Deus, a seguinte CONSTITUIÇÃO DA REPÚBLICA FEDERATIVA DO BRASIL (Brasil, 1988).

[4] Er wird geführt mit belastenden Bildern, manipulativen Emotionen und jeder Desinformationstechnik, die es gibt: Mit falschen Vor-Ort-Schilderungen, kompletten Lügen, alten Bildern und Videos, aus dem Kontext gerissenem Material, Computerspielsequenzen und KI-Animationen, die als Realität durchgehen sollen, falschen Schlüssen und Einordnungen

as diretrizes.

Dessa forma, embora essas redes contenham diretrizes sobre proteção de grupos vulneráveis e proibição de violência, elas diferem significativamente na aplicação dessas regras. Isso é especialmente evidente em relação à abordagem preventiva, às penalidades e à flexibilidade para considerar o contexto. O Tik Tok se destaca por sua postura mais incisiva e pelo uso avançado de inteligência artificial na moderação, enquanto o *Instagram* e o *Facebook* tendem a ser mais abertos a avaliações contextuais e à aplicação gradual de sanções.

Ressalta-se que apesar do exercício de moderação de conteúdo ser orientado pelos termos de uso e diretrizes, observou-se que em nenhum desses documentos explica a maneira como realmente é feita essa moderação. De acordo com o relatório "reclamações sobre o procedimento de moderação de conteúdo em redes sociais: o que pensam os usuários", 54,34% dos usuários de redes sociais queixaram-se da opacidade no processo de moderação, especialmente em relação à (IRIS, 2024):

- fundamentação inadequada de decisões de moderação (52,46%);
- contestação de decisão de moderação não respondida (22,54%);
- falta de notificação ou aviso sobre decisão de moderação (9,02%);
- ausência de ferramentas para contestar a decisão de moderação (7,38%);
- design da plataforma inacessível em relação aos mecanismos de revisão de decisão de moderação (4,51%);
- outros (4,1%)

A insatisfação dos usuários revela também a insuficiência de políticas combativas voltadas para a moderação de conteúdo das próprias plataformas, especialmente em relação ao discurso de ódio. As ferramentas de denúncia e revisão de conteúdos, bem como a remoção e possíveis restrições de contas não são feitas de maneira clara e transparente, operando de forma opaca impedindo aos usuários direito de contestação e compreensão sobre o conteúdo publicado.

Assim, a análise das interações nas redes sociais, baseadas nos sistemas complexos que moldam o comportamento dos usuários, revela uma dinâmica poderosa e pouco transparente. As plataformas digitais, ao personalizarem a experiência dos usuários com base em algoritmos que monitoram e processam seus comportamentos, acabam moldando a percepção e o engajamento de maneira imperceptível, mas significativa (Sassi, 2025). Como argumenta Shoshana Zuboff (2019), trata-se de um "capitalismo de vigilância", em que os dados coletados em interações triviais são utilizados para influenciar comportamentos e, em muitos casos, reforçar polarizações e bolhas ideológicas.

O processo de moderação de conteúdo, especialmente voltado ao combate ao discurso de ódio, deveria ser um mecanismo essencial para garantir um ambiente digital seguro. No entanto, as plataformas enfrentam críticas devido à opacidade de seus processos e à insuficiência de medidas combativas eficazes. A falta de clareza nas decisões de moderação e a incapacidade de fornecer *feedback* adequado aos usuários alimentam a desconfiança e questionam a real eficácia dessas políticas.

A análise do uso e dos termos das plataformas sugere que, apesar de sustentarem que realizam esforços para criar um ambiente acolhedor e seguro, o modelo de negócios baseado no engajamento acaba incentivando a disseminação de conteúdos polarizadores e emocionais, como o discurso de ódio. Esse conteúdo, muitas vezes, é difundido de maneira implícita, utilizando desinformação e apelos emocionais para normalizá-lo. Desse modo, conclui-se que, embora as plataformas afirmem adotar postura formal contra o discurso de ódio, suas medidas de moderação não são suficientemente transparentes ou rigorosas para enfrentar a complexidade desse tipo de conteúdo no ambiente digital.

4 CONSIDERAÇÕES FINAIS

A análise das interações nas redes sociais, sob a lógica dos sistemas complexos, revela uma dinâmica poderosa e pouco transparente, na qual os algoritmos moldam o comportamento dos usuários de forma imperceptível, mas significativa. As plataformas digitais personalizam as experiências com base em dados coletados de interações triviais, o que influencia a percepção dos usuários e reforça polarizações e bolhas ideológicas.

Diante disso, pequenas ações e publicações inocentes são consideradas para a atividade de recomendação dos algoritmos, que categorizam os conteúdos inserindo-os em bolhas informáticas, com grande potencial de radicalização dos usuários derivados do consumo de conteúdos prejudiciais. Esse consumo de conteúdos odientos tem gerado grandes efeitos, tanto na comunidade digital quanto fora dela, levando ao cometimento de atos ilícitos na sociedade. Tudo isso, em razão de um "capitalismo de vigilância", no qual os dados derivados das pequenas ações digitais são utilizados para moldar comportamentos, muitas vezes em benefício dos interesses comerciais das grandes empresas de tecnologia, que se aproveitam de conteúdos polarizadores para alavancar sua presença no cotidiano da população.

Nesse contexto, o processo de moderação de conteúdo, especialmente no combate ao discurso de ódio, deveria ser um mecanismo essencial para garantir um ambiente digital seguro e saudável. No entanto, as plataformas são frequentemente criticadas pela opacidade de seus processos e pela insuficiência de medidas eficazes.

A falta de clareza nas decisões de moderação, a ausência de *feedback* adequado aos usuários e a dificuldade em contestar essas decisões alimentam a desconfiança e questionam a real eficácia dessas políticas. Apesar dos esforços formais para criar um ambiente acolhedor, o modelo de negócios baseado no engajamento acaba incentivando a disseminação de conteúdos polarizadores e emocionais, como o discurso de ódio, muitas vezes difundido de maneira implícita por meio de desinformação e apelos emocionais.

Dessa forma, embora as plataformas sustentem, em seus termos de serviços, que adotam medidas para combater o discurso de ódio, essas ações não parecem suficientemente transparentes ou rigorosas para enfrentar a complexidade do ambiente digital. Para que a moderação de conteúdo seja efetiva e o ambiente online seja realmente inclusivo e seguro, é necessário um maior comprometimento das plataformas, com políticas mais claras e mecanismos de controle mais eficientes, pois do contrário seguirão servindo como poderosos instrumentos que auxiliam na propagação do ódio na forma do efeito borboleta.

REFERÊNCIAS

BAUMAN, Zygmunt. **Vida Líquida.** Rio de Janeiro: Zahar, 2005.

BAUMAN, Zygmunt. **44 Cartas do Mundo Líquido Moderno.** Rio de Janeiro: Zahar, 2011.

BRASIL. **Constituição da República Federativa do Brasil de 1988**. Brasília, DF. Disponível em: http://www.planalto.gov.br/ccivil_03/constituicao/ constituicao.htm. Acesso em: 09 set. 2024.

BRASIL. **Incitação à violência contra a vida na internet lidera violações de direitos humanos com mais de 76 mil casos em cinco anos, aponta Observa DH**. Disponível em: https://www.gov.br/mdh/pt-br/assuntos/noticias/2024/janeiro/incitacao-a-violencia-contra-a-vida-na-internet-lidera-violacoes-de-direitos-humanos-com-mais-de-76-mil-casos-em-cinco-anos-aponta-observadh#:~:text=Os%20crimes%20de%20%C3%B3dio%20na,Crimes%20Cibern%C3%A9ticos%2C%20da%20organiza%C3%A7%C3%A3o%20SaferNet. Acesso em: 02 set. 2024.

BELLTOWER NEWS. **Argumente gegen das Schweigen.** Disponível em: https://www.belltower.news/newsletter-editorial-argumente-gegen-das-schweigen-154039/. Acesso em: 03 set. 2024.

BRIGGS, John; PEAT, F. David. **Turbulent Mirror:** An Illustrated Guide to Chaos Theory and the Science of Wholeness. New York: Harper & Row, 1989.

BUTLER, Judith. **Discurso de ódio:** uma política do performativo. Tradução de Roberta Fabbri Viscardi. São Paulo: Editora Unesp Digital, 2021.

CASTELLS, Manuel. **A Sociedade em Rede.** São Paulo: Paz e Terra, 2021.

CASTELLS, Manuel. **Redes de Indignação e Esperança**. Rio de Janeiro: Zahar, 2012.

CGI.br. **Sistematização das Contribuições à Consulta sobre Regulação de Plataformas Digitais.** Textos de Juliano Cappi e Juliana Oms. São Paulo: Núcleo de Informação e Coordenação do Ponto BR, 2023. Disponível em: https://cgi.br/media/docs/publicacoes/1/20240227162808/ sistematizacao_consulta_regulacao_plataformas.pdf. Acesso em: 01 nov. 2024.

CHOMSKY, Noam. **Mídia:** Propaganda Política e Manipulação. São Paulo: Martins Fontes, 2001.

DATAREPORTAL. **Digital 2024:** Brasil. Disponível em: https://datareportal.com/reports/digital-2024-brazil. Acesso em: 02 set. 2024.

DOURADO, Bruna. **Ranking:** as redes sociais mais usadas no Brasil e no mundo em 2023, com insights, ferramentas e materiais. Disponível em: https://www.rdstation.com/blog/marketing/redes-sociais-mais-usadas-no-brasil/. Acesso em: 12 set. 2024.

FACEBOOK. **Termos de Serviço** [2024a]. Disponível em: https://pt-br.facebook.com/terms. Acesso em: 26 set. 2024.

INSTAGRAM. **Termos de Uso** [2024a]. Disponível em: https://pt-br.facebook.com/help/instagram/581066165581870. Acesso em: 26 set. 2024.

IRIS. **Reclamações sobre o procedimento de moderação de conteúdo em redes sociais:** o que pensam os usuários. Disponível em: https://irisbh.com.br/wp-content/uploads/2024/09/Reclamacoes-sobre-o-procedimento-de-moderacao-de-conteudo-em-redes-sociais-o-que-pensam-os-usuarios-IRIS.pdf. Acesso em: 03 set. 2024.

LESSIG, Laurence. **Code and Other Laws of Cyberspace.** New York: Basic Books, 2006

MERCURI, Karen Tank; LIMA-LOPES, Rodrigo Esteves. Discurso de Ódio em Mídias Sociais como Estratégia de Persuasão Popular. **Trabalhos em Linguística Aplicada**, [*S. l.*], v. 59, n. 2, p. 1216–1238, 2020. Disponível em: https://www.scielo.br/j/tla/a/5nXh3dFwFnRvJ fJXXydJXMj/abstract/?lang=pt. Acesso em: 08 set. 2024.

META. **Facebook** [2024a]. Disponível em: https://about.meta.com/br/technologies/facebook-app/. Acesso em: 12 set. 2024.

META. **Instagram** [2024b]. Disponível em: https://about.meta.com/br/technologies/instagram/. Acesso em: 12 set. 2024.

META. **Discurso de ódio** [2024c]. Disponível em: https://transparency.meta.com/pt-br/policies/community-standards/hate-speech/. Acesso em: 03 set. 2024.

RECUERO, Raquel. **Introdução à análise de redes sociais**. Salvador: EDUFBA, 2017.

RIBEIRO, Liara Maria Knaack Farah; FAVERO, Sabrina. Moderação de Conteúdo nas Redes Sociais: uma análise a partir da Medida Provisória nº 1068/2021. **Academia de Direito**, [*S. l.*], v. 6, p. 258-282, 2024. Disponível em: https://www.periodicos.unc.br/index.php/acaddir/article/view/4373/2170. Acesso em: 10 set. 2024.

ROMANINI, Vinicius. A comunicação como semiose e os desafios da sociedade da informação. *In*: PEREZ, Clotilde *et al*. (org.). **PPGCOM USP 50 anos:** entre o passado e o futuro, nosso percurso. Disponível em: https://repositorio.usp.br/item/003151063. Acesso em: 05 set. 2024.

ROXO, Luciana. **A difusão de informações e o fenômeno da "viralização" das notícias falsas nas redes sociais**. Disponível em: https://entremeios.com.puc-rio.br/media/Luciana%20Roxo.pdf. Acesso em: 06 set. 2024.

SAFERNET. **Safernet aponta que discurso de ódio cresceu nas duas últimas eleições.** Disponível em: https://new.safernet.org.br/content/safernet-aponta-que-discurso-de-odio-cresceu-nas-duas-ultimas-eleicoes. Acesso em: 08 set. 2024.

SARMENTO, Daniel. **Dignidade da pessoa humana:** conteúdo, trajetórias e metodologia. Belo Horizonte: Fórum, 2016.

SASSI, Ana Carolina. **Mídias cruzadas e discurso de ódio neonazista**: a proteção jurídica de adolescentes no Brasil e na Alemanha. Cachoeirinha: Editora Fi, 2025. Disponível em: https://www.editorafi.org/ebook/c136-midias-cruzadas-discurso-odio-neonazista. Acesso em: 02 jul. 2025.

SASSI, Ana Carolina; ROSA, Isabela Quartieri da. Relações de Poder e Redes de Comunicação: o discurso de ódio protagonizando engajamento. **Revista de Ciências do Estado**, Belo Horizonte, v. 9, n. 1, 2024. Disponível em: https://periodicos.ufmg.br/index.php/revice/article/view/e51384/e51384. Acesso em: 05 set. 2024.

SUNSTEIN, Cass. **#Republic:** Divided Democracy in the Age of Social Media. Princeton University Press, 2017.

TIK TOK. **Sobre o Tik Tok.** [2024a]. Disponível em: https://www.tiktok.com/about?lang=pt-BR. Acesso em: 12 set. 2024.

TIK TOK. **Diretrizes da Comunidade**. [2024b]. Disponível em: https://www.tiktok.com/community-guidelines/pt/overview. Acesso em: 13 set. 2024.

TIK TOK. **Termos de Serviço**. [2024c]. Disponível em: https://www.tiktok.com/legal/page/row/terms-of-service/pt-BR. Acesso em: 25 set. 2024.

TIK TOK. **Combater o discurso de ódio e o comportamento de ódio** [2024d]. Disponível

em: https://www.tiktok.com/safety/pt-br/countering-hate. Acesso em: 03 set. 2024.

VINHA, Telma. **Ataques de violência extrema em escolas no Brasil**: causas e caminhos. Disponível em: https://d3e.com.br/noticias/pesquisa-de-telma-vinha-sobre-ataques-de-violencia-em-escolas-traz-explicacoes-e-recomendacoes/. Acesso em: 01 out. 2024.

YOUTUBE. **Adultização** [2025]. Disponível em: https://www.youtube.com/watch?v=FpsCzFGL1LE. Acesso em: 07 ago. 2025.

ZUBOFF, Shoshana. **A Era do Capitalismo de Vigilância**. Rio de Janeiro: Intrínseca, 2019.