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

Ana Carolina Sassi
16-11-2024

RESUMO

As dinâmicas complexas e interconectadas das redes sociais têm contribuído para a criação de ambientes onde o discurso de ódio e a polarização se expandem rapidamente. Através 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 através 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 e a criação de políticas 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 comunicação e 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 obsoleta a existência de políticas combativas.

Palavras-chave: Redes sociais; Discursos de ódio; Moderação; Efeito Borboleta

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, dividir suas crenças e compartilhar valores. Não há dúvidas, que a internet trouxe muitos benefícios para a sociedade, mas também não se questiona os malefícios que viabiliza.

Nesse contexto de benefícios e malefícios sociais, pesquisas apontam para o aumento de denúncias de disseminação do ódio nas redes sociais, conforme o Observatório Nacional dos Direitos Humanos (GOV.BR, 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 e a criação de políticas contra o discurso de ódio, conseguem conter a propagação de ódio nas redes sociais?

Para responder ao questionamento, adotou-se uma abordagem interdisciplinar que visou investigar 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 da disseminação de discursos de ódio no ambiente das redes sociais. 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, por mais triviais que sejam, influenciam a comunicação, moldam a sociedade e impactam a economia.

O trabalho está dividido em dois capítulos. O primeiro aborda as interações sociais nas redes e sua relação com o efeito borboleta. Na segunda parte, 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, através da utilização de tecnologias de comunicação digital (CGI.br, 2023, p.28-29). 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 (CGI.br, 2023, p.30), o qual se desenvolve por meio de uma comunicação baseada em códigos e linguagem.

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, o que 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 as condições iniciais em sistemas não lineares, onde 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 que sistemas aparentemente simples podem ser altamente complexos e imprevisíveis.

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.

Romanini (2023, p.93-94) argumenta que as plataformas de redes sociais, como sistemas dinâmicos complexos, transformam as configurações sociais e os parâmetros culturais das comunidades. Essa transformação ocorre por meio da quebra de hábitos, introdução de novidades e interações em tempo real, capazes de gerar reverberações que se expandem com grande sensibilidade, exemplificando o "efeito borboleta". Assim, essas

interações podem levar sistemas à deriva ou até a catástrofes, evidenciando como mudanças sutis nos extratos mais básicos têm impactos amplos e duradouros.

Bauman (2005) utiliza a metáfora da liquidez para descrever a instabilidade e efemeridade das relações contemporâneas, sejam físicas ou virtuais, características que se alinham ao modelo interativo das redes sociais. Nessas plataformas, as conexões fragmentadas favorecem uma comunicação volátil, onde simples postagens podem gerar consequências imprevisíveis. A ideia de liquidez reforça o paradigma do efeito borboleta, pois, conforme Bauman (2011), pequenas ações digitais podem ser amplificadas exponencialmente, como demonstrado no fenômeno da viralização, que permite que conteúdos atinjam rapidamente vastas audiências, redefinindo a escala e impacto das interações.

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, p.14) 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 - divulgação e compartilhamento de informações falsas ou enganosas com a intenção de enganar, manipular ou prejudicar o público.

Gerou-se uma sociedade de tabloide, onde 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 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, p.152), 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 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. Levando a incitação de uma ação coletiva que propaga, escala e intensifica a repetição e o contágio, de forma inconsciente, que por meio da imitação e reprodução magnetiza crenças e desejos na rede. Na esteira da visibilidade, poder é significar, pertencer e liderar, nas redes sociais isso quer dizer 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. 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.

Castells (2021) destaca como as redes sociais reconfiguram estruturas de poder e interação social, promovendo a descentralização e permitindo mobilizações políticas e sociais de impacto global. Na sociedade em rede, as fronteiras entre o local e o global se dissolvem, e pequenas ações podem ser amplificadas rapidamente, como demonstrado na Primavera Árabe, em que protestos locais ganharam escala regional por meio das redes (Castells, 2012). Esse cenário evidencia o poder dos algoritmos na amplificação de demandas, transformando o ambiente digital em um espaço dinâmico e imprevisível de interações globais.

Dessa maneira, 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.

Butler (2021) refere 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, p.1218).

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. Noam Chomsky (2001) denuncia a relação entre comunicação e manipulação das redes sociais, que se utilizam de ferramentas de controle e propaganda política para 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 através 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 analisar os processos de interação sob a lógica dos sistemas complexos das redes sociais, é fundamental considerar a dinâmica comportamental imposta pelas plataformas, que incentivam os usuários a interagir ativamente por meio de publicações, compartilhamentos e validações como curtidas e reações. Essas interações geram uma personalização de conteúdos recomendados, alinhada aos interesses do indivíduo. Dado que as redes sociais são a principal fonte informativa da população, abrangendo temas como política, culinária, moda, dentre outros, essa lógica intensifica o consumo e molda as práticas informativas e comportamentais dos usuários.

Nessa senda, Shoshana Zuboff (2019, p.119) 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 da Meta, compartilham funcionalidades como publicação de imagens, vídeos e stories com duração de 24 horas, além de incentivar a interação por curtidas e comentários. O *Facebook*, mais antigo, inclui criação de grupos, páginas e um feed de notícias diversificado (Meta, 2024a). Já o *Instagram*, inicialmente focado em fotos, é popular entre jovens e relevante para influenciadores e marcas, com opções de anúncios e conteúdo patrocinado, promovendo interação por mensagens diretas e seguidores interessados. Ambos utilizam modelos de engajamento robustos (Meta, 2024b).

Dentre as plataformas mais utilizadas atualmente, o *Tik Tok* é a mais recente, gerida pela empresa chinesa ByteDance. 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 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* vendem a ideia de "proporcionar uma experiência personalizada através 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 modelo de negócios do Facebook foca em conectar pessoas e permitir a expressão individual por meio de experiências personalizadas, utilizando dados das interações para ajustar conteúdos, como publicações, anúncios e recomendações de conexões, grupos e eventos. A plataforma visa fortalecer comunidades por meio de laços relevantes e permite múltiplas formas de expressão, como atualizações, fotos, vídeos e tecnologias inovadoras, como realidade aumentada. Além disso, promove a descoberta de produtos e serviços por meio de anúncios personalizados, alinhando suas funcionalidades ao interesse dos usuários (Facebook, 2024a).

Já o *Instagram*, enfatiza oferecer experiências personalizadas para criação, conexão, comunicação, descoberta e compartilhamento, destacando a diversidade entre seus usuários. A plataforma oferece diferentes tipos de contas e recursos para ampliar a presença e a interação, promovendo relações relevantes por meio de experiências compartilhadas e recomendando conteúdos e conexões com base no comportamento dos usuários dentro e fora da plataforma. Além disso, utiliza dados para veicular anúncios e conteúdos patrocinados alinhados aos interesses individuais, garantindo que sejam tão relevantes quanto outras experiências no Instagram (Instagram, 2024a).

Seus 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. Tal processo ilustra a dinâmica com que as redes sociais potencializam pequenas atividades e as transformam em mercadorias valiosas.

Lawrence Lessig (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 conteúdos. 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.

A moderação de conteúdo é essencial para assegurar um ambiente informacional seguro e saudável nas plataformas digitais, operando com base em diretrizes internas e considerando legislações e padrões culturais. Além de controlar o que é permitido, desempenha um papel vital no gerenciamento do comportamento online (Zuboff, 2019). Ribeiro e Favero (2024) ressaltam que a moderação atua como um mecanismo de governança, estruturando a participação nas comunidades digitais, promovendo a cooperação e prevenindo abusos, como discursos de ódio e outras violações dos termos de uso.

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, onde 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. 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).

Ao analisar os termos de uso das três plataformas, constatou-se que o TikTok é a única que aborda, mesmo que de forma breve, a proibição de conteúdos criados para provocar, assediar ou causar desconforto a outros usuários. Especificamente, não permite a publicação de materiais que promovam bullying, ameaças de violência física, racismo ou discriminação com base em características como raça, religião, idade, gênero, deficiência ou orientação sexual, demonstrando preocupação explícita com comportamentos prejudiciais (TikTok, 2024c).

Consoante, 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[1]. A propagação de discursos de ódio objetiva inferiorizar aqueles e aquelas a quem é direcionado, desvalorizando-os como sujeito de direitos na sociedade.

A moderação de conteúdo também observa as diretrizes de comunidade das plataformas, através delas as plataformas estabelecem quais conteúdos são permitidos ou proibidos de divulgar. Conforme divulgam, 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" através da criação de padrões "que incluem diferentes pontos de vista e crenças" (Facebook, 2024b), proporcionando ao

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

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, p.104).

O discurso de ódio, por sua vez, 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” (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”, isto é, características pessoais com as quais se nasce, que são imutáveis ou que causariam danos psicológicos graves se você fosse forçado a mudá-las ou atacado por causa delas.

São discursos e comportamentos incompatíveis com a convivência saudável e harmoniosa 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 Bell Tower 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.[2] [tradução das autoras] (Bell Tower 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. 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.

Embora a moderação de conteúdo nas redes sociais seja guiada por termos de uso e diretrizes, não há transparência sobre como esse processo é efetivamente realizado. Segundo o relatório “Reclamações sobre o Procedimento de Moderação de Conteúdo em Redes Sociais” (IRIS, 2024, p. 05), 54,34% dos usuários criticaram a falta de clareza, apontando problemas como fundamentação inadequada de decisões (52,46%), ausência de respostas para contestações (22,54%), falta de notificação (9,02%), e mecanismos inacessíveis para revisão das decisões (4,51%), revelando significativa insatisfação com o sistema.

A insatisfação dos usuários revela a insuficiência de políticas combativas das plataformas digitais para moderar conteúdos, especialmente discursos de ódio. Ferramentas de denúncia e revisão são opacas, negando aos usuários clareza e direito à contestação. Além disso, as plataformas, ao personalizarem interações via algoritmos que analisam comportamentos, influenciam percepções de forma imperceptível, moldando engajamentos e reforçando polarizações. Essa lógica reflete o capitalismo de vigilância descrito por Zuboff

[2] 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

(2019), em que dados coletados são utilizados para manipular ações e criar dinâmicas sociais pouco transparentes e potencialmente prejudiciais.

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.

CONSIDERAÇÕES FINAIS:

A análise das interações nas redes sociais, sob a lógica dos sistemas complexos, revela uma dinâmica poderosa e pouco transparente, onde 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. Como aponta Shoshana Zuboff, trata-se de um "capitalismo de vigilância", no qual os dados são utilizados para moldar comportamentos, muitas vezes em benefício dos interesses comerciais das grandes empresas de tecnologia.

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.

Assim, conclui-se que, 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.

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

BRASIL. **Ataques às escolas no Brasil: análise do fenômeno e recomendações para a ação governamental**. Brasília: GT Especialistas em Violência nas Escolas, 2023a. Disponível em: https://www.gov. br/mec/pt-br/acesso-a-informacao/participacao-social/grupos-de- -trabalho/prevencao-e-enfrentamento-da-violencia-nas-escolas/resultados/relatorio-ataque-escolas-brasil.pdf. Acesso em 01 nov. 2024

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

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

CGI.br. **Sistematização das Contribuições à Consulta sobre Regulação de Plataformas Digitais** [livro eletrônico]. [editor] Núcleo de Informação e Coordenação do Ponto BR; [textos] Juliano Cappi, 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.

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.

GOV.BR. **Incitação à violência contra a vida na internet lidera violações de direitos humanos com mais de 76 mil casos em cinco anos, aponta ObservaDH**. 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.

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**, 59(2), p. 1216–1238, 2020. Disponível em: https://doi.org/10.1590/010318137609916202007 23. 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**, 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. [organizadores]. **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; ROSA, Isabela Quartieri da. Relações de Poder e Redes de Comunicação: o discurso de ódio protagonizando engajamento. **Revista de Ciências do Estado**, Belo Horizonte, vol.9, n.1, 2024. Disponível em: https://periodicos.ufmg.br/index.php/revice/article/view/e51384/e51384. Acesso em: 05 de 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.

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