

DISCURSO DE ÓDIO DE GÊNERO NAS REDES SOCIAIS

Laíse Milena Barbosa¹

Resumo: O presente artigo realiza uma análise dos termos de uso das cinco redes sociais mais utilizadas no Brasil na temática do discurso de ódio de gênero. O objetivo é verificar se há previsões para sua restrição e a incidência da sua moderação, tratando-se, portanto, de uma pesquisa qualitativa e interpretativa. Para tanto, inicialmente, foram investigadas interpretações e definições do discurso de ódio contra as mulheres, optando-se por denominá-lo de discurso de ódio de gênero. Ainda, através do relatório *Digital Brazil: 2021* foram selecionadas as plataformas e apresentadas algumas pesquisas sobre a moderação do *hate speech* que tem como grande alvo as mulheres. Finalmente, com base nesses elementos, analisou-se as regras das comunidades e alguns relatórios da moderação deste conteúdo, sendo apresentada uma crítica sobre a proteção das usuárias através do método dedutivo e da técnica de pesquisa documental. Como resultado, constatou-se que apenas o Facebook apresenta maior preocupação em detalhar os conteúdos de moderação sobre o discurso de ódio de gênero. Contudo, a tendência na extensão e especificidade dos termos das plataformas constitui uma potente alternativa para a proteção do discurso no ambiente virtual.

Palavras-Chave: gênero; discurso de ódio; redes sociais; termos de uso; moderação de conteúdo.

GENDER HATE SPEECH IN SOCIAL NETWORKS

¹ Mestranda em Direito do Estado pela Universidade Federal do Paraná (UFPR), com bolsa CAPES/PROEX. Bacharela em Direito pela Universidade Estadual de Ponta Grossa (UEPG).

Abstract: This article analyzes the terms of use of the five most used social networks in Brazil on the theme of gender hate speech. The objective is to verify if there are predictions for its restriction and the incidence of its moderation, being, therefore, qualitative and interpretative research. Therefore, initially, interpretations and definitions of hate speech against women were investigated, choosing to call it to gender hate speech. Also, through the report Digital Brazil: 2021, the target platforms for analysis were selected and some researches were presented on the moderation of hate speech, which has as a great target woman. Finally, based on these elements, the rules of the communities and some reports of the moderation of this content were analyzed, with a critique of the protection of users of the theme through the deductive method and the technique of document research being presented. As a result, it was found that only Facebook is more concerned with detailing the content of moderation on gender hate speech. However, the extension and specificity of the terms of the trend of platforms constitute a powerful alternative for the protection of discourse in the virtual environment.

Keywords: gender; hate speech; social networks; terms of use; content moderation.

INTRODUÇÃO



a atualidade, as redes oportunizam um ambiente de debate da política e de repercussão de opiniões. Como comunidades virtuais, possibilitaram desde a organização de movimentos sociais que levaram a derrubada de governo despótico, eleição de líderes populistas através dos algoritmos e, sobretudo, de circulação do discurso de ódio.

Dentre às vítimas do ódio online, mulheres e muçulmanos configuram o maior alvo do que qualquer outro gênero e comunidade (CHETTY; ALATHUR, 2018). Ainda, no Brasil, quase 70% das vítimas de discurso de ódio que procuram ajuda são as mulheres (SAFELAB, 2020).

Diante disso, este artigo se propõe a analisar os termos de uso e/ou padrões/acordos das cinco redes sociais mais utilizadas no Brasil segundo o *Digital Brazil: 2021* (YouTube, Facebook, WhatsApp, Twitter e TikTok) e verificar a previsão de restrição do discurso de ódio de gênero, por meio do método dedutivo.

Ainda, como trata-se de uma pesquisa qualitativa, interpretativa, foi utilizada a técnica documental que possui os termos e padrões das plataformas como fonte. Diante disso, o artigo encontra-se dividido em quatro partes.

O item 1 busca investigar o conceito e interpretações sobre o discurso de ódio de gênero que gera debates acerca do discurso de ódio sexista e misógino. Para a construção da argumentação e fundamentação da utilização das mídias sociais, tanto em âmbito global como nacional, foram apresentados dados coletados da *We Are Social* e *Hootsuite (Digital Brazil: 2021 e Digital 2021: Global Overview Report)* e pesquisas acerca das diferenças do discurso de ódio offline do online e as mulheres como principais vítimas.

Por conseguinte, através do acesso aos termos de uso de cada uma das plataformas nos sites originais, foram expostas as previsões de restrição do discurso de ódio de gênero e alguns relatórios da moderação deste conteúdo. Por fim, fora desenvolvida uma análise crítica destas declarações, levando em conta o perfil dos usuários, caráter educativo e procedimento de moderação.

À vista disso, com avanços tecnológicos e popularização das redes sociais, a pesquisa mostra-se atual e relevante pois, além da maioria dos usuários desconhecerem as regras

desenvolvidas pelas plataformas que guiam o processo de moderação, a presença do discurso de ódio de gênero online não só perpetua o patriarcado estruturante na sociedade, mas também a violência.

1. INTERPRETAÇÕES SOBRE O DISCURSO DE ÓDIO CONTRA AS MULHERES

O discurso de ódio é um fenômeno que possui uma pluralidade de interpretações, podendo ser estudado através da liberdade de expressão, tolerância e dignidade humana. Não obstante, diante das tentativas legais de sua regulamentação, surgiram grandes divergências sobre a forma de como cada país interpreta e aplica em seus casos concretos normativas sobre o *hate speech*.

O ódio se dirige para cima ou para baixo, em todo caso sempre em uma visão projetada verticalmente contra “os lá de cima” ou “os lá de baixo”; é uma categoria do “outro” que oprime ou ameaça o “próprio” (EMCKE, 2020), sendo o outro aquele que se deve desprezar ou denunciar.

Considerando esta e as definições do discurso odioso de ataque, estigmatização e silenciamento das minorias², esse discurso caracteriza-se como uma ameaçadora forma de comunicação que ultrapassa valores democráticos a partir do momento que impede a prevalência da harmonia social, depreciando grupos minoritários. Desta forma, legitima práticas sociais destinadas a promover ou preservar hierarquias entre grupos sociais (MOREIRA, 2019), como por exemplo a autenticação de valores patriarcais da sociedade através de um discurso de ódio com base no gênero e sexo.

² O presente artigo adota, também, as interpretações de Jeremy Waldron que compreende o discurso de ódio como algo além da esfera privada de ataque individual, pois a proteção contra esse discurso se baseia na dignidade humana. Para ele, isso se deve ao fato de que o *hate speech* promove e incita a discriminação e a violência, sendo dever do Estado e da sociedade o respeito por este direito (WALDRON, 2010)

Segundo Emcke (2020), o ódio não surge do nada, ele é cultivado, fabricado; é formado coletiva e ideologicamente. Desta forma, o discurso de ódio contra as mulheres é disseminado por uma cultura machista que permeia a nossa sociedade e reforça estigmas ao dar um ar de “neutralidade” ao lugar dos homens e mulheres nesse processo.

Contudo, a diferença é que parece haver um consenso muito mais alargado de que discursos racistas e homofóbicos constituem um discurso de ódio, sendo o de gênero ignorado ou então rejeitado, pois “enquanto o racismo e a homofobia permanecem ‘visíveis’, o sexismo foi tornado ‘invisível’” (LILLIAN, 2007, p. 736, tradução nossa)³.

Resultado disso são as divergências de interpretações e a escassez de conteúdos destinados à uma análise de um discurso de gênero como um discurso de ódio. Chetty e Alathur (2018) definem o discurso de ódio contra as mulheres quando ele é direcionado de maneira intencional geralmente para mulheres e meninas com base na característica do gênero ou sexo.

Um ponto de atenção necessário é de que, de fato, análises separadas sobre discurso de ódio dirigidos às mulheres são necessárias. Em adição à visão verticalizada que o ódio projeta descrito por Emcke, Richardson-Self (2018) destaca que o *hate speech* além de ser uma forma de violência promove uma cultura imperialista da subordinação e exploração. Nesse interim, a autora ressalta a necessidade de distinguir dois tipos de discursos patriarcais (o sexista e o misógino) em face da violência para melhor categorizar um discurso de ódio contra as mulheres.

O sexismo seria a crença idealizada de uma distinção naturalmente hierárquica de que homens são superiores às mulheres. Já a misoginia é caracterizada por um componente de hostilidade que molda as mulheres às normas patriarcais, também identificado pela coerção, estigmatização e difamação.

³ “While racism and homophobia remain ‘visible’, sexism has been rendered ‘invisible’” (LILLIAN, 2007, p. 736).

Como para a autora o discurso de ódio necessariamente deve apresentar a face da violência, o discurso sexista, por mais que seja opressor e moralmente inaceitável, não pode ser considerado um discurso de ódio por não transmitir essa hostilidade ou coação (RICHARDSON-SELF, 2018).

Nessa orientação, Mills (2008) também ressalta ser discutível que o sexismo possa ser visto como um discurso de ódio pois, por mais que em alguns casos ele pode ser interpretado como violência contra as mulheres em geral ou afirmação da violência contra as mulheres, não há como interpretar ou comprovar se a real intenção do agente era de incitar a violência. Para ela, esta distinção pode ser feita pois o discurso sexista pode ser visto como um confronto dentro da sociedade sobre a posição das mulheres na esfera pública, ao invés de simplesmente uma expressão de emoções negativas sobre as mulheres em geral.

Todavia, como o racismo e outras formas discriminatórias que eclodem no discurso de ódio, o discurso sexista também é resultante de forças sociais mais amplas, desigualdades institucionalizadas de poder e conflito sobre quem tem direitos a certas posições sócio-políticas e recursos econômicos. Sob este ponto de vista, “o sexismo é um índice do conflito em curso entre homens e mulheres” (D’ANGELO, 2020, p. 143, tradução nossa)⁴.

O Conselho Europeu de Estratégia de Igualdade de Gênero conceitua o discurso de ódio contra as mulheres como um discurso de ódio sexista por considerar o sexo. Assim, esse discurso além de representar uma forma de violência contra mulheres também contribui para a perpetuação da desigualdade de gênero

any supposition, belief, assertion, gesture or act that is aimed at expressing contempt towards a person, based on her or his sex or gender, or to consider that person as inferior or essentially reduced to her or his sexual dimension (...)

⁴ “Sexism is an index of the on-going conflict between men and women” (D’ANGELO, 2020, p. 143).

expressions which spread, incite, promote or justify hatred based on sex (COUNCIL OF EUROPE, p. 2)⁵.

Desta forma, o discurso de ódio de gênero está relacionado à representação das mulheres no meio social, que continua envolta de estereótipos arraigados de várias formas. Ameaças brutais e sexualizadas de morte, estupro e violência que objetificam o corpo feminino; comentários ofensivos sobre aparência, sexualidade, orientação sexual ou papéis de gênero em redes sociais que seguem um modelo machista pré-estabelecido de feminilidade; argumentos da diferença biológica para apontar características femininas de hipersensibilização que nada mais são do que o fruto da opressão; supostas piadas e utilização do humor que ridicularizam a vítima; uso não consensual de imagens, são alguns exemplos.

A preocupação com este discurso advém do fato de que ele por si só se configura como um dano, pois interfere nos sentimentos e modela a forma de como as mulheres agem em público, corroborando para sistemas de hierarquia de gênero na esfera pública (NIELSEN, 2002, p. 279).

Ainda, na nova dimensão do ambiente virtual e das redes sociais, a transmissão de ideias patriarcais, especialmente do discurso de ódio de gênero/sexo, pode possuir uma grande audiência. A popularização das redes funciona como fórum para localizar suas vozes discriminatórias online, mas a investigação deste discurso nestas ferramentas está apenas na fase infantil, especialmente nos países em desenvolvimento (ELIAS; GURBANOVA, 2018).

Diante de todo o exposto, o presente artigo reconhece as distinções existentes entre discurso sexista e misoginia e os debates acerca da composição dos conceitos dentro da definição do

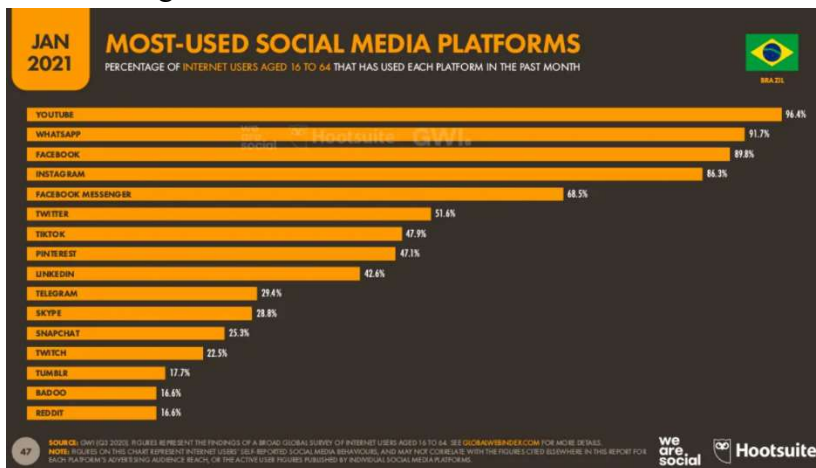
⁵ Tradução nossa: “qualquer suposição, crença, afirmação, gesto ou ato que visa expressar desprezo por uma pessoa, com base em seu sexo ou gênero, ou considerar essa pessoa como inferior ou essencialmente reduzida à sua dimensão sexual (...) inclui expressões que espalham, incitam, promovem ou justificam com base no ódio” (COUNCIL OF EUROPE, p. 2).

discurso de ódio. Portanto, opta por adotar a expressão de “discurso de ódio de gênero” para abarcar ambas as possibilidades. A escolha também se justifica pelas considerações que serão examinadas nas previsões dos termos e padrões das plataformas alvo de análise no último item.

2. PERFIL DAS REDES MAIS UTILIZADAS NO BRASIL E O DISCURSO DE ÓDIO ONLINE

A América do Sul é a quarta região com mais usuários de mídia social em comparação com a população total. Atrás apenas da América do Norte, Europa Ocidental e o Norte da Europa, cerca de 72% da população são usuários ativos nas mídias sociais. No Brasil, a estimativa se reduz para 70.3% da população total, com aumentos anuais de ao menos 7.1% de usuários (DIGITAL BRAZIL, 2021).

Estes são dados do relatório Digital Brazil: 2021 realizado pelo *We Are Social* e *Hootsuite*, o qual também delimitará o artigo nas pesquisas das cinco redes mais utilizadas no Brasil, conforme o gráfico abaixo:



DIGITAL BRAZIL: 2021 (2021)

Ainda, o tempo de utilização das plataformas chama a

atenção pois, segundo o relatório Global, o Brasil está em 3º lugar no ranking da média do tempo gasto, com 3h42min por dia (DIGITAL 2021).

Além da ampla participação das pessoas nas redes, um papel de destaque das ferramentas é na formação da opinião pública. Uma pesquisa realizada pela Câmara dos Deputados e pelo Senado Federal em novembro de 2019 mostrou, em porcentagens, qual a principal fonte de informação dos brasileiros. Dentre os principais veículos citados, em primeiro lugar está o WhatsApp com 79% dos brasileiros, logo após canais de televisão (50%), Youtube (49%), o Facebook (44%) e sites de notícias (38%) (VALENTE, 2019).

Estes dados demonstram como as redes sociais estão presentes no cotidiano da sociedade brasileira e acabam por interferir, até mesmo, no perfil e crenças dos seus usuários. Isto sobreleva ainda mais a importância de conhecer e interpretar os termos e padrões da plataforma que controlam a produção de conteúdo.

Retornando à temática do *hate speech*, antes de partir para a análise das declarações que compõe as plataformas, faz-se interessante compreender as diferenças existentes entre discurso de ódio online e offline. Brown (2018) destaca cinco: anonimato (1) que permite um discurso mais livre e desinibe os oradores; invisibilidade (2) caracterizada pela distancia do locutor e o público; comunidade (3) que a internet facilita ao atrair novos membros e garantir que se sintam conectados (própria das redes sociais); instantaneidade (4) que a Internet não apenas facilita, mas também incentiva, respostas instantâneas que são por natureza mais espontâneas e; (5) capacidade de ferir que, contudo, não há nenhuma evidência científica social que comprove os efeitos do discurso de ódio espontâneo em comparação com o discurso de ódio premeditado, podendo o online ter efeitos especialmente prejudiciais por ser feito na frente de um público maior, permanecendo de uma forma digital permanente.

Desta forma, diante das novas possibilidades tecnológicas da internet que facilitam a disseminação, surge o dever de se atentar para a maneira de como esse discurso pode se espalhar e as vítimas vulneráveis.

Tanto mulheres como meninas são alvos potenciais deste tipo de discurso online e offline. Cerca de 26% das mulheres entre 18 e 24 anos já foram perseguidas online e 25% enfrentaram assédio sexual online, sendo as afro-americanas e hispânicas mais propensas a sofrerem o assédio do que usuárias brancas (COUNCIL OF EUROPE, 2016). Relacionando a pauta de racial com gênero, Trindade (2018) notou que 81% das vítimas de discurso depreciativo nas redes sociais brasileiras são mulheres negras entre 20 e 35 anos e que ocupam cargos de relevância social, ou seja, que saem do padrão de subserviência.

No mais, no ano de 2016, a Organização Não-Governamental Comunica que Muda (2016) publicou o resultado de uma pesquisa que vasculhou diversas mensagens em redes sociais a fim de demonstrar em números o grau de intolerância na internet. O segundo tema com maior número de postagem foi o ódio contra mulheres, ficando atrás apenas dos xingamentos e discursos segregadores que incitavam o ódio no panorama político das eleições de 2014. Foram coletadas 49.544 citações que abordavam as desigualdades de gênero, sendo 88% delas com viés intolerante. Dentre as principais temáticas estão: “assédio, pornografia de vingança, incitação ao estupro e outras violências são, por vezes, travestidos de ‘piadas’” (COMUNICA QUE MUDA, 2016).

À vista disso, quem inventa ou é responsável por manter a tecnologia que facilite o dano e ações ilegais, especialmente se alguém lucrar com essa tecnologia, deve ser responsável pela introdução de medidas destinadas a prevenir ou limitar essas ações (BROWN, 2018). Nesse sentido, não limitando-se apenas ao bem estar dos usuários, mas também às exigências dos parceiros comerciais, algumas plataformas reformularam e

desenvolveram seus termos de uso para adaptação às exigências da era virtual.

Diante disso, no próximo tópico passaremos a analisar cinco termos de uso ou padrões/acordos dos usuários acerca do discurso de ódio de gênero.

3. ANÁLISE DOS CINCO TERMOS DE USO DAS REDES MAIS UTILIZADAS

Os Termos de Uso ou Padrões/Diretrizes/Acordo da Comunidade constituem um conjunto de regras das plataformas que descrevem o que é ou não permitido. Nesta pesquisa, as ferramentas alvo de análise serão: YouTube, Facebook, WhatsApp, Twitter e TikTok⁶.

Ainda, foram investigadas as versões em português de cada uma das plataformas, através da publicação em seus sites oficiais, e que foram atualizadas até a data da pesquisa (setembro de 2021). Ou seja, quaisquer alterações realizadas a partir dessa data não estão incluídas nesta análise.

3.1. YOUTUBE

Com mais de 127 milhões de usuários, dos quais 52.9% são mulheres (DIGITAL BRAZIL), a política da plataforma o YouTube (2021a) possui cinco tópicos específicos de proteção quanto ao conteúdo violento e perigoso, sendo um deles denominado “Política contra discurso de ódio”.

Nesse ponto, que teve sua última alteração no dia 05 de junho de 2019, a plataforma se posiciona pela remoção de todo conteúdo que promova a violência ou o ódio contra indivíduos ou grupos com base em várias características, como idade, raça,

⁶ Ressalta-se que, conforme será narrado posteriormente no tópico 3.2, as plataformas Messenger e Instagram que compõe o quarto e quinto lugar no relatório *Digital Brazil: 2021* constituem-se como produtos do Facebook. Desta forma, aplicam-se os mesmos Padrões da Comunidade desta rede, não sendo necessária a investigação em separado.

classe, religião, orientação sexual ou condição de veterano de guerra e, especialmente, sexo/gênero.

Ainda, deixa claro que esta política evita a publicação de conteúdos que incitem a violência ou o ódio contra pessoas ou grupos de acordo com as características, enquadrando-se nessa perspectiva a desumanização ou comparação com animais, insetos, pragas, doenças ou qualquer outra entidade não humana; afirmar que pessoas ou grupos são fisicamente ou mentalmente inferiores, deficientes ou doentes; defender a superioridade de um grupo sobre outros; condenar a atração emocional, romântica ou sexual de uma pessoa etc (YOUTUBE, 2021a).

Contudo, sob o argumento de manutenção da liberdade de expressão, a plataforma permite o discurso de ódio se este for relacionado com um conteúdo educativo, documental, científico ou artístico (YOUTUBE, 2021b).

No mais, o YouTube disponibiliza o número de vídeos removidos que violam sua política contra o *hate speech*. De abril a junho, 87.943 vídeos foram retirados. Ainda, também traz 12 exemplos de vídeos denunciados pela violação deste discurso, como um que contém uma música cuja letra promove a violência contra um grupo racial e outro que afirma que muçulmanos e refugiados são violentos e estão propensos a cometer crimes. Ambos foram removidos⁷ (YOUTUBE, 2021b), porém, nenhum deles diz respeito ao discurso de ódio de gênero/sexo.

3.2. FACEBOOK

No caso do Facebook, a ferramenta conta com mais de 130 milhões de usuários, sendo que 53.5% são mulheres (DIGITAL BRAZIL, 2021). Ressalta-se que o Instagram e o Facebook Messenger constituem serviços dos Produtos do Facebook

⁷ Interessante notar que o período trimestral com maior número de vídeos removidos na temática do discurso de ódio foi de abril a junho de 2019, com cerca de 111.960 (YOUTUBE, 2021b), logo após a última alteração da sua política sobre o tema.

(FACEBOOK, 2021a) e, por este motivo, por mais que existam dinâmicas diferentes na troca de informações, os termos de uso das três plataformas são iguais, possuindo um tópico específico sobre discurso de ódio.

Contudo, se comparado com as disposições do YouTube, os padrões do Facebook descrevem com maiores detalhes o conceito e características do *hate speech* alvo de moderação. Seu texto prevê que a remoção do conteúdo acontecerá quando houver um incitamento direto à violência contra o que denominam de “características protegidas”, sendo elas: raça, etnia, nacionalidade, religião, orientação sexual, casta, identidade de gênero, doença grave ou deficiência e, no caso em tela, sexo e gênero (FACEBOOK, 2021b).

Além disso, o Facebook também categoriza o discurso com base em três níveis, mas não deixa clara a existência de alguma hierarquia entre eles. Para a pesquisa atual, foi restringido os comentários que, com base nas fundamentações do primeiro capítulo, enquadram-se como um discurso de ódio de gênero.

Quanto ao primeiro nível, um importante ponto de previsão para a pesquisa são as proibições em casos de comparações, generalizações ou declarações na questão de gênero que descrevem “mulheres como objetos domésticos ou referência à mulher como propriedade ou ‘objeto’” (FACEBOOK, 2021b).

No que tange ao segundo nível, incluem-se os termos pejorativos relacionados a atividades sexuais, como exemplo vagabunda, vadia e declarações de inferioridade com expressões sobre ser melhor/pior do que outra característica protegida, incluindo, entre outras: “Eu acredito que os homens são superiores às mulheres”. Por fim, o terceiro nível, restringe-se ao conteúdo que apoie a segregação, a exclusão de grupos, de maneira explícita, política, econômica e social (FACEBOOK, 2021b).

Na mesma seção a plataforma também apresenta alguns dados. O primeiro deles é o percentual de probabilidade de um usuário observar um discurso de ódio a cada 10 mil

visualizações, sendo estimado o número de 05 entre abril e junho de 2021. Ainda, acerca dos conteúdos que o Facebook tomou medida sobre o tema, no mesmo período de tempo, foram de 31.5 milhões, sendo que 97.60% do conteúdo foi sintonizado pela própria plataforma e o restante pelos usuários (FACEBOOK, 2021b).

Ademais, com mais de 99 milhões de usuários, na qual 58.2% são mulheres (DIGITAL BRAZIL, 2021), o Instagram configura-se como a plataforma da pesquisa de maior número de usuárias do sexo feminino do Brasil. Nas suas próprias diretrizes da comunidade, o Instagram (2021) direciona a remoção de conteúdo sobre discurso de ódio às diretrizes do Facebook expostas anteriormente.

No tocante à disponibilização de dados sobre a moderação do *hate speech*, a ferramenta tomou medidas sobre o tema, no período de abril a junho de 2021, em 9.8 milhões de postagens, sendo deste total sinalizado pela própria plataforma a porcentagem de 95.10% (número um pouco menor se comparado ao do Facebook) (FACEBOOK, 2021b).

3.3. WHATSAPP

Para além de constituir uma fonte de comunicação de mensagens instantâneas, o WhatsApp já ocupou grandes espaços nas campanhas eleitorais de 2018 no Brasil e, conforme observado no capítulo anterior, também constitui como uma grande fonte de informação para os brasileiros.

O Instituto OPUS (2020) realizou uma pesquisa para descobrir o perfil dos usuários e constatou que, assim como as já mencionadas, a plataforma é mais popular entre as mulheres do que entre os homens no Brasil.

Porém, em comparação com as ferramentas supramencionadas, o WhatsApp prevê nos seus Serviços situações extremamente superficiais do que não é lícito e aceitável ao usuário,

como por exemplo o item “b” que preconiza a impossibilidade de utilização

“de forma ilícita, obscena, difamatória, ameaçadora, intimidadora, assediante, odiosa, ofensiva em termos raciais ou étnicos, ou que instigue ou encoraje condutas que sejam ilícitas ou inadequadas, como a incitação a crimes violentos, a exploração de crianças ou outras pessoas, a ação de colocá-las em perigo, ou a coordenação de danos reais” (WHATSAPP, 2021)

Por fim, não foram encontrados relatórios e dados referentes à moderação de conteúdo, principalmente na temática da pesquisa.

3.4. TWITTER

No Acordo do Usuário do Twitter (2021), o qual compreende seus Termos de Serviço, a plataforma descreve que o seu objetivo é de proporcionar o diálogo público, sendo impedidos discursos que impeçam essa concretização através da violência, assédio e outros comportamentos semelhantes.

Desta forma, para garantia do diálogo com liberdade e segurança, a plataforma prevê em um tópico específico a conduta de propagação de ódio, dispondo que “também não é permitido promover violência, ameaçar ou assediar outras pessoas com base em raça, etnia, nacionalidade, orientação sexual, gênero, identidade de gênero, religião, idade, deficiência ou doença grave” (TWITTER, 2021a).

Ainda, no seu *help center* o Twitter reconhece que pesquisas tem mostrado que alguns grupos de pessoas historicamente marginalizadas são alvos desproporcionais do abuso online, incluindo, entre eles, as mulheres, pessoas de cor e lésbicas. Portanto, descreve que calúnias ou linguagens sexistas e racistas não consensuais que pretendam desumanizar, degradar ou reforçar estereótipos negativos ou prejudiciais a alguma das categorias são proibidas (TWITTER, 2021b).

Alguns relatórios breves e gerais sobre solicitações de

remoção e informações sobre a temática do *hate speech* são apresentados pela plataforma. No 17º Relatório de Transparência (TWITTER, 2020), último publicado referente ao período de janeiro a junho de 2020, foi observada uma redução de 35% no número de contas que sofreram medidas corretivas sob política de ódio. Porém, o Twitter também ressalta as interrupções no fluxo de trabalho por conta da Covid-19.

3.5. TIKTOK.

Nas suas diretrizes para a comunidade, o TikTok apresenta um tópico específico para comportamento de ódio. A plataforma destaca sua prioridade à “segurança, diversidade, inclusão e autenticidade”. Diante disso, ressalta que não tolera a discriminação ou qualquer conteúdo que envolva comportamento de ódio, estando os usuários passíveis de serem suspensos ou banidos (TIKTOK, 2021a).

Da mesma forma que nas demais plataformas, o TikTok (2021a) bane o discurso de ódio através de atributos protegidos que abrangem: raça, etnia, origem nacional, religião, casta, orientação sexual, sexo, gênero, identidade de gênero, doença grave, deficiência e condição migratória.

No mais, dentro deste rol traz previsões sobre comparações com entidades não humanas, promoção ou justificativa de exclusão, inferioridade (TIKTOK, 2021a), mas não apresenta exemplos ou pontos específicos sobre discurso de ódio de gênero.

No seu último relatório de transparência do primeiro trimestre de 2021, a plataforma destaca que dos vídeos removidos, 2,3% violaram essa política. Ainda, dentro desse percentual, 67,3% dos vídeos de comportamento de ódio foram removidos antes de serem denunciados, embora 83,9% tenham sido removidos 24 horas após a postagem (TIKTOK, 2021b).

Em comparação com o ano de 2020, houve um aumento

de 0.3% na proteção pois, de acordo com a plataforma, foi instaurado um modelo mais aprimorado na detecção deste tipo de discurso. Porém, ainda assim, continua sendo um desafio detectar (TIKTOK, 2021b).

4. A PROTEÇÃO DAS USUÁRIAS FRENTE AO DISCURSO DE ÓDIO DE GÊNERO NAS REDES SOCIAIS: UMA ANÁLISE

Conforme foi observado no item anterior, há uma ampla cobertura das plataformas pelo discurso de ódio. Com exceção do WhatsApp, que apresenta apenas um parágrafo amplo e disperso sobre indicações que não são recomendadas aos usuários, as demais reconhecem e prescrevem explicitamente a proteção da categoria sexo ou gênero dentro do *hate speech*.

Por outro lado, somente o Facebook aponta alguns exemplos e contextos que se enquadrariam no conceito do discurso de ódio de gênero descrito no item 1 do artigo (o qual engloba o sexista e a misoginia), conforme demonstrado nas situações de inferioridade frente à figura masculina e comparações com objetos domésticos.

O Twitter até reconhece em uma comunicação separada no seu centro de ajuda que mulheres, pessoas de cor e lésbicas constituem um grupo específico que são recorrentemente alvos de linguagens sexistas e racistas, mas ao mesmo tempo não transfere sua “confissão” ao seu Acordo de Usuários que esclareceria e advertiria sua comunidade.

No mais, o YouTube ao apresentar algumas informações dos vídeos denunciados dentro da temática, sem esclarecer a escolha, restringe-se às descrições de ataques às categorias de muçulmanos, refugiados, gays e comentários pró-nazismo (YOUTUBE, 2021b).

Diante disso, a primeira observação que deve ser considerada é que, ao expor orientações curtas, sem inserir os usuários

em conjunturas e padrões que possam ser alvo de remoção, dificulta o processo de ação pedagógica educativa. Como documentos que regem não apenas a dinâmica do ambiente digital das plataformas, mas limitam o direito fundamental da liberdade de expressão, o zelo por padrões bem escritos e fundamentados é indispensável, notadamente no conteúdo sobre discurso de ódio de gênero.

Adicionado a isso, diante da tendência dos usuários de não lerem as políticas de privacidade e os termos de serviço das plataformas (OBAR; OELDORF-HIRSCH, 2018), não há como esperar que por iniciativa própria as pessoas conheçam os padrões da comunidade. Assim, através de algum mecanismo de alerta, as ferramentas podem incentivar a busca pela informação.

Outro ponto interessante para destaque é que a ampla participação do público feminino nas três primeiras plataformas reforça a importância das ferramentas de ampliarem sua proteção e políticas educacionais para, especialmente, as usuárias que hoje são a maioria no contexto brasileiro.

Portanto, deve ser ressaltado o compromisso e responsabilidade das empresas em alertar os usuários acerca dos padrões já existentes. Da mesma maneira, se as pessoas pensarem que sua mensagem e conteúdo podem ser removidos e, mais importante, que seu usuário pode ser fechado por empresas de Internet, elas podem pensar duas vezes antes de postar determinado conteúdo (BROWN, 2018).

Apesar disso, como observado nos dados coletados, grande parte da remoção do conteúdo sobre discurso ódio (o qual infelizmente não é catalogado por incidência de categoria protegida, como o gênero) ocorre de maneira automática pelas plataformas, antes mesmo da advertência do usuário. Essas sinalizações automáticas são fundamentais para ajudar na identificação de conteúdo com violações que, logo após, submetem-se à uma revisão humana para a adoção das medidas adequadas.

Contudo, aplicar os termos que dizem respeito ao *hate*

speech de forma geral e em grande escala é uma tarefa complexa, porque exige uma compreensão diferenciada de idiomas e contextos locais (DOUEK, 2018), como a interpretação de termos e palavras específicas. Desta forma, a importância dos usuários em se manterem sempre atualizados dos termos e políticas das plataformas é indispensável não apenas para realizar denúncias próprias dos conteúdos, mas também para fiscalizar e monitoramento feito pelos algoritmos.

Por fim, o último ponto é da necessidade de maior transparência e *accountability* na moderação de conteúdo por parte das plataformas. Algumas alternativas expostas por Bowers e Zittrain (2020) são de apresentação de relatórios das atividades, liberação de dados aos pesquisadores e formulação de regras claras.

Desta forma, ressalta-se que através da pesquisa foi possível observar que as plataformas apresentaram relatórios da sua moderação (com exceção do WhatsApp), mas que não foram suficientes para conhecer a incidência da moderação sobre a tema do discurso de ódio de gênero.

Diante de todo o exposto, os Termos de Uso das ferramentas ao concederem a permissão para a moderação de conteúdo podem servir como “documentos constitucionais” (SUZOR, 2018) para salvaguardar algumas garantias dos usuários, mas também sustentar decisões arbitrárias que dão grande poder às plataformas. Assim, torna-se extremamente relevante o conhecimento das regras dispostas e processo de aplicação da restrição pela comunidade no geral.

CONCLUSÕES

As análises e dados apresentados demonstraram que o virtual está propagando a mesma lógica de exclusão e opressão que a esfera offline. Mesmo diante do maior número de usuárias mulheres nas redes sociais, o grupo ainda é o grande alvo do *hate*

speech.

As regras das plataformas certificaram que a própria demonstração e definição da moderação do discurso de ódio de gênero é muito incipiente. O WhatsApp se limita em não incentivar o uso dos seus serviços para propagação de mensagens odiosas (em âmbito geral), enquanto que o YouTube e o Twitter possuem apenas a breve previsão de restrição do discurso relacionado ao gênero/sexo.

Conforme exposto, somente o Facebook apresentou maior preocupação em detalhar os conteúdos sobre o discurso contra as mulheres. Ainda, um destaque atual e relevante da plataforma é a recente fundação do Comitê de Supervisão que, independentemente, realiza uma análise das decisões de remoção de conteúdo da ferramenta, abrangendo diversos casos sobre o *hate speech*.

A ironia é que as empresas de internet aparentam ser a resposta capaz de fornecer aos oradores de ódio ferramentas online cada vez mais poderosas para espalhar suas mensagens. Entretanto, como foi observado nos relatórios de moderação, também estão em melhor posição para policiar o uso dessas ferramentas antes mesmo dos usuários.

Desta forma, por mais que se acredite no potencial e responsabilização das plataformas pela produção do conteúdo, a necessidade dos seus usuários de se manterem atentos aos termos e padrões de uso é imprescindível. Mesmo com o avanço visível das regras em se tornarem mais extensas e específicas, ainda assim devem ser objeto de constante questionamento e controvérsia, especialmente a maioria das usuárias que são vítimas do discurso de ódio online.



REFERÊNCIAS BIBLIOGRÁFICAS

- BOWERS, John. ZITTRAIN, Jonathan. Answering impossible questions: Content governance in an age of disinformation. **Harvard Misinformation Review**. 14 de janeiro de 2020. Disponível em: <<https://misinformationview.hks.harvard.edu/article/content-governance-in-an-age-of-disinformation/>>. Acesso em: 20 set. 2021.
- BROWN, Alexander. What Is so Special about Online (as Compared to Offline) Hate Speech? **Ethnicities** 18 (3), p. 297-326, 2018.
- COMUNICA QUE MUDA. Redes da Intolerância. 03 agosto 2016. Disponível em: <https://www.comunicaquemuda.com.br/redes-da-intolerancia/>. Acesso em: 02 set. 2021.
- CHETTY, Naganna; ALATHUR, Sreejith. Hate Speech Review in the Context of Online Social Networks. **Aggression and Violent Behavior**, 40, p. 108-180, 2018.
- COUNCIL OF EUROPE. Background Note on Sexist Hate Speech. 01 fevereiro 2016. Disponível em: <https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=090000168059ad42>. Acesso em: 02 set. 2021.
- COUNCIL OF EUROPE. Combating Sexist Hate Speech. Disponível em: <https://rm.coe.int/1680651592>. Acesso em: 02 set. 2021.
- D'ANGELO, Francesca. Freedom of Expression and Sexist Hate Speech: Where is the Boundary? La camera blu. **Journal of Women's and Gender Studies**, v. 23, 2020.
- DIGITAL BRAZIL. Digital 2021: Brazil. 21 de fevereiro de 2021. Disponível em: <<https://datareportal.com/reports/digital-2021-brazil>>. Acesso em: 02 set. 2021.
- DIGITAL 2021. Digital 2021: Global Overview Report. 27 de janeiro de 2021. Disponível em: <<https://datareportal.com/reports/digital-2021-global-overview-report>>.

Acesso em: 02 set. 2021.

DOUEK, Evelyn. Facebook's Oversight Board: move fast with stable infrastructure and humility. **North Carolina Journal of Law and Tecnology**, v. 21, 2021.

ELIAS, Stanley.; GURBANOVA, Nubar. Relocating Gender Stereotypes Online: Critical Analysis of Sexist Hate Speech in Selected Social Media. In: **International Conference on Language Phenomena in Multimodal Communication (KLUA)**, Advances in Social Science, Education and Humanities Research (ASSEHR), v. 228, p. 272-277, 2018. Disponível em: <https://www.atlantispress.com/proceedings/klua-18/25900116>. Acesso em: 25 ago. 2021.

EMCKE, Carolin. **Contra o ódio**. São Paulo: Âyiné, 2020.

FACEBOOK. O que são os Produtos do Facebook? 2021a. Disponível em: <https://www.facebook.com/help/1561485474074139?ref=igtos>. Acesso em: 10 set. 2021.

FACEBOOK. Discurso de ódio. 2021b. Disponível em: <https://transparency.fb.com/pt-br/policies/community-standards/hate-speech/> Acesso em: 02 set. 2021.

INSTAGRAM. Diretrizes da Comunidade. 2021. Disponível em: https://help.instagram.com/477434105621119/?helpref=hc_fnav. Acesso em: 11 set. 2021.

INSTITUTO OPUS. Perfil dos usuários das redes sociais: WhatsApp, Facebook e Instagram. 12 de julho de 2021. Disponível em: <https://www.opuspesquisa.com/blog/estudo/perfil-usuarios-redes-socais-whatsapp-facebook-instagram/>. Acesso em: 11 set. 2021.

LILLIAN, Donna L. A thorn by any other name: sexist discourse as hate speech. **Discourse & Society**, v. 18, n. 6, p. 719–740, 2007.

MILLS, Sara. **Language and Sexism**. Cambridge: Cambridge

- University Press, 2008.
- MOREIRA, Adilson. **Racismo Recreativo**. São Paulo: Pólen, 2019.
- NIELSEN, Laura Beth. Subtle, Pervasive, Harmful: Racist and Sexist Remarks in Public as Hate Speech. **Journal of Social Issues**, v. 58, n. 2, p. 265-280, 2002.
- OBAR, Jonathan A.; OELDORF-HIRSCH, Anne. The Biggest Lie on the Internet: Ignoring the Privacy Policies and Terms of Service Policies of Social Networking. **Services, Information, Communication & Society**, p. 1-20, 2018.
- RICHARDSON-SELF, Louise. Woman-Hating: On Misogyny, Sexism, and Hate Speech. **Hypatia**, v. 33, n. 2, p. 256 – 272, 2018.
- SAFERLAB. Denúncias de ódio recebidas pela SaferNet Brasil. 2020. Disponível em: <<http://saferlab.org.br/o-que-e-discurso-de-odio/>>. Acesso em: 28 ago. 2021.
- SUZOR, Nicolas. Digital Constitutionalism: Using the Rule of Law to Evaluate the Legitimacy of Governance by Platforms. **Social Media + Society**, 2018.
- TRINDADE, Luiz Valério P. It is not that funny. Critical analysis of racialideologies embedded in racialized humour discourses on social media in Brazil. **University of Southampton** (tese de doutorado), 2018. Disponível em: <https://www.academia.edu/52681990/It_is_not_that_funny_Critical_analysis_of_racial_ideologies_embedded_in_racialized_humour_discourses_on_social_media_in_Brazil_PhD_Thesis_>. Acesso em: 10 set. 2021.
- TIKTOK. Diretrizes da Comunidade. 2021a. Disponível em: <https://www.tiktok.com/community-guidelines?lang=pt_BR#38>. Acesso em: 12 set. 2021.
- TIKTOK. Relatório de Transparência do TikTok. 2021b. Disponível em:

- <<https://www.tiktok.com/safety/resources/tiktok-transparency-report-2021-h-1?lang=pt-BR>>. Acesso em: 12 set. 2021.
- TWITTER. Acordo do usuário de Twitter. 2021a. Disponível em: <https://cdn.cms-twodigitalassets.com/content/dam/legal-twitter/site-assets/tos-aug-19th-2021/Twitter_User_Agreement_PT.pdf>. Acesso em: 11 set. 2021.
- TWITTER. Hateful conduct policy. 2021b. Disponível em: <<https://help.twitter.com/en/rules-and-policies/hateful-conduct-policy>>. Acesso em: 11 set. 2021.
- TWITTER. Destaques do 17º Relatório de Transparência do Twitter. 2020. Disponível em: <https://blog.twitter.com/pt_br/topics/company/2020/destaques-do-17-relatorio-de-transparencia-do-twitter>. Acesso em: 11 set. 2021.
- VALENTE, Jonas. WhatsApp é principal fonte de informação do brasileiro, diz pesquisa. **Agência Brasil**. 10 dezembro de 2019. Disponível em: <https://agenciabrasil.ebc.com.br/geral/noticia/2019-12/whatsapp-e-principal-fonte-de-informacao-do-brasileiro-diz-pesquisa#:~:text=Uma%20pesquisa%20realizada%20pela%20C3%A2mara%20dos%20Deputados%20e,a%20plataforma%20mais%20popular%20juntamente%20com%20o%20Facebook>. Acesso em: 24 ago. 2021.
- WALDRON, Jeremy. Dignity and defamation: the visibility of hate. **Harvard Law Review**, v. 123, n. 1.596, p. 1.597-1.657, 2010. Disponível em: https://harvardlawreview.org/wp-content/uploads/pdfs/vol123_waldron.pdf. Acesso em: 05 set. 2021.
- WHATSAPP. Termos de Serviço do WhatsApp. 4 de janeiro de 2021. Disponível em: <<https://www.whatsapp.com/legal/updates/terms-of-service>>. Acesso em: 11 set. 2021.

YOUTUBE. Política contra discurso de ódio. 2021a. Disponível em: https://support.google.com/youtube/answer/2801939?hl=pt-BR&ref_topic=9282436. Acesso em: 10 set. 2021.

YOUTUBE. Políticas em destaque. 2021b. Disponível em: https://transparencyreport.google.com/youtube-policy/featured-policies/hate-speech?hl=pt-BR&policy_removals=period:2019Q2&lu=policy_removals. Acesso em: 10 set. 2021.