



# MODERAÇÃO DE CONTEÚDO EM GRUPOS BRASILEIROS DO FACEBOOK

AUTORES

João Victor Archegas  
Lucas Henrique Conceição

COLABORADORA

Marina Estarque

REVISÃO

Celina Bottino  
Christian Perrone



Instituto  
de Tecnologia  
& Sociedade  
do Rio

# SUMÁRIO

<b>Resumo Executivo</b>	<b>PG. 3</b>
<b>1. Introdução e Metodologia</b>	<b>PG. 5</b>
<b>2. Ascensão dos Grupos de Facebook: da praça pública à sala de estar</b>	<b>PG. 7</b>
<b>3. Ferramentas de Moderação de Conteúdo e Divisão de Trabalho nos Grupos</b>	<b>PG. 10</b>
<b>3.1. Sistematização das Inovações Tecnológicas de Moderação em Grupos e a Atuação do Facebook</b>	<b>PG. 11</b>
<b>3.2. As Inovações Tecnológicas em Uso: A Moderação em Grupos no Brasil</b>	<b>PG. 15</b>
<b>3.3. Eficiência da Possibilidade de Monetização das Comunidades</b>	<b>PG. 18</b>
<b>4. Desafios Particulares dos Grupos e Interação com a Plataforma</b>	<b>PG. 19</b>
<b>5. Perspectivas de Administradores e Moderadores</b>	<b>PG. 23</b>
<b>6. Considerações Finais</b>	<b>PG. 30</b>
<b>ANEXO 1 - QUESTIONÁRIO DE PESQUISA</b>	<b>PG. 32</b>
<b>SOBRE OS AUTORES</b>	<b>PG. 40</b>

## RESUMO EXECUTIVO

Como é feita a moderação de conteúdo em grandes grupos no Facebook Brasil? Quais são os desafios particulares enfrentados por essas comunidades? Essas são algumas das perguntas que serviram de base para o desenvolvimento da pesquisa a seguir. Partindo de uma série de entrevistas realizadas com líderes de grupos da rede social, o ITS Rio apresenta, pela primeira vez, um estudo sobre as complexidades e nuances do tema no Brasil, considerando como recorte temático representantes de grupos vinculados a minorias políticas e sociais no Brasil. O presente relatório está subdividido da seguinte forma:

No **capítulo 1** é feita uma breve introdução ao tema de pesquisa, além de ser apresentada em maior profundidade a metodologia que guiou o desenvolvimento deste trabalho;

No **capítulo 2** é feita uma retomada histórica do Facebook enquanto rede social, dando destaque a uma mudança de estratégia de seus executivos de se dissociar da figura de “praça pública” da era digital e se aproximar do papel de “sala de estar”;

No **capítulo 3** é feita uma exposição e análise das diversas inovações tecnológicas introduzidas pelo Facebook para facilitar a moderação em grupos. A efetividade das novas inovações é analisada em contraste com os primeiros resultados das entrevistas realizadas com líderes de grandes grupos brasileiros. Os administradores entrevistados representavam os seguintes grupos:

- > LGBTQI+ Resistência pela Democracia;
- > LGBT Brasil;
- > Pais Pretos Presentes
- > Sou do Nordeste mesmo e com orgulho;
- > Empreendedorismo Feminino;
- > [Mulheres que Viajam Sozinhas](#);
- > Feministrampos;
- > Não é Coincidência, é Racismo;
- > Porque Representatividade Importa, e Muito;
- > Umbandistas;
- > A Ponte para Pretxs.

No **capítulo 4** trata-se dos desafios particulares dos grupos no Facebook, como, por exemplo, ataques coordenados de ódio baseados no uso de reações emocionais como risadas ou raiva. Ademais, também é feita uma análise da interação entre os líderes das comunidades e a própria plataforma em face desses desafios;

No **capítulo 5**, por sua vez, são apresentados os principais resultados da fase de avaliação dos entrevistados, passando desde como eles e elas avaliam a atuação do Facebook até a sua própria atuação dentro das comunidades. Por fim, são feitas algumas considerações finais.



# 1 INTRODUÇÃO E METODOLOGIA

O presente relatório busca preencher, ainda que parcialmente, uma significativa lacuna nos estudos sobre a moderação de conteúdo em plataformas digitais de redes sociais no Brasil: a moderação em grupos, ou seja, nos espaços compartilhados na plataforma para o compartilhamento de interesses de determinadas pessoas, no âmbito do Facebook do Brasil. A maior parte das análises na área têm como cerne a formulação e aplicação dos padrões da comunidade da rede social em relação às interações públicas dos usuários. Nada obstante, hoje uma considerável parcela da atividade online dos perfis no Facebook se dá dentro de comunidades, muitas delas fechadas, também conhecidas como grupos. Em 2019, por exemplo, a empresa informou que [1.4 bilhão de usuários](#) estavam ativos em grupos do Facebook mensalmente.

Os Padrões de Comunidade do Facebook já foram analisados mais a fundo em outro [relatório](#) produzido pelo ITS-Rio. Naquela oportunidade, analisou-se o documento publicado em 2018, que contém diretrizes sobre conteúdos e ações que são permitidos ou não na rede social, pautando a forma que a plataforma lida com conteúdos que veiculam violência, assédio, discurso de ódio, notícias falsas, nudez e terrorismo. O relatório também explorou a relação entre a estrutura de princípios e normas internas da plataforma com a pauta regulatória dos Estados Unidos, Alemanha e França. Este relatório apresenta uma continuação daquela pesquisa, contudo analisando especificamente a relação entre a aplicação dos Padrões de Comunidade pela plataforma e pelos administradores de comunidades, considerando inclusive a possibilidade destes gestores estipularem normas específicas em seus grupos.

Portanto, o relatório parte da premissa de que desconsiderar a realidade da moderação de conteúdo em grupos - que, como será visto neste relatório, é pautada por uma dinâmica própria, onde os administradores e moderadores das comunidades têm maior flexibilidade para interpretar as regras comuns da plataforma e, inclusive, estipular novos padrões de conduta específicos às suas comunidades - é não dar atenção para uma importante dimensão das interações na rede social. Esta tarefa se torna ainda mais indispensável quando se leva em consideração que o Facebook, desde pelo menos 2020, está mudando o foco dos seus serviços e vem apostando em interações mais privadas entre os seus usuários em comunidades específicas, muitas delas fechadas, ao invés de interações públicas em páginas e no feed de notícias. Como se verá no tópico subsequente, trata-se de uma transição da rede social de um papel de “praça pública” da era digital para um papel de “sala de estar”, valorizando trocas entre pessoas que já se conhecem ou que possuem interesses em comum.

Portanto, o relatório adota uma metodologia primordialmente qualitativa de investigação, através da realização e análise de entrevistas com os administradores que fundaram onze grupos brasileiros, realizadas ao longo de 2021. Os administradores entrevistados foram aqueles que criaram as comunidades aqui referenciadas, mas a pesquisa reconhece que cada comunidade possui mecanismos internos de governança específicos, com a possibilidade do compartilhamento da função de administrador com mais pessoas assim como a definição de usuários que atuam somente como moderadores, estes últimos sem acesso às ferramentas que permitem a definição de normas específicas a cada grupo. As perspectivas e opiniões analisadas, portanto, refletem a perspectiva dos fundadores destas comunidades e não do grupo em si.

A metodologia qualitativa permite extrair conhecimentos específicos referentes às práticas de moderação utilizadas em diferentes comunidades, considerando a moderação de conteúdo sobre a perspectiva de diversos administradores e moderadores. Para entender melhor estas complexidades e nuances, o ITS Rio selecionou comunidades com um elevado grau de engajamento entre seus membros, sempre priorizando recortes temáticos vinculados a minorias políticas no Brasil de forma a mapear os principais desafios enfrentados por esses grupos. As onze comunidades pesquisadas, assim, podem ser categorizadas dentro dos seguintes recortes: população LGBTQIA+, negras e negros, mulheres, nordestinas e nordestinos e praticantes de religiões de matriz africana. A partir destas entrevistas, buscou-se compreender, então, (1) como essas comunidades se organizam, (2) como se dá e quais são os desafios da moderação nos grupos, (3) quais são os impactos da moderação sobre os líderes desses grupos e (4) como os entrevistados avaliam a atuação do Facebook e a sua própria atuação dentro das comunidades.

Um questionário padrão foi elaborado e utilizado como guia para a condução de todas as entrevistas. O documento foi subdividido em quatro partes. Na primeira parte, os líderes foram questionados sobre aspectos organizacionais de suas comunidades, como, por exemplo, o número de administradores, moderadores e membros. Na segunda parte, os líderes responderam uma série de perguntas sobre a moderação em suas comunidades, passando, por exemplo, pelos desafios que enfrentam, pelo processo de tomada de decisão quando da revisão de um determinado conteúdo, pelo volume de conteúdos que revisam diariamente e pela relação com o Facebook. Na terceira parte, os líderes foram indagados sobre os impactos pessoais causados pela moderação, como, por exemplo, a existência ou não de desgaste emocional. Por fim, na quarta e última parte, os entrevistados foram convidados a avaliar (dando uma nota de um a cinco) diferentes aspectos da moderação nos grupos e no Facebook como um todo, considerando a atuação da empresa na moderação de conteúdo em grupos e páginas, a transparência da plataforma quando do emprego de medidas de moderação nestes espaços, a

efetividade das medidas, a clareza e apresentação dos padrões de comunidade a seus usuários, e as expectativas que moderadores e administradores possuem em referência à efetividade do Comitê de Supervisão estabelecido pela Meta.

## 2 ASCENSÃO DOS GRUPOS DE FACEBOOK: DA PRAÇA PÚBLICA À SALA DE ESTAR

O Facebook foi fundado em 2004 com o objetivo de conectar estudantes da Universidade de Harvard. Conforme foi ganhando popularidade entre os acadêmicos, seu fundador decidiu estender seus serviços para outras universidades da Ivy League, o grupo de elite dos centros de educação superior nos Estados Unidos. Neste primeiro momento, só poderia criar um perfil na plataforma quem tivesse um endereço de e-mail institucional com final “.edu”. A demanda, entretanto, cresceu de forma exponencial e, em pouco tempo, o Facebook não só estava disponível em todas as universidades dos EUA, como também em diversos países ao redor do globo e para todas as pessoas, estudantes ou não.

A missão da empresa, desde que se tornou global, é, nas palavras de seu fundador e CEO, “dar voz às pessoas”. Refletindo sobre os primeiros anos do Facebook e a guerra dos EUA contra o Iraque, Mark Zuckerberg afirmou em sua célebre [palestra](#) na Universidade de Georgetown em 2019 que se recorda de pensar que “se mais pessoas tivessem voz para compartilhar suas experiências, talvez as coisas tivessem acontecido de forma diferente”. Seu objetivo, portanto, era oferecer essa oportunidade para os menos favorecidos e “empurrar a sociedade para que se torne melhor ao longo do tempo”. Como ele mesmo reconhece, um fundamento de seu projeto é a liberdade de expressão e a possibilidade de transformar o mundo por meio do discurso.

Entre 2010 e 2011, o mundo então testemunhou o potencial da missão do Facebook (e de tantas outras empresas de tecnologia, em especial o [Twitter](#)). Durante a “Primavera Árabe”, protestantes usaram as redes sociais para superar barreiras de coordenação política impostas por regimes autoritários no Oriente Médio, culminando na [democratização](#) (ainda que temporária) de alguns países da região. Ou seja, concretizando a profecia de seus fundadores, as redes sociais serviram como motores do que Francis Fukuyama um dia chamou de “[o fim da história](#)”, movendo as correntes da política em direção ao Estado Democrático de Direito. O pressuposto para esta missão, portanto, era justamente construir espaços online propícios ao debate público, relegando as interações privadas ao segundo plano.

Os efeitos desta opção estrutural podem ser identificados até mesmo nos julgados dos principais tribunais do mundo, como os da Suprema Corte dos EUA. No caso *Packingham v. North Carolina* de 2017, os ministros da Corte Americana julga-

ram inconstitucional, por violar a primeira emenda à Constituição estadunidense, a legislação do estado da Carolina do Norte que proibia que “agressores sexuais condenados” acessassem redes sociais. Escrevendo em nome da Corte, o *Justice* Anthony Kennedy ponderou que redes sociais como Facebook e Twitter são as “praças públicas” da era digital. Assim, ao proibir que certas pessoas tenham acesso aos seus serviços, o estado da Carolina do Norte estaria “prevenindo os usuários de participar do exercício legítimo de seu direito [à liberdade de expressão]”.

Como era de se esperar, esta percepção de que o Facebook seria o equivalente a uma “praça pública” deixou a empresa mais vulnerável a críticas a respeito da gestão daquele ambiente online. Dois episódios em especial abalaram as estruturas da rede social. Primeiro, em 2016, o Facebook foi usado por *trolls* russos e por uma [agência](#) ligada ao Kremlin, a *Internet Research Agency* - IRA, levando a uma polarização do debate eleitoral nos EUA, potencialmente com a intenção de enfraquecer as instituições democráticas do país. Segundo, em 2018, o jornal britânico *The Guardian* [revelou](#) que a empresa Cambridge Analytica teve acesso a dados pessoais de milhões de usuários do Facebook sem obter regularmente seu consentimento, explorando uma vulnerabilidade na API reservada pela plataforma para desenvolvedores de aplicativos, e que estava usando essas informações para oferecer serviços de assessoria política para candidatos como Donald Trump.

Em menos de uma década, o sentimento de entusiasmo pelas transformações sociais oferecidas pelas redes sociais foi substituído por um sentimento de preocupação com os efeitos negativos causados à sociedade, em especial em termos de desinformação e privacidade. Desde então, diversos países passaram a debater e adotar leis para regular as redes sociais, buscando neutralizar alguns dos seus principais riscos. E é justamente neste contexto que Mark Zuckerberg [anunciou](#), no início de 2019, uma verdadeira mudança de paradigma para a missão institucional do Facebook. A partir daquele momento, ele buscava reorientar sua empresa, afastando-a da figura de “praça pública” e aproximando-a de uma “sala de estar digital”.

No mesmo sentido, menos de dois anos depois, após a insurreição armada protagonizada por apoiadores de Donald Trump no Capitólio em 6 de janeiro de 2021, Zuckerberg [afirmou](#) que estava pensando em novas formas de diminuir a circulação de conteúdo político no feed de notícias. Em suas palavras, “as pessoas não querem que a política e a briga assumam a dianteira do serviço”. As novas mudanças no algoritmo foram implementadas em agosto de 2021 nos EUA e em países como Brasil, Canadá, Suécia e Espanha. Mais recentemente, a empresa [anunciou](#) que passaria a testar o novo algoritmo em mais de 80 países, transformando este projeto que nasceu de uma reação aos eventos eleitorais nos EUA em uma verdadeira experiência global.



Ou seja, ao invés de promover postagens públicas, especialmente políticas, que dão a impressão de que o feed de notícias é uma versão digital dos espaços públicos que conhecemos no “mundo offline”, o Facebook passou a focar em comunicações privadas e até mesmo criptografadas (como ocorre no WhatsApp, também subsidiária da Meta). A ideia, portanto, é compartimentalizar a rede social, fazendo com que os usuários interajam cada vez mais com grupos menores de pessoas que eles conhecem ou que compartilham dos seus interesses. Nestes espaços, Zuckerberg disse que espera que as pessoas tenham discussões não apenas mais efêmeras, mas também protegidas de terceiros. Para [ele](#), “as comunicações focadas na privacidade se tornarão ainda mais importantes do que as plataformas abertas de hoje”.

Os grupos do Facebook, então, passaram a receber um cuidado especial por parte da empresa e seu CEO. Em [entrevista](#) para o The New York Times em abril de 2019, o fundador da rede social anunciou um “redesign” do aplicativo para promover conversas e interações em grupos ao invés de postagens no feed de notícias. Ao abrir o Facebook, o usuário já encontra um atalho para a função “grupos” na aba principal. Para Zuckerberg, o fortalecimento de comunicações seguras em espaços privados terá o efeito de “criar uma plataforma mais confiável” para os usuários, na tentativa de superar as chagas deixadas pelos escândalos dos últimos anos. Outro aspecto dessa mudança é a aposta da empresa em métodos de comunicação passageiros, como é o caso dos stories do Instagram que desaparecem 24 horas depois de serem publicados pelo usuário.

Entretanto, para alguns especialistas como [Jennifer Grygiel](#) da Universidade de Syracuse, esta é, na verdade, uma estratégia para diminuir a quantidade de conteúdos públicos que a companhia precisa moderar e, assim, se eximir de uma parcela de sua responsabilidade. Ademais, como o próprio Zuckerberg reconhece, a mudança para meios mais privados de comunicação envolve um sensível *trade-off*, uma vez que torna-se mais difícil identificar e neutralizar conteúdos danosos e problemáticos, justamente em razão da descentralização da moderação de conteúdo, isto é, do processo pelo qual a plataforma aplica seus termos de uso e padrões da comunidade. Embora a empresa continue monitorando seus serviços, especialmente por meio de seus algoritmos, a linha de frente passa a ser ocupada pelos administradores e moderadores dos grupos sendo este o objeto central do presente relatório.

Um [exemplo](#) envolvendo o próprio CEO ajuda a ilustrar este último ponto. Em julho de 2021, Zuckerberg entrou em um grupo do Facebook reservado para amantes de cachorros, onde os membros trocam dicas e experiências sobre a convivência com seus animais domésticos. Entretanto, em seu primeiro post no grupo, o fundador da rede social se limitou a anunciar novas funcionalidades que sua empresa havia introduzido aos grupos e, assim, violou uma das regras daquela comunidade:

sempre poste uma foto de seu cachorro. Os membros, então, começaram a debater nos comentários sobre o possível banimento de Zuckerberg. Uma das usuárias comentou: “Podemos bani-lo? Imagina o poder disso!”. No mundo descentralizado dos grupos do Facebook, que busca resgatar a natureza privada das comunicações interpessoais na Internet, até mesmo o dono da plataforma pode ser banido por violar as regras específicas de uma dada comunidade.

Mas é claro que os problemas vão além disso. Como [noticiou](#) o The Washington Post ainda em 2019, “a estratégia de promover interações fechadas dentro dos grupos está cada vez mais colidindo com a capacidade da empresa de monitorar aqueles que violam suas regras contra discurso de ódio, assédio e outros males”. Ademais, ao contrário de postagens públicas que podem ser monitoradas e examinadas por meio de ferramentas como o *Crowd Tangle*, grande parte das interações em grupos fechados e secretos estão fora do radar de acadêmicos e especialistas. Isto é problemático quando se leva em conta que, segundo números da própria empresa em 2019, cerca de [1.4 bilhão de usuários](#) interagem em grupos mensalmente. Isto propiciou, até pouco tempo atrás, a proliferação de comunidades que espalhavam [teorias da conspiração](#) sobre as eleições nos EUA, a vacina contra a COVID-19, entre diversas outras questões. Ademais, segundo [acadêmicos](#) que estudam desinformação, usuários em grupos com interesses comuns geralmente não denunciam o conteúdo de seus pares, diminuindo a probabilidade de identificação de violações por parte do Facebook.

### 3 FERRAMENTAS DE MODERAÇÃO DE CONTEÚDO E DIVISÃO DE TRABALHO NOS GRUPOS

Os desafios identificados na seção anterior não significam, entretanto, que os administradores e moderadores de grupos não podem agir de forma eficiente diante de possíveis violações aos padrões da comunidade. Pelo contrário, o movimento de descentralização que marcou o processo de ascensão de grupos no Facebook nos últimos anos também foi acompanhado da implementação de novas soluções tecnológicas por parte da plataforma para facilitar a moderação de conteúdo nestas comunidades.

Nesta terceira seção, assim, antes de apresentar alguns dos resultados das entrevistas conduzidas com líderes de grupos brasileiros na plataforma, haverá espaço para uma sistematização das principais inovações introduzidas pelo Facebook com o objetivo de tornar o monitoramento nos grupos mais eficiente, além da própria atuação da plataforma para evitar violações aos padrões da comunidade no ecossistema de grupos. Na sequência, será analisado o dia a dia da moderação de conteúdo pelos moderadores e administradores dos grupos entrevistados, dando destaque ao uso destas ferramentas disponibilizadas pela plataforma.

### 3.1 Sistematização das Inovações Tecnológicas de Moderação em Grupos e a Atuação do Facebook

O Facebook vem demonstrando uma maior preocupação com a saúde da interação entre os membros de seus grupos desde 2020, quando a empresa introduziu uma ferramenta chamada "[assistente do administrador](#)" ou "*admin assist*" no mês de outubro. Segundo a própria plataforma, trata-se de uma funcionalidade presente em todos os grupos do Facebook que "permite configurar controles de conteúdo que são executados automaticamente". Trata-se de ferramenta que permite uma melhor integração entre os Padrões de Comunidade do Facebook e as regras internas definidas pelos administradores dos grupos. O principal bônus da ferramenta é a redução de análises manuais que os administradores e moderadores precisam fazer em seus grupos.

Uma das funções do "assistente do administrador" - a qual, como ficará claro a seguir, é utilizada pela maioria dos grupos entrevistados para esta pesquisa - é a filtragem de postagens e comentários por meio da definição prévia de palavras-chave. Ou seja, a depender da realidade e do contexto no qual uma determinada comunidade está inserida, o administrador do grupo pode selecionar quais são as palavras-chave que indicam conteúdos que devem, a princípio, ser moderados.

Isso é especialmente útil para minorias políticas que podem se sentir lesadas pelo uso de determinados termos, contribuindo para a construção de um ambiente digital seguro. Ainda, segundo [reportagem](#) do TechCrunch, "isso pode ser usado para filtrar certos conteúdos inapropriados, como linguagem chula, mas também para barrar a aprovação de posts que não sejam pertinentes, o uso de palavras e hashtags associadas a movimentos perigosos, como o QAnon, ou até aqueles posts que querem usar o grupo para fazer anúncios".

Na mesma época, o Facebook também [apresentou](#) outras duas novas ferramentas. A primeira, conhecida como "*administrador de colaborador de marcas*" ou "*brand collabs manager*", permite que grupos públicos colaborem com marcas que pretendem anunciar seus produtos e serviços naquele espaço, promovendo, assim, a monetização de sua audiência. A segunda consiste na introdução de um currículo por parte do Facebook para certificar administradores de grupos. Ou seja, a plataforma passaria a oferecer uma certificação de "administração de comunidades" ou "*community management*" que indica a competência de um determinado administrador para construir e apoiar comunidades digitais na rede social.

No final de 2020, o Facebook voltou a apresentar novidades para os administradores e moderadores de grupos. Em novembro veio o [anúncio](#) de uma nova função batizada de "perguntas e respostas" com o intuito de promover engajamento entre os usuários. Isso permite que os grupos organizem sessões de perguntas e respostas

em tempo real, o que é especialmente útil para grupos que possuem uma finalidade prática, como o compartilhamento de conhecimento sobre um determinado assunto. Nas palavras da empresa, as perguntas e respostas “são colocadas em uma pilha deslizável de cartões que qualquer pessoa acompanhando pode ler e interagir”.

Já em dezembro a empresa [informou](#) a disponibilização dos “alertas de moderação” como uma seção separada dentro da aba de ferramentas do administrador nos grupos. O primeiro alerta que faz parte desta seção já foi mencionado anteriormente; trata-se do “alerta de palavras-chave”, por meio do qual o administrador é notificado sempre que palavras ou frases previamente selecionadas aparecem no grupo. Já o segundo alerta foi introduzido pela primeira vez em dezembro de 2020 e é conhecido como “alerta de engajamento”. Segundo o Facebook, ele “permite que você receba notificações quando um conteúdo em seu grupo está recebendo várias reações e comentários”. Isso permite que o administrador preste atenção àquelas postagens que estão crescendo em engajamento rapidamente e que precisam ser moderadas com um maior cuidado.

Estas ferramentas têm como objetivo primordial tornar a moderação dos grupos pelos seus administradores mais eficiente. Isso ajuda a manter estas comunidades seguras, com uma menor incidência de conteúdos abusivos, como spam e discurso de ódio. Ademais, é uma forma de garantir que os administradores não se sintam sobrecarregados pelo trabalho de moderação e, com isso, abandonem suas comunidades. Nesse sentido, ainda em 2020 e com o intuito de facilitar o acesso a essas ferramentas, o Facebook [anunciou](#) a criação de um novo dashboard conhecido como “*admin home*” para centralizar todas as funcionalidades disponíveis aos administradores.

Dentre as tecnologias disponíveis no dashboard está a possibilidade do administrador restringir certos perfis de postarem e comentarem no grupo, como pessoas que acabaram de ingressar no Facebook ou então usuários que violaram as regras do grupo anteriormente. Outra funcionalidade consiste na imposição de limites à frequência de comentários que cada membro pode formular. Por exemplo, o administrador pode estipular que cada membro só é autorizado a publicar um novo comentário a cada dez minutos, o que pode servir como um freio para conversas conflituosas, promovendo, assim, debates mais civilizados dentro da comunidade.

Dentro do “*admin home*”, os administradores também podem acessar um relatório individualizado sobre a atividade de cada membro, como o número de postagens e comentários feitos ao longo do tempo e quantas vezes foram identificadas violações das regras do grupo naqueles conteúdos. Isso pode ser usado pelo administrador para remover certas pessoas da comunidade. O Facebook também informa

ao administrador neste *dashboard* quantas vezes os conteúdos compartilhados no grupo violaram os padrões da comunidade da plataforma, permitindo a apresentação de recursos para uma revisão da decisão de moderação.

Para além dessas funcionalidades, a própria plataforma adota algumas medidas preventivas para evitar a violação de suas regras no ecossistema de comunidades na rede social. Em 2020, por exemplo, o Facebook [anunciou](#) que administradores e moderadores que tenham seus grupos banidos por violar as regras da plataforma seriam impedidos de criar novos grupos por um período de 30 dias. Ainda, membros de grupos que violarem os padrões da comunidades precisarão de aprovação prévia de suas postagens pelo administrador ou moderador do grupo pelos próximos 30 dias. Por fim, a empresa também informou que passaria a exigir que todos os grupos na plataforma tenham administradores ativos sob pena de serem arquivados.

Em 2021 foram anunciadas ainda mais novidades em relação ao ecossistema de grupos no Facebook. Em abril, a empresa [introduziu](#) a possibilidade dos conteúdos compartilhados nos grupos serem organizados em tópicos que seriam criados com base numa *hashtag* ou palavra-chave contida na publicação. Outra novidade foi a possibilidade dos administradores e moderadores criarem [salas de chat](#) no Messenger vinculadas ao grupo. Estas salas podem ser usadas para duas principais finalidades. Primeiro, possibilitar interações mais reservadas entre os membros do grupo. Segundo, oferecer um espaço dedicado onde o administrador e sua equipe de moderadores podem interagir em tempo real para discutir casos de moderação.

De acordo com uma [análise](#) do TechCrunch, os chamados “*community chats*” ou “chats da comunidade” são úteis para grupos maiores que possuam um “time de pessoas que precisa estar na mesma página sobre decisões de administração da comunidade, incluindo aquelas que devem ser feitas rapidamente em reação a conteúdos que precisem do seu *input*, monitoramento e remoção”. Como será abordado abaixo, essa era uma necessidade apontada por diversos líderes de comunidades no Facebook Brasil que antes precisavam se valer de outras ferramentas, como grupos no WhatsApp, para alinhar protocolos de moderação de conteúdo entre administradores e moderadores.

Ainda, em junho de 2021, o Facebook apresentou um terceiro tipo de alerta para a funcionalidade “alertas de moderação”. Trata-se do “alerta de conflito”, uma versão mais sofisticada do “alerta de engajamento” que, com o apoio de uma Inteligência Artificial (IA), é capaz de identificar e avisar o administrador sobre atritos que estejam acontecendo entre membros da comunidade nos comentários de uma determinada postagem. Nada obstante, como bem [notou](#) James Vincent em matéria para o The Verge, esta tecnologia parece seguir a lógica do algoritmo do Facebook responsável por identificar conteúdos abusivos na rede social. Esta ferramenta, como já é notório, possui uma série de limitações, especialmente quando se trata

de conteúdo humorístico e irônico. É dizer, o algoritmo não é eficiente quando se trata de análises contextuais complexas, como a distinção entre uma sátira e um conteúdo de ódio a partir de certos parâmetros culturais e sociais.

Na sequência, em julho de 2021, a rede social [adicionou](#) aos grupos talvez a sua atualização mais polêmica até então. Trata-se da possibilidade dos administradores em alguns grupos selecionados reconhecerem membros mais experientes de suas comunidades como “especialistas do grupo”. Estes especialistas ganham um selo especial, facilitando sua identificação por outros membros. Nas palavras da própria empresa, “os administradores podem contar com especialistas do grupo para organizar sessões de perguntas e respostas, compartilhar perspectivas sobre algum tópico ou tirar dúvidas no grupo”.

Nada obstante, como [apurou](#) a Vice em agosto de 2021, diversos grupos que espalham desinformação sobre a vacina no Facebook têm membros adeptos de teorias da conspiração que receberam o selo de “especialistas”. Isso pode dar um ar de legitimidade e credibilidade para os conteúdos que são compartilhados por essas pessoas, principalmente porque os demais membros não recebem nenhuma informação sobre como funciona o processo de atribuição do selo de “especialista do grupo”. Sem informações contextuais, os usuários podem achar que a própria empresa endossa o ponto de vista daquelas pessoas, gerando um risco real à saúde pública.

Já em outubro de 2021 a empresa fez dois importantes [comunicados](#). Em primeiro lugar, o Facebook passou a rebaixar conteúdos postados nos grupos por usuários que tenham violado as regras da plataforma de forma reiterada, ainda que fora dos grupos. Isso demonstra, em parte, a centralidade que os grupos assumiram na estrutura da rede social nos últimos anos. Em segundo lugar, a empresa também passou a promover uma maior integração entre a moderação de conteúdo feita com base nos padrões da comunidade e a moderação de conteúdo feita com base nas regras do grupo.

Isso se dá por meio da função “*flagged by Facebook*”, onde o administrador consegue ver quais conteúdos do grupo potencialmente violam as regras da plataforma, permitindo, assim, que ele mesmo remova o conteúdo ou solicite uma revisão. Afinal, sempre que o Facebook remove um conteúdo do grupo por violar suas regras, aquela comunidade recebe um “strike” e, com o passar do tempo, o acúmulo de “strikes” pode levar à suspensão ou até mesmo ao banimento do grupo. Essa função, portanto, promove uma maior sinergia entre as duas dimensões de moderação na plataforma e evita a aplicação de penalidades ao grupo como um todo.

É certo que a rede social continuará a desenvolver tecnologias que facilitam e tornam mais eficiente a administração dos grupos. Ainda assim, este panorama das mudanças apresentadas entre 2020 e 2021 demonstra uma crescente preocupação

da empresa com interações mais privadas dentro da plataforma, dando seguimento ao projeto de seu CEO de se afastar de uma imagem de “praça pública” e se aproximar de algo que pareça mais com uma “sala de estar”. Para que isso aconteça, é preciso oferecer aos administradores e moderadores as ferramentas necessárias para, por um lado, fomentar o engajamento entre os membros de seus grupos e, por outro, monitorar e moderar os conteúdos ali compartilhados de acordo com regras específicas da comunidade.

A seguir, serão apresentados alguns dados coletados pelo ITS Rio ao longo de 2021 por meio de uma série de entrevistas com administradores de grandes grupos no Facebook Brasil. Como mencionado anteriormente, os grupos selecionados para este estudo são voltados para minorias políticas no país, como a população LGBTQIA+, negras e negros, mulheres, nordestinas e nordestinos e praticantes de religiões de matriz africana. O panorama de inovações tecnológicas apresentado acima servirá como fio condutor para as discussões a seguir, auxiliando na compreensão das complexidades e nuances envolvidas no dia a dia da moderação de conteúdo nesses grupos.

### 3.2 As Inovações Tecnológicas em Uso: A Moderação em Grupos no Brasil

Em geral, os grupos pesquisados possuem múltiplos administradores e moderadores designados. Os administradores são responsáveis pela organização da comunidade como um todo, ditando, por exemplo, quais serão suas regras e tópicos de discussões. Os moderadores, por sua vez, são escolhidos pelos administradores e auxiliam no monitoramento e moderação do grupo, em especial das postagens e dos comentários. Cada grupo possui autonomia para definição de sua estrutura de administração e moderação, com comunidades com diversos administradores que atuam concomitantemente e outras que apresentam somente um administrador(a), normalmente seu criador, e um time de moderadores com funções limitadas de gestão. Assim, é comum que, conforme os grupos cresçam em número de membros e em engajamento, novos administradores e moderadores sejam incorporados aos seus quadros para dar conta do volume de novos conteúdos compartilhados todos os dias pelos seus milhares de membros.

Dos onze grupos que fizeram parte do estudo, dois são administrados e moderados pela mesma pessoa. É o caso dos grupos “Porque representatividade importa e muito” e “Não é coincidência, é racismo”. Embora já tenha tentado incorporar mais pessoas como administradores e moderadores, a administradora destes grupos relata que há falta de interesse e que muitas pessoas acabam cansando rapida-



mente. Ademais, ela afirma que, em razão da missão de seus grupos, não quer que “venha uma pessoa que torne isso um negócio”.

A administradora dos grupos acima mencionados criou a primeira comunidade quando estava morando em Niterói porque, embora visse pessoas negras nas ruas, sua *“timeline era toda branca”*. Nada obstante, os membros começaram a debater racismo e a pedir que ela compartilhasse sua experiência com os demais. Daí surgiu a ideia de criar uma segunda comunidade dedicada ao debate sobre discriminação e preconceito. Seu objetivo era, portanto, criar um espaço seguro onde as pessoas pudessem denunciar casos de racismo e receber apoio de outros membros. O grupo possui um alto grau de engajamento, mas, segundo a administradora, isso é uma faca de dois gumes: “Tem vários dias que eu penso que queria só descansar, que ninguém tivesse nenhuma situação de racismo, discriminação e intolerância. Infelizmente é muita coisa, muita coisa mesmo. Todo dia tem conteúdo”.

O volume de conteúdos que precisam ser aprovados e moderados foi destacado como um dos principais desafios enfrentados no dia a dia dos grupos entrevistados. O criador do grupo “Pais Pretos Presentes”, uma comunidade cuja “proposta é a criação de filhos pretos, de crianças pretas e o empoderamento dessas crianças”, confessou ter se surpreendido quando descobriu quantas horas dedicava por dia para moderar sua comunidade. O próprio aplicativo do Facebook possui uma função que indica o tempo online do usuário. “Quando eu vi isso, eu fiquei horrorizado que eram umas cinco horas por dia”, afirma. Depois disso, ele passou a ter mais controle sobre seu tempo online e hoje usa o aplicativo por cerca de duas horas por dia para monitorar o grupo.

Em média, os entrevistados afirmaram que gastam entre duas e três horas por dia com moderação em seus grupos. Mas há exceções, como é o caso da administradora do grupo privado “Feministrampos”, uma comunidade que “tem como objetivo o fortalecimento da rede de conexão entre mulheres através de busca e oferecimento de serviços e produtos e repasse de vagas de emprego”. O grupo hoje conta com mais de 120 mil membros. Segundo a administradora, são aprovados cerca de 300 posts por dia, o que exige um mínimo de quatro horas diárias de dedicação. “Me dedico todos os dias da semana, isso inclui final de semana, feriado, férias, etc. Todos os posts precisam passar por aprovação”, assevera.

**Com base nas entrevistas, é possível concluir que há um forte incentivo para que os administradores aprove todas as publicações dos membros de seus grupos.**

O administrador do grupo “LGBTQI+ Resistência pela Democracia”, disse ter feito uma experiência com um grupo de postagem aberta, mas que, em razão do grande volume de conteúdos inadequados por dia, isso não funcionou. Segundo



ele, “até por questão de exclusão de conteúdo e das denúncias de postagem inadequada que poderiam comprometer o grupo, nós fechamos as publicações para que fossem moderadas”. Ou seja, uma elevada incidência de conteúdos inadequados pode levar o Facebook a suspender ou até mesmo banir aquela comunidade, criando um incentivo para que os administradores sejam mais cautelosos a respeito do que é disponibilizado no grupo.

Muitos administradores, inclusive, acreditam que suas reputações estão, de alguma forma, vinculadas ao sucesso de seus grupos. Nas palavras de um dos administradores, “isso gera uma certa ansiedade e então eu preciso estar olhando de vez em quando, dando uma geral [para ver] se tem conteúdo denunciado, se alguém denunciou comentário ou postagem, porque de certa forma é o meu nome em jogo”. O administrador do grupo “LGBT Brasil”, que começou como uma comunidade de Orkut em 2010, disse que “já aconteceu do Facebook retirar [um conteúdo] e dizer que pode remover o grupo do ar”. Ou seja, o administrador é notificado sobre eventuais postagens que tenham sido removidas pela própria plataforma e que podem colocar o seu grupo em risco, o que aumenta a pressão para que continue monitorando a comunidade de uma forma cada vez mais eficiente e ostensiva.

Outros entrevistados se tornaram verdadeiras celebridades digitais e relatam que já foram reconhecidos na rua. O administrador do grupo “A Ponte Para Pretxs!”, é uma dessas pessoas. O grupo privado, criado em 2014 e reservado para pessoas negras, hoje conta com quase 40 mil membros e “trabalha pela promoção de pessoas negras a partir de oportunidades profissionais e educacionais”. Sua comunidade é reconhecida pelo próprio Facebook como uma das mais ativas e significativas do Brasil. Em razão desse reconhecimento, o administrador desta comunidade já foi convidado pela rede social para compartilhar sua experiência durante o [Facebook Communities Summit](#) de 2020, um evento online global onde a empresa anuncia e debate as principais novidades relacionadas ao universo dos grupos na plataforma.

Ademais, para superar os principais desafios da moderação, as equipes dos diferentes grupos entrevistados criaram diferentes soluções e formas de divisão de trabalho. Perguntado sobre a atuação dos demais administradores e moderadores, Daniel afirmou que eles optaram por um modelo de “segmentação”, onde ele fica responsável pela aprovação das postagens pendentes, outros focam na moderação de comentários e um terceiro grupo se debruça sobre as denúncias recebidas dos membros do grupo. Além disso, o grupo “LGBTQI+ Resistência pela Democracia” também conta com um endereço de email específico para o recebimento de denúncias. Embora o próprio Facebook disponibilize uma ferramenta para isso, o administrador desta comunidade comenta que chegava a receber mais de 50 mensagens por dia em seu inbox com questionamentos sobre a moderação, o que o fez perceber a necessidade de criar um canal dedicado.

Já outros grupos - como é o caso do "Umbandistas", uma comunidade para praticantes e simpatizantes da religião de matriz africana no Brasil - criaram grupos de WhatsApp entre os administradores e moderadores para melhor coordenar suas atividades. Segundo a administradora do grupo, a equipe do grupo usa o aplicativo de mensageria privada para discutir casos de moderação de conteúdo. Ou seja, sempre que há dúvida sobre o que fazer, o administrador ou moderador levanta a questão no aplicativo de mensageria privada para que todos possam comentar e discutir. Como visto acima, recentemente o Facebook identificou essa necessidade por parte dos líderes de comunidades e introduziu a funcionalidade "*community chats*", o que permite que administradores e moderadores organizem suas atividades de forma colaborativa e dialógica na própria plataforma, sem precisar recorrer a outras ferramentas.

Por fim, cumpre ressaltar que um tema que surgiu em diversas entrevistas foi o da importância da monetização em comunidades do Facebook. Como visto acima, muitos líderes de grupos dedicam horas de seus dias para realizar a administração de moderação de suas comunidades. Perguntado se, ao realizar estas atividades, estaria "trabalhando" para a plataforma, o administrador do grupo "A Ponte para Pretxs" afirmou: "sim, porque a qualidade e o sucesso do meu grupo é a qualidade e o sucesso do Facebook, então eu trabalho para ele sim". No mesmo sentido, a criadora e administradora da [comunidade multimídias](#) "Mulheres que Viajam Sozinhas" que aborda o tema do "empoderamento feminino através das viagens", relata que gosta de desempenhar a tarefa de moderação, mas que "às vezes é bem desestimulante" e que deveria ser recompensada de alguma forma pela plataforma. É nesse sentido que o Facebook vem introduzindo ferramentas como o "*brands collab manager*" e a loja do grupo, permitindo que líderes de comunidades possam monetizar sua atuação na rede social.

### 3.3 Eficiência da Possibilidade de Monetização das Comunidades

Desvinculado das ferramentas que auxiliam na administração e moderação de conteúdos em grupos, em novembro de 2021, o Facebook apresentou um mecanismo de monetização dos grupos. A possibilidade de monetizar as atividades das comunidades é uma demanda recorrente por parte de alguns administradores e reflete parte dos anseios dos administradores de comunidades mais ativas da plataforma. A monetização permitida pela plataforma se vincula a oportunidade de abrir uma loja dentro dos grupos para vender diferentes produtos relacionados aos interesses daquela comunidade. Os administradores se tornam os responsáveis pela gestão da loja e, também, pela destinação do dinheiro arrecadado. Alguns grupos, por exemplo, podem optar por doar esses valores para organizações não governa-

mentais, enquanto outros podem usar a função para remunerar seus administradores e moderadores pelo tempo dedicado ao crescimento daquela comunidade digital.

Contudo, a utilização de lojas como forma de monetização das comunidades não foi referenciada pelos administradores quando questionados sobre a sua atuação. Perguntado se, ao realizar estas atividades, estaria “trabalhando” para a plataforma, o administrador do grupo “A Ponte para Pretxs” afirmou: “sim, porque a qualidade e o sucesso do meu grupo é a qualidade e o sucesso do Facebook, então eu trabalho para ele sim”. No mesmo sentido, a criadora e administradora da comunidade multimídias “Mulheres que Viajam Sozinhas” que aborda o tema do “empoderamento feminino através das viagens”, relata que gosta de desempenhar a tarefa de moderação, mas que “às vezes é bem desestimulante” e que deveria ser recompensada de alguma forma pela plataforma.

Observa-se, portanto, que por um lado o Facebook busca atender estes anseios de forma terceirizada, introduzindo ferramentas como o “brands collab manager” e a loja do grupo, com foco no empreendedorismo. Em contrapartida, se observa em uma parcela significativa dos entrevistados, a insatisfação com o aspecto voluntário da administração dos grupos e a busca por uma remuneração por meio de investimentos diretos da plataforma.

## 4 DESAFIOS PARTICULARES DOS GRUPOS E INTERAÇÃO COM A PLATAFORMA

A interação entre os usuários em grupos gera uma série de desafios particulares que muitas vezes escapam às medidas de governança e moderação estabelecidas pela plataforma. Por exemplo, o administrador do grupo “Sou do Nordeste Mesmo e com Muito Orgulho” reportou a ocorrência de postagens em seu grupo com conteúdos que veiculavam violências contra pessoas e animais, assim como comentários de cunho preconceituosos em referência à cultura nordestina.

*A especificidade e necessária contextualização dos comentários e conteúdos compartilhados muitas vezes impede o reconhecimento da violação dos Padrões de Comunidade pelo algoritmo da plataforma, o que leva a uma maior necessidade de atuação e supervisão pelos administradores e moderadores dos grupos.*

O aumento da especificidade das violações aos padrões de comunidade da plataforma, inclusive, é apontado como um dos principais desafios enfrentados pelos gestores. O administrador do grupo “Pais Pretos Presentes” revela as dificuldades inerentes à sua atividade de manutenção de um espaço coletivo seguro ao passo que recebe um grande número de postagens inadequadas de usuários. Nesse contexto,



os administradores entrevistados concluem majoritariamente que o sistema de filtragem de conteúdos a partir da definição de palavras-chave é de extrema utilidade, uma vez que permite o exercício da moderação de conteúdo de forma particularizada em relação às nuances dos ataques recebidos. Contudo, em que pese a eficiência desta ferramenta, ainda há relatos de desafios particulares à gestão de grupos com alto engajamento na rede social. Primeiro, especificamente em grupos vinculados a pautas políticas e sociais, muitas vezes se observam ataques coordenados com o intuito de ridicularizar o conteúdo de outros membros. Conteúdos e atividades que causam danos a grupos minoritários circulam na plataforma através de diferentes métodos, não necessariamente vinculados somente a mensagens, comentários, ou postagens. Muitas vezes são utilizados símbolos, memes e emojis que podem mascarar discursos de ódio e ataques aos membros do grupo.

São exemplos a utilização da reação de risada ou raiva em publicações e comentários que trazem discussões e relatos importantes e vinculados ao tema central de um grupo. Analisada de forma isolada, a utilização de reações não necessariamente constituem uma violação da liberdade de expressão, contudo, quando utilizados de forma coordenada, podem levar a efeitos incisivos que impedem o livre exercício desse direito fundamental no ambiente digital, uma vez que tende a coibir a plena participação de usuários que preferem se autocensurar para não sofrer ataques coordenados (algo entendido como uma forma de "*chilling effect*", efeito inibidor de expressão). Neste contexto, a ferramenta de "alerta de engajamento", disponibilizada pelo Facebook, traz maior eficiência no tempo de resposta a estes ataques, mas de toda forma induz os administradores e moderadores à necessidade de uma vigilância constante e exaustiva das atividades desenvolvidas pelos usuários e membros de seus respectivos grupos.

A vinculação da métrica de engajamento com a utilização de reações na plataforma condiz com o mecanismo interno utilizado pela plataforma entre 2017 e 2019 para definição dos conteúdos apresentados diretamente na *timeline* de seus usuários. Como noticiado pelo [Washington Post](#) em sua cobertura dos documentos internos vazados à mídia no final de 2021, o Facebook estabeleceu um algoritmo que define o alcance de determinados conteúdos em sua rede de acordo com o uso dos emojis de reação angariados. Por exemplo, reações de risada e raiva eram consideradas até cinco vezes mais relevantes pois mantêm usuários mais engajados com a rede.

Dessa forma, conteúdos que levassem a reações mais obtusas de seus usuários passaram a ser promovidos com maior frequência e alcance no *feed* individual de cada usuário causando um efeito cascata que aumentava a visibilidade de conteúdos controversos na plataforma. A problemática desta estratégia de impulsionamento de conteúdo se vincula ao fato de que, de acordo com os cientistas de dados do

próprio Facebook, postagens controversas e com alto uso de reações diversas da tradicional 'curtida' são mais propensas à veiculação de desinformação, conteúdos tóxicos e notícias de qualidade duvidosa.

No que concerne à moderação de conteúdo em grupos privados, a utilização dos diferentes ícones de reação em determinadas postagens não se vincula à propagação direta de conteúdos essencialmente danosos, mas está interligada com ataques que buscam deslegitimar pautas e posicionamentos importantes no âmbito interno de cada grupo. Um exemplo de ataque ocorreu no grupo "Sou do Nordeste Mesmo e com Orgulho" que teve uma *live* invadida por uma ação coordenada de um grupo de "comédia" que se organizou através de um evento similar em outra plataforma para comentar e reagir ao conteúdo.

Problemas de 'invasão' como o narrado acima iluminam um segundo desafio reportado pelos administradores entrevistados. A administração de grupos coletivos na plataforma está sempre sujeita à possibilidade de sequestros de conteúdo, seja de forma direta ou indireta. O administrador do grupo "Sou do Nordeste Mesmo e com Orgulho", por exemplo, indica a dificuldade de lidar com usos inadequados do espaço para a promoção de eventos e pautas divergentes do objetivo geral do grupo. O administrador relata a publicação de convites para eventos em outras plataformas, postagens fomentando discussões de cunho político-partidário alheias ao objeto do grupo, além do compartilhamento de links para sites de captura de dados pessoais por membros da comunidade. A utilização do engajamento e participantes do grupo para pautas que estão aquém dos objetivos e padrões de comunidade estabelecidos pelos líderes do grupo muitas vezes gera insatisfação entre os criadores e administradores destas comunidades.

Tentativas mais diretas de invasão de grupos também fazem parte do relato de alguns entrevistados. O administrador do grupo "LGBT Brasil", relata como sua comunidade foi "hackeada" através de um membro que havia sido elevado à categoria de administrador do grupo. Com os privilégios de administrador, o usuário retirou os demais administradores e moderadores e alterou o nome do grupo para um título com conteúdo homofóbico. Ou seja, o grupo sofreu uma espécie de "sequestro digital" e teve suas pautas ressignificadas para atacar uma minoria política que antes se organizava a partir daquele espaço para debater pautas de pertencimento e orgulho.

Frente a estes "ataques", o Facebook possui mecanismos limitados para prevenção e remediação, [sugerindo](#) que o papel de moderador e administrador seja designado somente para contas de pessoas conhecidas e de confiança. Nada obstante, a perda de poderes administrativos em grupos normalmente ocorre através do hackeamento dos perfis de outros administradores e não do grupo em si. Para a recuperação do acesso, as vítimas devem contatar diretamente a plataforma, tanto

para redefinir as senhas e padrões de segurança de sua conta pessoal quanto para alertar a plataforma acerca de atividade ilegítima na administração do grupo.

As limitações da plataforma neste sentido iluminam um aspecto que os entrevistados enxergam como paradoxal nas suas relações com a plataforma. Notam existir um misto de apreciação e tensão entre a sua atuação como administradores e a rede social como um todo. Se, por um lado, os investimentos realizados pelo Facebook para facilitar a moderação de conteúdo pelos administradores são vistos com bons olhos; por outro, a participação conjunta do Facebook na moderação de conteúdo, através de suas ferramentas algorítmicas (como nota-se abaixo), muitas vezes gera frustrações.

O administrador do grupo “LGBTQI+ Resistência pela Democracia”, por exemplo, indica descontentamento com o que entende como unilateralidade das decisões de moderação de conteúdo realizadas pela plataforma com base nos Padrões de Comunidade. No pensamento do administrador, a moderação do discurso de ódio indica problemas algorítmicos, uma vez que o Facebook exclui rapidamente conteúdos quando verifica a presença de certas palavras, muitas vezes sem analisar o contexto em que elas foram utilizadas. No âmbito de sua comunidade, o administrador observa que muitas vezes usuários adequam o seu discurso, evitando o uso de determinadas palavras mesmo que mantendo, substancialmente, uma posição de ataque e ódio. Em contrapartida, o mesmo observa que usuários que buscam combater o discurso de ódio de forma transparente, identificando as palavras e termos pejorativos, têm seu conteúdo excluído indevidamente.

A problemática não parece estar adstrita a um grupo específico. A administradora do grupo “Mulheres que Viajam Sozinhas”, também relata insatisfação com a retirada automática de conteúdo sem a apresentação de motivos claros e fundamentação para a moderação. Em sua perspectiva, apesar do Facebook realizar um significativo investimento no uso de soluções algorítmicas para uma melhor moderação das suas mais variadas comunidades, a empresa deve investir no componente humano que integra a plataforma, considerando a necessidade de contextualização para lidar com as múltiplas sutilezas da linguagem, assim como investir diretamente no reconhecimento do trabalho realizado por administradores de grandes grupos.

O problema que subjaz a interação com a plataforma decorre da busca de um maior reconhecimento, pelos administradores, do trabalho realizado para a manutenção das comunidades. Entre os entrevistados, parece haver um consenso de que a moderação de conteúdo pelo Facebook em grupos coletivos deve ser vista como uma forma de intervenção que depende do acompanhamento e avaliação dos administradores e moderadores de sua pertinência. Ao passo que o Facebook mantém suas atividades de moderação mesmo em referência aos conteúdos dispostos em

comunidades administradas, na visão dos entrevistados essa prerrogativa deve ser exercida com a consideração e *input* dos administradores.

Em contrapartida, o Facebook reconhece administradores de grupos de grande impacto em sua plataforma com a classificação de '*power adminis*'. Estes, são administradores reconhecidos pelo caráter dinâmico de suas comunidades e são incluídos em grupos fechados geridos por funcionários do Facebook. Nestes grupos, '*power adminis*' têm acesso à novas ferramentas implementadas pela plataforma para a gestão de comunidades, e podem inclusive testar e dar *feedback* quanto à efetividade e possíveis melhorias às novas ferramentas. Atualmente, o Facebook possui quatro grupos de '*power admins*' distribuídos regionalmente entre América do Norte, Reino Unido e Irlanda, Brasil, e DACH (Alemanha, Áustria, e Suíça), sendo o Brasil o único país com um grupo exclusivo.

Portanto, observa-se uma tentativa da plataforma de estabelecer um vínculo direto com os administradores de grandes comunidades, dando-lhes uma forma de reconhecimento, ainda que os entrevistados e entrevistadas entendam que mais pode ser feito aprofundar essa relação e serem melhor reconhecidos.

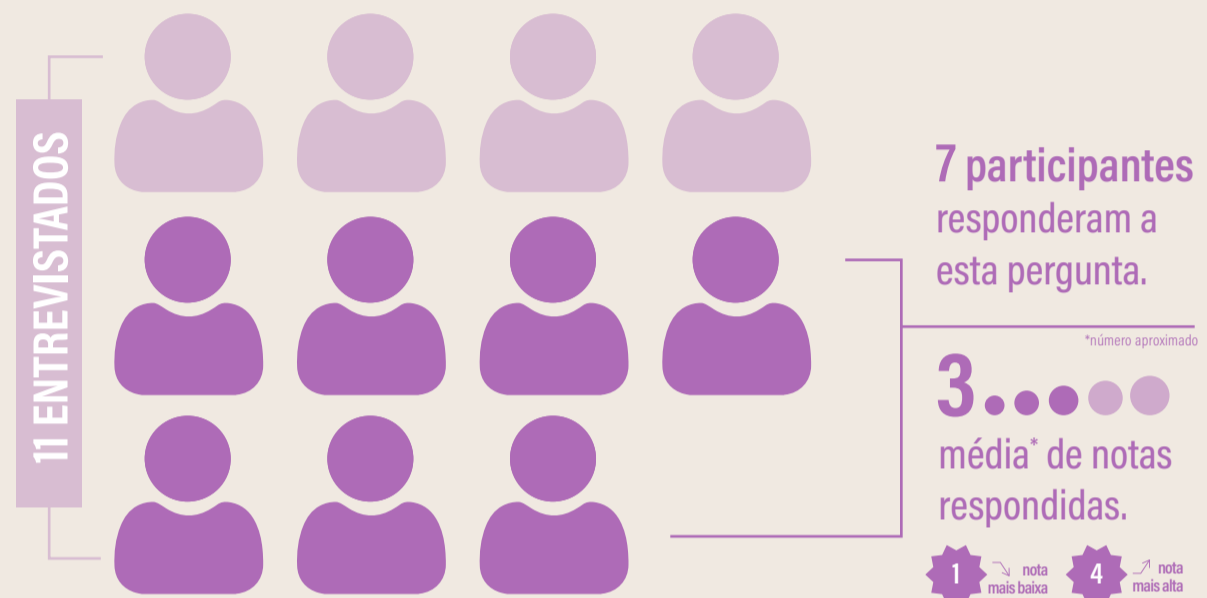
## 5 PERSPECTIVAS DE ADMINISTRADORES E MODERADORES

Como visto na primeira seção deste relatório, ao final das entrevistas os líderes das comunidades que fizeram parte desta pesquisa tiveram a oportunidade de avaliar diversos aspectos da moderação de conteúdo em grupos e no Facebook como um todo. É importante destacar, desde já, que alguns participantes não se sentiram confortáveis em responder algumas das perguntas de avaliação. Todas elas pediam que os entrevistados dessem uma nota de um a cinco. Embora os resultados não sejam necessariamente generalizáveis - em razão do número limitado de participantes e da falta de algumas respostas -, passa-se agora a uma análise dos principais pontos que foram levantados na parte final das entrevistas e quais são os aprendizados que podem ser deduzidos para o ecossistema da moderação de conteúdo em plataformas digitais.

Em primeiro lugar, os participantes foram perguntados sobre a atuação do próprio Facebook em termos de moderação de conteúdo em grupos e páginas na plataforma, principalmente em situações que envolvem violência, assédio, discurso de ódio, dentre outros. Considerando os sete participantes que deram uma nota para a atuação da plataforma, a média das respostas ficou em aproximadamente 3 pontos, sendo 1 a nota mais baixa e 4 a nota mais alta.

## PERGUNTA 01:

Atuação do Facebook na moderação de conteúdo em situações que envolvem violência, assédio, e discurso de ódio.



A partir das entrevistas foi possível perceber que os líderes de comunidades ainda possuem muitas ressalvas a respeito da moderação desempenhada pela própria rede social, principalmente quando há o envolvimento de algoritmos. O administrador do grupo “LGBT Brasil”, se queixou do fato de que o Facebook “apaga umas coisas que você vê que é um robô, então tem um palavrão e ele apaga”. Nada obstante, segundo o administrador, em outros casos termos pejorativos são usados, em especial contra usuários LGBTQIA +, mas por não estarem no radar dos algoritmos da plataforma acabam não sofrendo qualquer restrição imediata.

Na sequência, os entrevistados foram questionados sobre o grau de transparência do processo de moderação de conteúdo desempenhado pelo Facebook. Considerando os seis participantes que deram suas notas, a média das respostas ficou em aproximadamente 3 pontos, sendo 1 a nota mais baixa e 5 a nota mais alta.





## PERGUNTA 02:

Grau de transparência do processo de moderação de conteúdo.



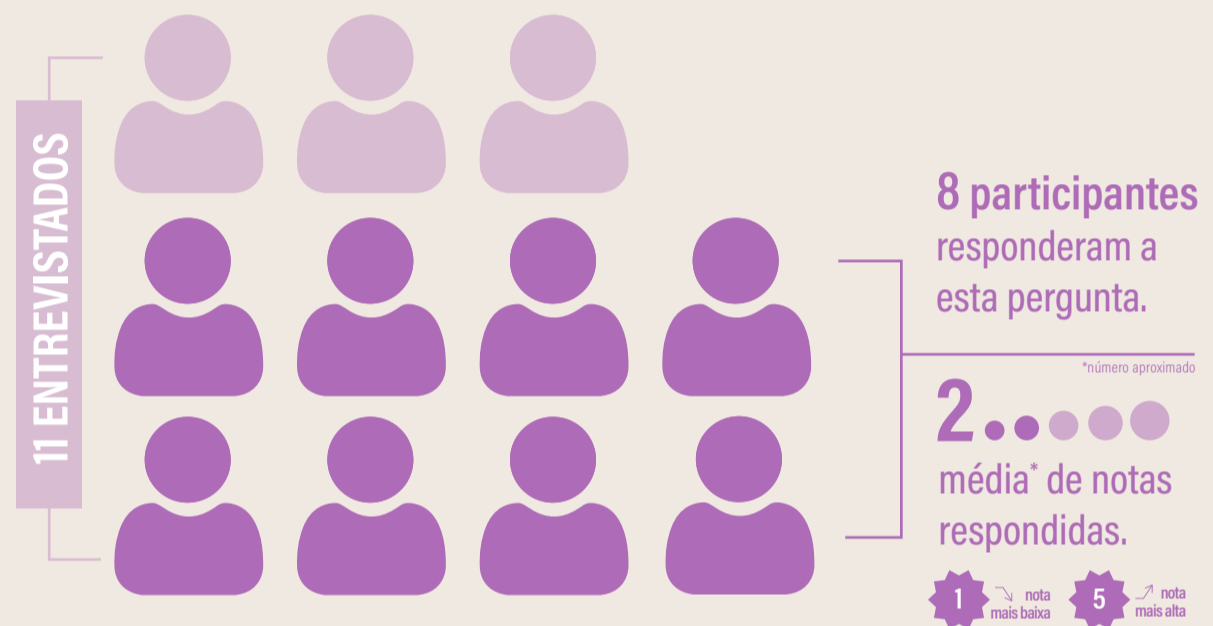
O administrador do grupo “Pais Pretos Presentes”, afirmou que “sempre sabe o que acontece” em termos de moderação na plataforma, o que considera ser algo importante e positivo. Outros líderes, entretanto, mencionaram que é comum ver conteúdos removidos de suas comunidades e não contar com informações suficientes por parte da plataforma que justifiquem a remoção. O administrador do grupo “Sou do Nordeste Mesmo e com Orgulho”, afirmou que o Facebook encaminha um aviso “dizendo que tirou um conteúdo por ser ofensivo”, mas que, ao ler o aviso completo, não consegue entender a razão da remoção dada sua generalidade.

Os entrevistados também foram perguntados sobre o grau de preocupação da plataforma em relação aos usuários que são alvo de discurso de ódio online. Considerando os oito participantes que deram suas notas, a média das respostas ficou em aproximadamente 2 pontos, sendo 1 a nota mais baixa e 5 a nota mais alta.



## PERGUNTA 03:

Grau de preocupação da plataforma em relação aos usuários que são alvo de discurso de ódio online.

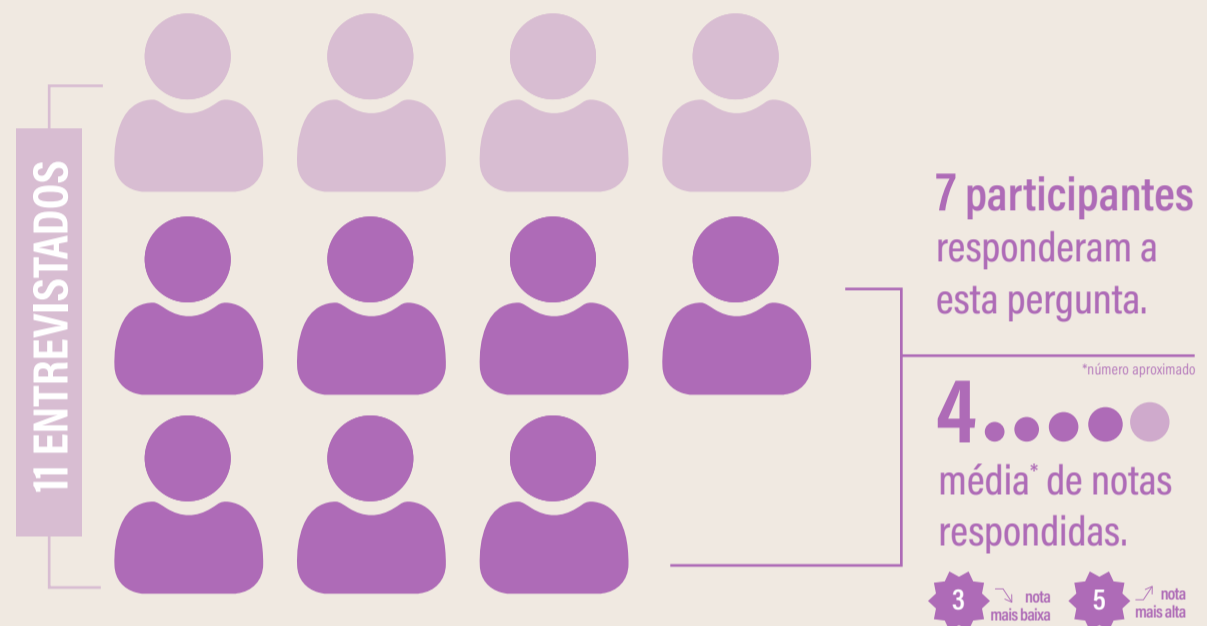


A percepção da maioria dos entrevistados em relação ao grau de preocupação do Facebook ainda é relativamente negativa, mas muitos mencionaram que percebem um esforço da plataforma para melhorar sua atuação. A Administradora da comunidade “Mulheres que Viajam Sozinhas”, por exemplo, diz acreditar, pela sua experiência na plataforma, “que o Facebook vem investindo nisso”. O administrador do grupo “A Ponte Para Pretxs!”, por sua vez, mencionou que a empresa está começando a publicar relatórios e estudos sobre o impacto da desinformação e do discurso de ódio na plataforma, o que considera ser um desdobramento positivo, e que já foi procurado pela plataforma para discutir esse assunto em reunião.

Sobre a atuação de administradores e moderadores em grupos, os entrevistados foram questionados sobre a efetividade da moderação de conteúdo para enfrentar desafios como discurso de ódio em suas comunidades. Considerando os sete participantes que deram suas notas, a média das respostas ficou em aproximadamente 4 pontos, sendo 3 a nota mais baixa e 5 a nota mais alta.

## PERGUNTA 04:

Efetividade da moderação de conteúdo por administradores e moderadores para enfrentar desafios como discurso de ódio.



De todas as perguntas feitas aos entrevistados na fase de avaliação do questionário, essa foi a que recebeu as maiores pontuações, indicando que os participantes valorizam sobremaneira a moderação que é feita pelos líderes de comunidades e acreditam se tratar de uma forma eficiente de enfrentar diversos problemas na plataforma. Nada obstante, algumas ressalvas foram feitas. Segundo a administradora do grupo “Feministrampos”, embora a moderação em grupos seja efetiva, é preciso um apoio direto da plataforma para evitar a sobrecarga dos administradores e moderadores. Outros participantes encaram a eficiência na moderação como uma obrigação. “Eu acho que se você cria um grupo, tem que ter responsabilidade por ele”, asseverou o administrador do grupo “Sou do Nordeste Mesmo e com Orgulho”.

Ainda, os entrevistados tiveram a oportunidade de se manifestar sobre os padrões da comunidade do Facebook e o quanto as políticas da plataforma estão dispostas de forma clara aos usuários. Entre todas as perguntas feitas na fase de avaliação desta pesquisa, essa foi a que contou com menos respostas.

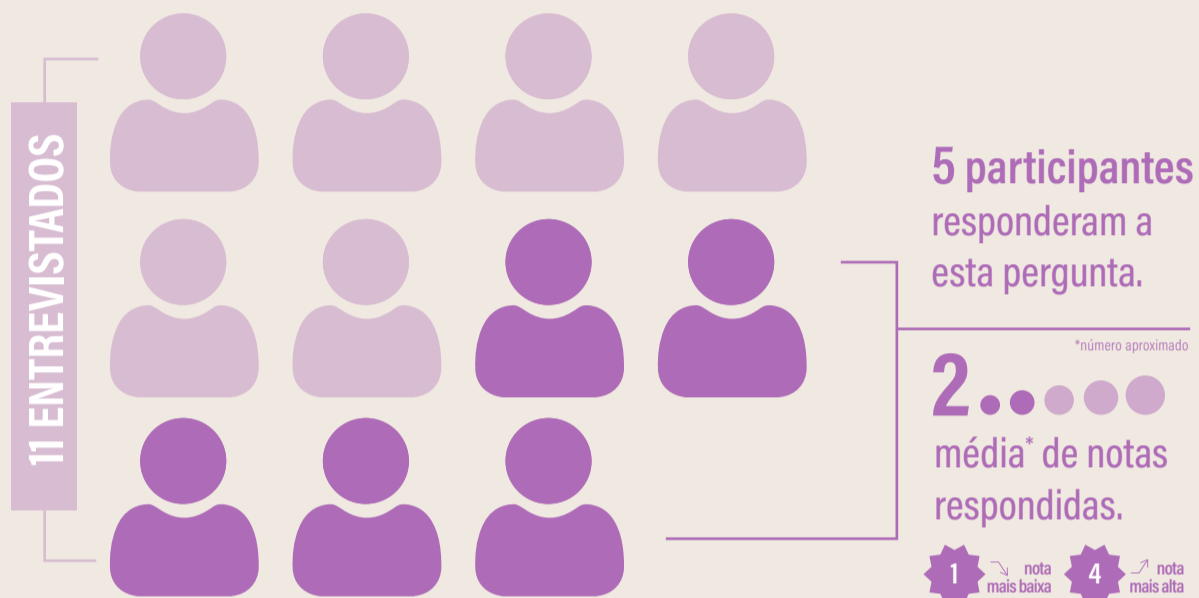
**Metade dos entrevistados alegaram desconhecer os padrões da comunidade e, por isso, preferiram não responder.**



Considerando os cinco participantes que deram suas notas, a média das respostas ficou em aproximadamente 2 pontos, sendo 1 a nota mais baixa e 4 a mais alta.

## PERGUNTA 05:

Clareza dos Padrões de Comunidade da plataforma.

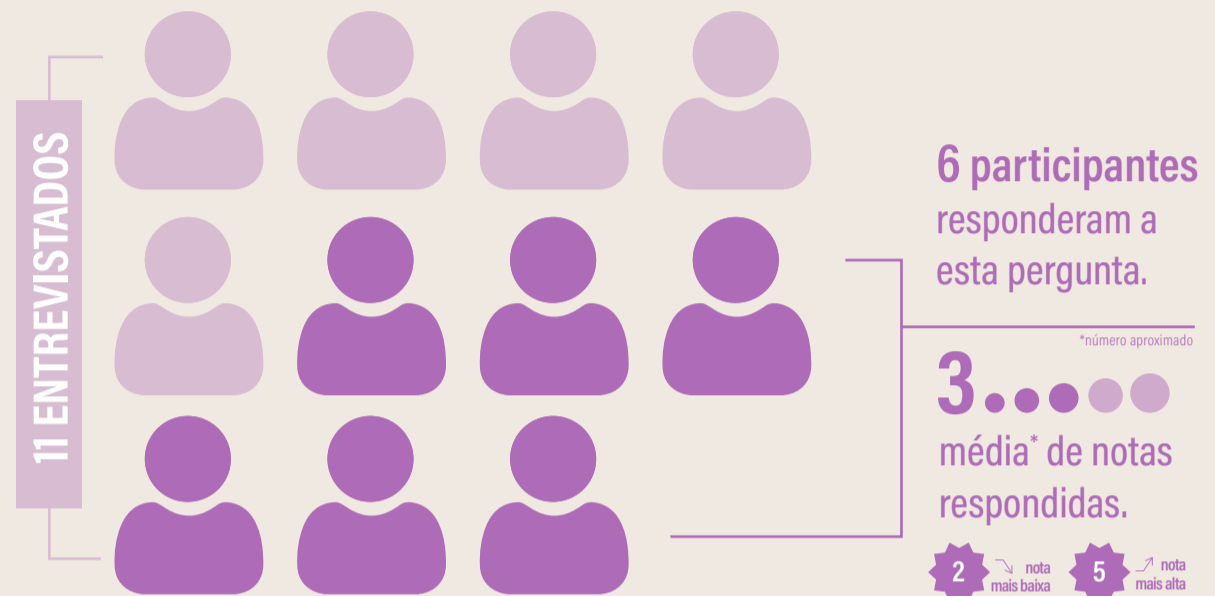


Alguns participantes alegaram não saber onde encontrar os padrões da comunidade e outros se queixaram da falta de clareza em torno das políticas da rede social. Isso pode indicar uma falta de convergência entre a moderação feita pelo Facebook e aquela feita nos grupos pelos líderes de comunidade, uma vez que estes mencionam desconhecer os padrões da empresa. O administrador do grupo "Sou do Nordeste Mesmo e com Orgulho" também indicou um problema de tradução no documento. "Eu acho que eles escrevem em inglês e quando traduzem para cá não faz sentido", deduziu. Trata-se de um problema que foi explorado mais a fundo em outro [relatório](#) do ITS Rio.

Por fim, os entrevistados também foram perguntados sobre suas expectativas a respeito do [Comitê de Supervisão](#), órgão independente criado em 2020 para decidir em última instância casos de moderação de conteúdo envolvendo usuários do Facebook e Instagram. Considerando os seis participantes que deram suas notas, a média das respostas ficou em aproximadamente 3 pontos, sendo 2 a nota mais baixa e 5 a mais alta.

## PERGUNTA 06:

Expectativas a respeito do Comitê de Supervisão.



Sobre o funcionamento do Comitê, o administrador do grupo “Pais Pretos Presentes” destacou o aspecto deliberativo e coletivo da instituição. “Eu acho importante, eu acredito muito nesse crescimento através do outro, a filosofia africana fala muito isso, [...] da importância do coletivo”, avaliou. Outros entrevistados, entretanto, foram mais céticos. É o caso do administrador do grupo “LGBT Brasil” que disse acreditar que a rede social pode simplesmente delegar certas responsabilidades para o Comitê com o intuito de não prestar mais contas. “Eu sou um pouco cético quanto a [...] como vai se dar esse controle independente, [...] quais são os critérios, como isso vai chegar nas pessoas?”, indagou.



## 6 CONSIDERAÇÕES FINAIS

Os esforços do Facebook para incentivar uma maior interação entre usuários através de diversas comunidades fomentou o desenvolvimento de novas ferramentas para auxiliar moderadores e administradores de grupos. Os últimos dois anos foram marcados por inovações implementadas para auxiliar administradores e moderadores de comunidades do Facebook a manterem seus grupos seguros de acordo com suas particularidades.

Dentre as inovações, destacam-se a implementação do “assistente do administrador” (“*admin assist*”), que permite a automação de controles de conteúdo pelos administradores de grupos na plataforma, como por exemplo o estabelecimento de palavras-chave para filtragem de postagens e comentários. Esta ferramenta foi sinalizada por diversos entrevistados como um efetivo instrumento que facilita a moderação de conteúdo em seus grupos, permitindo uma maior eficiência e granularidade em sua atividade de moderação.

Em contrapartida, em que pese a plataforma ter introduzido funcionalidades especificamente voltadas à monetização dos grupos por seus administradores, ainda se observa uma insatisfação entre os entrevistados quanto à possível remuneração pelo trabalho realizado. Muitos líderes dedicam horas de seus dias na manutenção de suas comunidades e alguns entrevistados demonstraram incômodo com o fato de que a plataforma os reconhece somente como voluntários.

Essa insatisfação também se reflete na forma que os administradores entrevistados enxergam o seu papel no contexto maior das atividades de moderação de conteúdo na plataforma. Alguns entrevistados relataram descontentamento com outras ferramentas utilizadas diretamente pelo Facebook para moderar o conteúdo em grupos. A moderação algorítmica é criticada por sua incapacidade de captar a especificidade e contextualização de comentários e conteúdos nestes espaços, o que muitas vezes impede o reconhecimento da violação dos Padrões de Comunidade ou leva a retiradas e bloqueios de conteúdos de forma equivocada.

Por fim, as entrevistas também forneceram um entendimento da percepção dos administradores frente às demais práticas de moderação de conteúdo realizadas pelo Facebook. Em situações de moderação que envolvem violência, assédio e discurso de ódio, os entrevistados demonstraram ressalvas a respeito da moderação desempenhada por algoritmos. Este receio se espelha também quanto ao grau de transparência do processo de moderação, com alguns administradores alegando não receber informações suficientes que embasaram a decisão de remoção pela plataforma.

Em suma, o presente relatório demonstra uma maior preocupação da plataforma em estabelecer relações entre os usuários através de comunidades que compartilham determinados interesses e valores. Neste sentido, observa-se uma interessante dinâmica entre os padrões de comunidade estabelecidos pelo Facebook e a atuação dos administradores das comunidades no Brasil, que aplicam de forma concorrente aquelas normas em relação a estas, as quais, por sua vez, foram criadas especificamente para as suas comunidades.

Como se observou, metade dos entrevistados alegaram desconhecer os padrões de comunidade que pautam a moderação de conteúdo realizada pelo Facebook, o que reflete atritos entre os administradores dos grupos e a atuação da plataforma na moderação de conteúdos compartilhados nestes espaços. Esta dinâmica também pauta parte das insatisfações dos administradores entrevistados, que passam a enxergar a sua relação com a plataforma como algo que excede a diferenciação entre prestador e usuário, se aproximando à categoria de colaborador ou até mesmo funcionário.



## ANEXO 1 - QUESTIONÁRIO DE PESQUISA

O questionário abaixo serviu de base para a condução das entrevistas realizadas no âmbito das investigações do ITS-Rio quanto à moderação de conteúdo no Facebook do Brasil. O fio condutor das perguntas foi a análise da moderação realizada por administradores e moderadores de grupos de redes sociais e a aplicação dos Padrões de Comunidade estabelecidos pela plataforma. As entrevistas foram realizadas entre Novembro de 2020 e Maio de 2021 e refletem as perspectivas dos administradores daquelas comunidades naquele período. Cada pessoa entrevistada consentiu com a divulgação de seu nome completo, assim como do grupo que administra. Ao início de todas as entrevistas, os administradores foram questionados se foram vítimas de qualquer forma de discurso de ódio. Ato seguinte, as entrevistas seguiram com a seguinte estrutura:

### SOBRE A COMUNIDADE

- » Diga o nome do seu grupo no Facebook. Explique a função da comunidade e os principais temas abordados.
- » O seu grupo tem quantos moderadores?
- » O seu grupo tem quantos administradores?
- » Quantos membros fazem parte?
- » Qual é o engajamento da sua comunidade? É uma comunidade muito ativa?



## SOBRE A MODERAÇÃO

---

- » Quantas horas você dedica para a moderação de conteúdo do seu grupo ou página no Facebook?
- » Em geral, quantas horas por dia ou por semana você dedica para o grupo em geral?
- » Você aprova todos os conteúdos de membros antes da publicação ou qualquer membro pode publicar sem aprovação anterior?
- » Como você escolhe o que moderar?
- » Os usuários fazem denúncias automáticas pela plataforma? Ou por mensagem inbox?
- » Você recebe algum tipo de aviso do Facebook sobre conteúdo denunciado?
- » Você ativamente lê as publicações em busca de problemas?
- » Que tipo de problemas você costuma enfrentar no grupo? Que tipo de conteúdo precisa da sua intervenção?
- » Você costuma encontrar muito conteúdo com violência, assédio, discurso de ódio, notícia falsa, nudez, terrorismo, vendas não autorizadas, spam, suicídio ou automutilação? Se sim, quais?
- » Quando encontra algum conteúdo problemático, o que faz?
- » Você debate com os outros moderadores sobre o que fazer?
- » Você se preocupa em tratar igualmente os usuários, isto é, caso um membro faça a mesma coisa que outro, você procura aplicar sempre a mesma regra ou punição?
- » Com qual frequência você exclui conteúdo (por dia ou semana)?



- » Você costuma conversar com quem publicou o conteúdo problemático para que a pessoa reveja sua posição e apague a publicação?
- » Você já excluiu membros do grupo ou puniu algum deles (silenciou temporariamente, por exemplo)? Como foi essa experiência?
- » Você recebe muitas reclamações de usuários insatisfeitos com a moderação? Se sim, qual a frequência (dia, semana, mês)?
- » Como lida com essas reclamações? Isso te abala emocionalmente? Você responde?
- » O Facebook já excluiu conteúdo ou marcou alguma publicação como falsa dentro do ou página sem que você participasse da decisão? Se sim, conte o que aconteceu.
- » Se sim, você concordou com a remoção do conteúdo ou marcação de uma publicação como falsa feita pelo Facebook? Caso não tenha concordado, você recorreu? Qual foi o resultado?
- » Você conhece a função “Qualidade do Grupo do Facebook”? Se sim, com qual frequência você a utiliza?
- » Quando entrou na opção “Qualidade do Grupo do Facebook”, você concordou com as publicações que tinham sido retiradas pelo Facebook? Por quê?
- » Ao notar que uma publicação do grupo tinha sido retirada pelo Facebook que, na sua opinião, não deveria ter sido removida, você entrou em contato com o membro que publicou para que ele recorresse?
- » O seu grupo já recebeu algum ataque ou denunciou conteúdo de terceiros (por conter violência, assédio, discurso de ódio, notícia falsa, nudez, terrorismo, vendas não autorizadas, spam, suicídio, automutilação, entre outros motivos)? Se sim, conte o que aconteceu.



- » Como foi o ataque e/ou o conteúdo denunciado?
- » Qual foi a resposta do Facebook? O problema foi resolvido? Se sim, em quanto tempo?
- » Se a resposta foi negativa, você recorreu? Qual foi o resultado?
- » Caso tenha recorrido e não tenha conseguido reverter a decisão, a empresa explicou os motivos da sua decisão?
- » Na sua percepção, os motivos apresentados pela empresa foram suficientes?
- » Como esse ataque sofrido ou conteúdo denunciado te afetaram emocionalmente e a sua vida?
- » O que você sentiu com essa experiência, inclusive com interação com o Facebook?
- » Você mantém algum contato direto com o Facebook (reuniões, mensagens)?
- » Se sim, com qual frequência você entra em contato com a empresa?
- » Em conjunto com a empresa, vocês estabelecem metas de novos membros do grupo/página a serem batidas?
- » O Facebook já te ofereceu alguma ferramenta (como alarme para palavras-chave) ou qualquer ajuda para facilitar o seu trabalho de moderação (curso, capacitação)?



## SAÚDE MENTAL

---

- » Você sente que esse trabalho de moderação atrapalha a sua vida de alguma forma? Como?
- » Você gosta de fazer essa moderação ou se pudesse não faria?
- » Você considera esse trabalho te toma mais tempo do que você gostaria?
- » Você considera esse trabalho estressante e desgastante emocionalmente?
- » Se tivesse que classificar, de 1 a 5, o quanto você diria que o trabalho de moderação é estressante e desgastante emocionalmente? (sendo 5 muito estressante e desgastante emocionalmente)
- » Você já considerou excluir o seu grupo para evitar o trabalho de moderá-lo ou administrá-lo?



## AVALIAÇÃO

- A) Após essas experiências e também pela sua vivência no Facebook, como avalia a atuação da empresa na moderação de conteúdo nos grupos e páginas, especialmente quando envolve violência, assédio, discurso de ódio, notícia falsa, nudez, terrorismo, vendas não autorizadas, spam, suicídio ou automutilação?
- » Se tivesse que dar uma nota, de 1 a 5, como avalia a atuação do Facebook na moderação de conteúdo, especialmente quando envolve violência, assédio, discurso de ódio, notícia falsa, nudez, terrorismo, vendas não autorizadas, spam, suicídio ou automutilação?
- B) Na sua opinião, o facebook deveria intervir mais ou menos na moderação dos grupos e páginas? Isto é, você está satisfeito de ter o papel de principal moderador o conteúdo no seu grupo ou página?
- C) Você considera que o Facebook dá aos moderadores e administradores todas as ferramentas tecnológicas necessárias para manter a saúde das comunidades em grupos e páginas? Se não, do que sente falta?
- D) Você acha que o Facebook deveria fazer alguma coisa para melhorar ou facilitar a sua experiência e o seu trabalho de moderação de conteúdo no grupo ou comunidade? Se sim, o que?
- E) Você considera que o Facebook apoia o suficiente os moderadores/administradores de grupos, principalmente ajudando a lidar com o investimento de tempo e o custo emocional do trabalho de moderação?

F) Você sente que trabalha para o Facebook?

G) Você acha que, pelo seu trabalho de moderação, o Facebook deveria te compensar de alguma forma (remuneração financeira, auxílio psicológico, suporte tecnológico)?

H) Quais das medidas abaixo você acha que seriam importantes para melhorar a moderação de conteúdo nos grupos e páginas? É possível marcar mais do que uma opção

- (1) Mudar ou detalhar mais as regras de moderação, ou seja, os Padrões da Comunidade
- (2) Investir em inteligência artificial, para que conteúdos nocivos sejam automaticamente reconhecidos e retirados pela plataforma
- (3) Investir na contratação de mais Moderadores de Conteúdo do Facebook e na capacitação desses funcionários
- (4) Outros - Indicar qual

I) Na sua percepção, o processo de moderação de conteúdo do Facebook é transparente?

- » Se tivesse que dar uma nota, de 1 a 5, como avalia a transparência do processo de moderação do Facebook?

J) Na sua percepção, o Facebook se preocupa com usuários que são alvo de discurso de ódio na plataforma?

- » Se tivesse que dar uma nota, de 1 a 5, como avalia o nível de preocupação do Facebook com usuários que são vítimas de discurso de ódio na plataforma?



K) Na sua percepção, o processo de moderação de conteúdo pelos próprios líderes dos grupos e páginas é uma forma efetiva de lidar com o discurso de ódio na plataforma?

- » Se tivesse que dar uma nota, de 1 a 5, como avalia a efetividade do processo de moderação de conteúdo por parte dos administradores/moderadores para lidar com o discurso de ódio nos grupos e páginas?

L) Você já conhecia os padrões da comunidade (regras sobre o que é ou não permitido no facebook)? Acha que está claro para o usuário quais são essas regras?

- » Se tivesse que dar uma nota, de 1 a 5, como avalia a clareza e apresentação dessas regras para o usuário?
- » Quando vai fazer a moderação de conteúdo, você se baseia de alguma forma nos padrões da comunidade ou na sua própria opinião?

M) Em 2018, o Facebook anunciou a criação de um Comitê de Supervisão, com especialistas independentes, para julgar as decisões da empresa sobre moderação de conteúdo. Nas palavras do fundador do Facebook, Mark Zuckerberg, o comitê funcionaria como “um tribunal superior”. Como você avalia a criação desse comitê? Acha que é uma boa ideia? Você acredita que o Comitê vai ser eficiente no combate à desinformação e ao discurso de ódio no Facebook?

- » Se tivesse que dar uma nota, de 1 a 5, qual a sua expectativa sobre a atuação e eficiência do Comitê? (Sendo 1 pouca expectativa e 5 muita expectativa)



# **SOBRE OS AUTORES**

## **João Victor Archegas**

Mestre em Direito Constitucional Comparado por Harvard. Gammon Fellow por mérito acadêmico na Harvard Law School. Professor de Direito na FAE (Curitiba). Pesquisador sênior da área de Direito e Tecnologia do Instituto de Tecnologia e Sociedade do Rio de Janeiro (ITS Rio).

## **Lucas Henrique Conceição**

Doutorando em Direito Internacional na Università Bocconi - Milão. Mestre, summa cum laude, em Direito Constitucional pela Birkbeck College, University of London. Ex-aluno e bolsista por mérito acadêmico do Columbia Summer Program in American Law, na Universidade de Leiden. Pesquisador da área de Direito e Tecnologia do Instituto de Tecnologia e Sociedade do Rio de Janeiro (ITS Rio).

## **Marina Estarque**

Jornalista e pesquisadora, cobre transparência e liberdade de imprensa na América Latina para o Centro Knight de Jornalismo nas Américas, da Universidade do Texas em Austin. Como repórter, trabalhou para veículos como Folha, Estadão, Deutsche Welle, Agência Lupa, Rádio da ONU em Nova York, entre outros. É mestre em edição jornalística pela Universidade da Coruña.

## **Celina Bottino**

Graduada em direito pela PUC-Rio, mestre em direitos humanos pela Universidade de Harvard. Especialista em direitos humanos e tecnologia. Foi pesquisadora da Human Rights Watch em Nova York. Supervisora da Clínica de Direitos Humanos da FGV Direito-Rio. Foi consultora da Clínica de Direitos Humanos de Harvard e pesquisadora do ISER. Associada do Centro de Defesa dos Direitos da Criança e Adolescentes do Rio de Janeiro. Atualmente desenvolve pesquisas na área de direitos humanos e tecnologia coordenando projetos na área de liberdade de expressão e privacidade. É afiliada ao Berkman Klein Center de Harvard e diretora de projetos do ITS.

## **Christian Perrone**

Advogado, Consultor de Políticas Públicas. Pesquisador Fulbright (Universidade de Georgetown, EUA). Doutorando em Direito Internacional (UERJ); Mestre em Direito Internacional (L.L.M/Universidade de Cambridge, Reino Unido). Ex-Secretário da Comissão Jurídica Interamericana da OEA. Coordenador da área de Direito e Tecnologia no Instituto de Tecnologia e Sociedade do Rio de Janeiro (ITS Rio).





Acesse nossas redes



[itsrio.org](http://itsrio.org)