

Reclamações sobre o procedimento de moderação de conteúdo em redes sociais: o que pensam os usuários



iris

INSTITUTO
DE REFERÊNCIA
EM INTERNET
E SOCIEDADE

Reclamações sobre o procedimento de moderação de conteúdo em redes sociais: o que pensam os usuários

AUTORIA

Fernanda dos Santos Rodrigues Silva
Júlia Maria Caldeira Gertrudes
Luiza Correa de Magalhães Dutra
Rafaela Ferreira Gonçalves da Silva

Revisão interna

Paloma Rocillo
Paulo Rená da Silva Santarém

PROJETO GRÁFICO, CAPA, DIAGRAMAÇÃO, EDIÇÃO DE IMAGEM E FINALIZAÇÃO

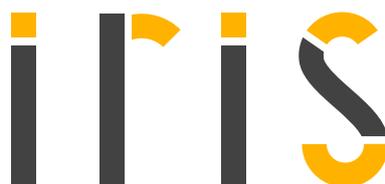
Felipe Duarte
Imagens: Freepik

PRODUÇÃO EDITORIAL

IRIS - Instituto de Referência em Internet e Sociedade

COMO REFERENCIAR EM ABNT

SILVA, Fernanda dos Santos Rodrigues; GERTRUDES, Júlia Maria Caldeira; DUTRA, Luiza Correa de Magalhães; SILVA, Rafaela Ferreira Gonçalves da. **Reclamações sobre o procedimento de moderação de conteúdo em redes sociais: o que pensam os usuários**. Belo Horizonte: Instituto de Referência em Internet e Sociedade, 2024. Disponível em: <https://bit.ly/4enMVKy>. Acesso em: dd mmm aaaa.



**INSTITUTO
DE REFERÊNCIA
EM INTERNET
E SOCIEDADE**

DIREÇÃO

Ana Bárbara Gomes

Paloma Rocillo

MEMBROS

Felipe Duarte | Coordenador de Comunicação

Fernanda Rodrigues | Coordenadora de Pesquisa e Pesquisadora

Glenda Dantas | Pesquisadora

Júlia Caldeira | Pesquisadora

Júlia Tereza Koole | Estagiária de pesquisa

Luisa Melo | Estagiária de pesquisa

Luiza Correa de Magalhães Dutra | Pesquisadora

Paulo Rená da Silva Santarém | Pesquisador

Thais Moreira | Analista de comunicação

Wilson Guilherme | Pesquisadore

irisbh.com.br

Sumário

Sumário Executivo	<u>5</u>
Apresentação	<u>7</u>
Introdução	<u>8</u>
1.1. Como identificar as principais reclamações de usuários sobre o procedimento de moderação de conteúdo em plataformas de rede social?	<u>9</u>
1.1.1. Como os dados foram coletados?	<u>11</u>
1.1.2. Como analisar os dados?	<u>12</u>
1.2. Quais as principais reclamações sobre o procedimento de moderação de conteúdo em relação à remoção de postagens e suspensão/bloqueio de contas?	<u>16</u>
1.3. Quais as demais reclamações sobre moderação de conteúdo em plataformas de redes sociais não relacionadas a remoção de postagens e/ou suspensão/bloqueio de contas?	<u>20</u>

Sumário Executivo

- O presente estudo buscou compreender quais as principais reclamações de usuários frente à moderação de conteúdo em redes sociais, com base na análise de queixas em relação ao Facebook, Instagram, Twitter, TikTok e Youtube feitas na plataforma Reclame Aqui;
- Do total de 1081 reclamações coletadas, foram descartadas reclamações que não versavam sobre moderação (28,03%), versavam sobre hackeamento (23,03%) e aquelas em que não foi possível identificar se era caso de moderação (7,4%), resultado em uma amostra final de 449 reclamações efetivamente analisadas.
- Dentre as reclamações analisadas, 54,34% tratavam sobre o objeto da pesquisa, qual seja, sobre **reclamações em torno do procedimento de moderação no caso remoção de postagens e suspensão/bloqueio de conta**, apontando para o quanto a opacidade no processo de moderação contribui para a desconfiança dos usuários, impactando negativamente a percepção pública das plataformas.
 - Considerando somente esse universo, as reclamações mais frequentes são, em ordem decrescente:
 - fundamentação inadequada de decisões de moderação (52,46%);
 - contestação de decisão de moderação não respondida (22,54%);
 - falta de notificação ou aviso sobre decisão de moderação (9,02%);
 - ausência de ferramentas para contestar a decisão de moderação (7,38%);
 - design da plataforma inacessível em relação aos mecanismos de revisão de decisão de moderação (4,51%);
 - outros (4,1%).
 - Dentre as reclamações analisadas, os outros 45,66% tratavam sobre moderação de conteúdo, inclusive sobre procedimento, mas não sobre nosso objeto. Neste universo, as reclamações versavam sobre:
 - reclamações genéricas sobre a moderação de conteúdo das plataformas (36,1%);
 - pedidos de moderação de conteúdo de terceiro (30,24%);
 - problemas com monetização (10,24%);
 - problemas com a recomendação ou o alcance do conteúdo (10,24%);
 - restrição de funcionalidades na plataforma em virtude de decisão de moderação (10,24%);
 - problemas gerados por limitação etária da plataforma (1,95%);
 - pedidos de moderação de anúncios (1,46%).
- Diante disso, **recomenda-se** que as plataformas digitais promovam transparência

e fundamentação nas decisões de moderação de forma adequada. É importante o fornecimento de explicações nítidas sobre as razões por trás das ações tomadas, como a remoção de conteúdo ou a suspensão de contas. A transparência ajuda a assegurar que as decisões não pareçam arbitrárias ou injustas, permitindo que os usuários compreendam melhor as regras e como elas são aplicadas. Além disso, a fundamentação nas decisões de moderação promove um ambiente mais equitativo e pode reduzir a sensação de censura ou injustiça por parte dos usuários;

- **Recomenda-se também que** as plataformas aumentem a eficácia dos processos de recurso e revisem suas políticas internas para abordar efetivamente as necessidades dos usuários e garantir um ambiente digital seguro e confiável. Os usuários devem ter acesso facilitado a mecanismos eficazes para contestar decisões de moderação que considerem injustas e esses processos devem ser simples e eficientes. Isso inclui garantir que os apelos sejam revisados de maneira justa e imparcial, com feedback sobre os resultados, sem deixar solicitações não respondidas e utilizar a automatização de respostas com parcimônia.



Apresentação

O IRIS é um centro de pesquisa independente e interdisciplinar dedicado a produzir e comunicar conhecimento científico sobre os temas de internet e sociedade com o objetivo de fomentar políticas públicas que avancem os direitos humanos na área digital. Há mais de 5 anos o IRIS se dedica em específico à pauta da moderação de conteúdo, com pesquisas científicas publicadas sobre transparência,¹ devido processo² e estratégias de combate à desinformação.³

O presente documento registra alguns dos primeiros resultados encontrados no projeto de pesquisa **Entre posts e polêmicas: estratégias de resolução de conflitos na moderação de conteúdo em plataformas digitais**, realizado com apoio financeiro da Google e conduzido de forma independente, com o objetivo de compreender os problemas relacionados à remoção de conteúdos legítimos que circulam em plataformas digitais, especialmente de redes sociais. Este relatório focou em identificar quais as **principais reclamações de usuários frente ao procedimento de moderação**, entendendo esta informação como fundamental para o aprimoramento das atividades das plataformas e maior segurança aos usuários.

1 KURTZ, Lahis Pasquali; DO CARMO, Paloma Rocillo Rolim; VIEIRA, Victor Barbieri Rodrigues. **Transparência na moderação de conteúdo: tendências regulatórias nacionais**. Belo Horizonte: Instituto de Referência em Internet e Sociedade, 2021. Disponível em: <https://bit.ly/3xjAUka>. Acesso em: 6 set. 2024.

2 SILVA, Fernanda dos Santos Rodrigues; GERTRUDES, Júlia Maria Caldeira; SILVA, Rafaela Ferreira Gonçalves da. **Regulação de plataformas e devido processo na moderação de conteúdo: perspectivas em 5 continentes**. Belo Horizonte: Instituto de Referência em Internet e Sociedade, 2024. Disponível em: <https://irisbh.com.br/publicacoes/devido-processo-na-moderacao-de-conteudo-em-5-continentes/>. Acesso em: 6 set. 2024.

3 PEREIRA, Ana Bárbara Gomes. SILVA, Fernanda dos Santos Rodrigues; GERTRUDES, Júlia Maria Caldeira; SILVA, Rafaela Ferreira Gonçalves da. **Cartilha de Enfrentamento à Desinformação em Redes Sociais**. Belo Horizonte: Instituto de Referência em Internet e Sociedade, 2024. Disponível em: <https://irisbh.com.br/wp-content/uploads/2024/08/Cartilha-de-Enfrentamento-a-Desinformacao-em-Redes-Sociais.pdf>. Acesso em: 6 set. 2024.

Busca-se fornecer um entendimento baseado em evidências sobre como as polêmicas relacionadas à moderação de conteúdo podem ser geridas de maneira mais eficaz e transparente no contexto digital brasileiro. Esta pesquisa é um passo necessário para ampliação da liberdade de expressão online, segurança dos usuários e confiabilidade às plataformas digitais.

Introdução

A falta de informações sobre o histórico de decisões de moderação de conteúdo em plataformas digitais, principalmente redes sociais, impede uma visualização adequada acerca dos principais problemas enfrentados pelos usuários durante esse procedimento. Apesar da existência de publicação de relatórios de transparência, a maioria se destina apenas a contabilizar números totais, como de conteúdos moderados, recursos e cumprimento de decisões judiciais. As situações concretas e racionalidade decisória que orienta a atuação da plataforma não são publicizadas.

Essa opacidade, no entanto, não parece contribuir para nenhuma das partes envolvidas. De um lado, os usuários não têm segurança sobre como se dará o procedimento de moderação e como a plataforma atua em casos específicos; de outro, as plataformas sofrem com especulações em torno de como operam, o que pode levar à desconfiança e descredibilidade frente ao público. Embora abrir completamente todos os procedimentos e mecanismos de moderação de conteúdo possa servir também de forma negativa, para que pessoas mal-intencionadas utilizem dessas informações para tentar burlar o sistema, o mínimo de transparência e observância de um devido processo parecem ser um caminho que equilibra os interesses de ambos os lados.⁴

Conforme apresentado em pesquisa publicada pelo IRIS em 2024, diferentes regulações voltadas para plataformas ao redor do mundo têm trazido parâmetros mínimos que se encaixam na ideia de um devido processo.⁵ Dever de notificar ação de moderação, prazo para recurso e resposta pela empresa, dentre outros, são algumas das possibilidades que normas internacionais têm previsto para dar segurança jurídica ao cenário. No entanto, uma informação fundamental para auxiliar no aprimoramento dessa atividade pelas plataformas e na previsão de políticas mais objetivas, permanece oculta: quais as **principais reclamações de usuários frente à moderação de conteúdo em redes sociais? Nosso estudo aponta evidências cientificamente embasadas sobre essa pergunta.**

4 SILVA, Fernanda dos Santos Rodrigues; GERTRUDES, Júlia Maria Caldeira; DUTRA, Luiza Correa de Magalhães; SILVA, Rafaela Ferreira Gonçalves da. **Guia Informativo: Devido Processo na regulação da moderação de conteúdo ao redor do mundo.** Belo Horizonte: Instituto de Referência em Internet e Sociedade, 2023. Disponível em: <https://bit.ly/3smC0i0>. Acesso em: 6 set. 2024.

5 SILVA, Fernanda dos Santos Rodrigues; GERTRUDES, Júlia Maria Caldeira; SILVA, Rafaela Ferreira Gonçalves da. **Regulação de plataformas e devido processo na moderação de conteúdo: perspectivas em 5 continentes.** Belo Horizonte: Instituto de Referência em Internet e Sociedade, 2024. Disponível em: <https://irisbh.com.br/publicacoes/devido-processo-na-moderacao-deconteudo-em-5-continentes/>. Acesso em: 30 ago 2024.

1.1. Como identificar as principais reclamações de usuários sobre o procedimento de moderação de conteúdo em plataformas de rede social?

Como já comentado, não existe hoje um repositório oficial de decisões e/ou reclamações sobre o procedimento de moderação de conteúdo online. Os números absolutos apresentados em relatórios de transparência não permitem uma visualização mais detalhada dos principais problemas enfrentados pelos usuários, o que poderia contribuir para uma melhora na realização da atividade. Pensando nisso, foi necessário buscar um canal de acesso público que as pessoas tivessem adotado como ferramenta principal para expressar suas críticas. No Brasil, frequentemente esse local é a plataforma **Reclame Aqui**, que, como o próprio nome já diz, serve para que consumidores possam fazer reclamações acerca de bens vendidos ou serviços prestados por determinadas empresas. O seu papel é tão relevante na área consumerista, que, juridicamente, tribunais pelo país têm reconhecido sua validade para fins de prova em ações envolvendo direito do consumidor.⁶

Na área acadêmica, estudos abordam diferentes questões relativas ao uso desse recurso. No que diz respeito ao consumidor, há pesquisas que exploram como a plataforma pode promover seu empoderamento, ainda que com restrições.⁷ Quanto às empresas, pesquisas analisaram seu *ethos* discursivo nas respostas às reclamações⁸ e o papel da plataforma no gerenciamento reputacional e de imagem.⁹ E por fim, quanto ao próprio Reclame Aqui, é possível localizar trabalhos que analisam a evolução da natureza da plataforma.¹⁰ De uma forma ou de outra, fato é que este espaço tem sido reconhecido como importante para compreender a apresentação de queixas de cidadãos e cidadãs frente a empresas de bens e serviços.

6 BRASIL. Tribunal de Justiça de Minas Gerais (10ª Câmara Cível). Agravo de Instrumento nº 1.0000.22.069254-5/001. DIREITO CIVIL E PROCESSUAL CIVIL. AGRAVO DE INSTRUMENTO. AÇÃO DE OBRIGAÇÃO DE FAZER COMPEDIDO DE TUTELA PROVISÓRIA DE URGÊNCIA. LIMINAR INDEFERIDA. RETIRADA DO NOME DA EMPRESA AUTORA DE RECLAMAÇÃO POSTADA NO SITE “RECLAME AQUI”. REQUISITOS LEGAIS. NÃO CONFIGURAÇÃO. PONDERAÇÃO DE PRINCÍPIOS. PREPONDERÂNCIA DA LIBERDADE DE EXPRESSÃO E DE PENSAMENTO. Relator(a): Des.(a) Jaqueline Calábria Albuquerque, 25 de outubro de 2022. Disponível em: https://www5.tjmg.jus.br/jurisprudencia/pesquisaNumeroCNJEspelhoAcordao.do;jsessionid=B7BB6EDBF0ACA98E483A990DFA60BA76.juri_node2?numeroRegistro=1&totalLinhas=1&linhasPorPagina=10&numeroUnico=1.0000.22.069254-5%2F001&pesquisaNumeroCNJ=Pesquisar. Acesso em: 2 set. 2024.

7 KOZINETS, Robert V.; FERREIRA, Daniela Abrantes; CHIMENTI, Paula. How Do Platforms Empower Consumers? Insights from the Affordances and Constraints of Reclame Aqui. **Journal of Consumer Research**, v. 48, n. 3, out./2021, pp. 428-455. Disponível em: <https://academic.oup.com/jcr/article-abstract/48/3/428/6171149>. Acesso em: 28 ago. 2024.

8 SÁ, Mirlene Batista. **O ethos discursivo em respostas de empresas no site Reclame Aqui**. Dissertação (Mestrado em Letras) - Universidade Federal de Rondônia, Porto Velho, 2021. 97p.

9 SEIVA, Carlos Eduardo Marques. **A apropriação da plataforma Reclame Aqui como balizadora na gestão de imagem e reputação organizacional na era da informação**: um estudo exploratório. Monografia (Curso de Administração) - Universidade do Vale do Rio dos Sinos, São Leopoldo, 2023. 75p.

10 SELL, Cleiton Lixieski; CORRÊA, Jeano Saraiva. “Reclame Aqui” no Brasil: uma análise dos limites e possibilidades frente a perspectiva de um movimento social e seu viés em relação ao ativismo digital. **Revista Faculdade de Direito da UFC**, Fortaleza, v. 39, n. 2, pp. 51-74, jul./dez. 2018. Disponível em: <https://core.ac.uk/download/pdf/480546611.pdf>. Acesso em: 28 ago. 2024.

Com a escolha pelo Reclame Aqui como o portal para verificar reclamações de usuários frente ao procedimento de moderação de conteúdo, surgiu um novo desafio: as páginas de plataformas digitais de redes sociais não são verificadas, ou seja, não possuem o selo de verificação do Reclame Aqui que, segundo o próprio site,¹¹ observa as seguintes informações, dentre outras:

- “Ter a empresa cadastrada gratuitamente na plataforma;
- Manter uma boa reputação no Reclame AQUI (Sem Reputação Definida, Regular, Bom, Ótimo e RA1000);
- Ter os canais de atendimento ativos e o CNPJ no Reclame AQUI;
- Ter contratado a RA Brand Page, a nossa solução de posicionamento de marca dentro da plataforma;
- Ser aprovado pela verificação mensal para conquistar o Selo de Empresa Verificada.”

Mesmo diante da ausência do selo de verificação, foi possível perceber que **as páginas das redes sociais a serem analisadas eram efetivamente utilizadas por pessoas que acreditavam conversar diretamente com a empresa**. Embora nenhuma das plataformas escolhidas tivesse sua página verificada no Reclame Aqui, a descrição da página, foto de perfil e/ou o teor e quantidade de reclamações indicava que se tratava de um espaço relacionado à respectiva plataforma,¹² o que orientava os usuários a prestarem suas queixas ali mesmo. Assim, por uma questão fática, **foi possível concluir que as reclamações encontradas diziam respeito a situações envolvendo as referidas redes sociais, ainda mais pela perspectiva dos usuários**.

Diante disso, a metodologia utilizada para chegar aos resultados dessa pesquisa passou pelas seguintes etapas:

11 RECLAME AQUI. Selo RA Verificada do Reclame AQUI verifica a existência e confiança das empresas. Disponível em: <https://blog.reclameaqui.com.br/selo-ra-verificada-certifica-credibilidade-da-empresa/>. Acesso em: 28 ago. 2024.

12 A página das plataformas no Reclame Aqui podem ser acessadas nos seguintes links: Instagram - <https://www.reclameaqui.com.br/empresa/instagram/>; Facebook - <https://www.reclameaqui.com.br/empresa/facebook/>; Twitter - <https://www.reclameaqui.com.br/empresa/twitter/>; TikTok - <https://www.reclameaqui.com.br/empresa/tiktok/>; e Youtube - https://www.reclameaqui.com.br/empresa/youtube_194558/sobre/.

1.1.1. Como os dados foram coletados?

- Plataformas de rede social investigadas: Instagram, Facebook, Twitter, Tiktok e Youtube;
- Justificativa da seleção das plataformas: são as mais utilizadas por candidatos a cargos políticos,¹³ demonstrando seu possível impacto em um dos principais processos do Estado democrático de direito;
- Amostra de coleta prevista: 250 reclamações de cada plataforma;
- Forma de coleta:
 - A plataforma do Reclame Aqui apresentava algumas limitações, como o fato de não permitir selecionar um lapso temporal para pesquisar reclamações e não ser possível acessar além da página 50 de queixas. Em virtude disso, e considerando o grande fluxo de entrada de novas reclamações a cada hora, buscou-se coletar as reclamações das primeiras 25 páginas de cada plataforma, contendo 10 reclamações cada uma, o que totalizava 250 reclamações por rede social;
 - Foram selecionadas reclamações apresentadas na categoria “Conteúdo”, que era gerada pelo próprio Reclame Aqui;
 - A coleta foi feita por duas pesquisadoras a partir do registro em print da página da reclamação, link da reclamação e conteúdo;
 - Após a coleta das reclamações selecionadas, observamos que algumas estavam repetidas, de modo que as repetições foram excluídas.
- Amostra total analisada: 1081 reclamações
- Amostra final por plataforma, já excluídas as repetições:
 - a) Instagram: 247; b) YouTube: 249; c) TikTok: 85; d) Facebook: 250; e e) Twitter: 250
- Limitações:
 - Por ser a página criada mais recentemente na plataforma¹⁴ e, conseqüentemente,

13 MALI, Tiago. Sobe número de candidatos nas redes sociais; saiba as preferidas. **Poder 360**. 18 ago. 2022. Disponível em: <https://www.poder360.com.br/eleicoes/sobe-numero-de-candidatos-nas-redes-sociais-saiba-as-preferidas/>. Acesso em: 6 set. 2024.

14 A página do TikTok no Reclame Aqui consta como cadastrada há 4 anos, enquanto as demais têm entre 7 a 13 anos.

com menor número de reclamações na categoria “Conteúdo”, o TikTok teve um número inferior de queixas analisadas;

- É possível que reclamações sobre o objeto da pesquisa, isto é, sobre o procedimento de moderação de conteúdo nas plataformas escolhidas, estivessem também classificadas em outras categorias que não a de “Conteúdo”, de modo que não foram coletadas nesta pesquisa. No entanto, esta foi uma limitação que se impôs, diante da impossibilidade de revisar todo o material do site. Assim, nos testes realizados, o fato da categoria “Conteúdo” retornar temas principais relacionados ao nosso objetivo demonstrava que ela era um recorte pertinente e apropriado para ser trabalhado.

1.1.2. Como analisar os dados?

Para análise dos dados, foram criadas categorias iniciais para classificar as reclamações, a partir de estudo anterior¹⁵ sobre os elementos de um direito ao devido processo na moderação de conteúdo. Ou seja, parâmetros mínimos para assegurar justiça e legitimidade no procedimento de moderação das plataformas frente aos usuários foram a lente de análise das reclamações. Considerando que a moderação de conteúdo pode ser tanto intervenção no conteúdo de terceiros, como através de remoção de postagens e suspensão de contas, assim como de recomendação de conteúdos no feed, optou-se por restringir o objeto da pesquisa a analisar somente o primeiro caso, ou seja, **reclamações sobre procedimentos de remoção de postagens e suspensão/bloqueio de contas.**

Dessa forma, a primeira versão da lista de categorias foi testada por três pesquisadoras da equipe em duas rodadas, sendo feitas reuniões a cada uma delas para discutir a classificação e refletir sobre a necessidade de alteração da lista inicial. Ao final dessa etapa, a segunda versão da lista de categorias foi a seguinte: *ausência de ferramentas de recurso/contestação/revisão; design inacessível da plataforma; contestação de decisão de moderação não respondida; usuário não foi informado sobre decisão de moderação, mas percebeu conteúdo removido; plataforma informou sobre decisão de moderação de conteúdo, mas não fundamentou; não é sobre moderação; é sobre moderação, mas não sobre nosso objeto; não foi possível identificar se é moderação; e outros.*

De forma resumida, o procedimento de análise das reclamações à luz do devido processo ocorreu da seguinte forma:¹⁶

15 SILVA, Fernanda dos Santos Rodrigues; GERTRUDES, Júlia Maria Caldeira; SILVA, Rafaela Ferreira Gonçalves da. **Regulação de plataformas e devido processo na moderação de conteúdo: perspectivas em 5 continentes.** Belo Horizonte: Instituto de Referência em Internet e Sociedade, 2024. Disponível em: <https://irisbh.com.br/publicacoes/devido-processo-na-moderacao-deconteudo-em-5-continentes/>. Acesso em: 30 ago 2024.

16 As três pesquisadoras da equipe se dividiram na análise das plataformas, o que foi feito em três rodadas: (I) uma rodada de análise inicial de cada plataforma por sua respectiva pesquisadora; (II) uma segunda rodada em que cada pesquisadora revisava a classificação das reclamações em duas ou

1. Divisão das reclamações para cada pesquisadora;
2. Análise inicial conforme divisão;
3. Revisão da análise feita por outra pesquisadora;
4. Análise das divergências em uma terceira rodada.

Após análise de todas as reclamações coletadas, identificou-se que o grande número envolvendo contas hackeadas demandava uma reclassificação enquanto uma categoria por si só, o que ensejou uma revisão de categorização de todas as plataformas por duas pesquisadoras. Nesse sentido, considerando a amostra total de 1081 reclamações e o objeto da pesquisa, também foram descartadas reclamações que não versavam sobre moderação (28,03%), que versavam sobre hackeamento (23,03%) e aquelas em que não foi possível identificar se era caso de moderação (7,4%). O resultado é uma amostra final de 449 reclamações sobre moderação, que foram efetivamente analisadas.

Destas, 54,34% eram sobre o nosso objeto de pesquisa, repita-se, o procedimento de moderação nos casos de remoção de postagens e suspensão/bloqueio de contas, enquanto 45,66% eram sobre outros aspectos da moderação. Assim, as reclamações analisadas foram classificadas da seguinte forma:

uma plataforma em que não havia trabalhado antes; e (III) uma terceira rodada de revisão em que cada pesquisadora pegava uma nova plataforma que não havia trabalhado e analisava somente os casos de divergência entre as duas primeiras para tomada de decisão final. Após, todas apresentaram coletivamente seus pareceres da terceira rodada, a fim de sanar quaisquer desencontros e verificar eventual necessidade de ajuste.

QUAIS OS PRINCIPAIS PROBLEMAS NA MODERAÇÃO DE CONTEÚDO NAS REDES SOCIAIS?

Design inacessível da plataforma

Casos em que não foi possível localizar como contestar/revisar/recorrer da decisão

1

Ausência de ferramentas de recurso/contestação/revisão

Casos em que a plataforma não disponibilizou local para recorrer / revisar / contestar decisão

2

Contestação de moderação não respondida

Casos em que o recurso foi apresentado, uma ou mais vezes, e plataforma nunca retornou, ou em que existia uma demora injustificada para retornar o pedido de revisão

3

Usuário não foi informado sobre decisão de moderação, mas percebeu conteúdo removido

Casos em que o usuário percebeu que teve conteúdo derrubado por não o ter encontrado mais, mas a plataforma não o avisou em nenhum momento.

4

Plataforma informou sobre a moderação de conteúdo, mas não fundamentou adequadamente

Casos em que usuário reclamou que não teve explicação do porquê seu conteúdo foi removido.

5

Outros

Situações que questionaram a forma de resolução de conflitos, mas não se enquadram em nenhuma das categorias listadas

6

É sobre moderação, mas não sobre nosso objeto

Casos de moderação de terceiros por conteúdo danoso; de monetização; de moderação por terceiros por falta de identidade; de recomendação de conteúdo; de reclamações sobre moderação/decisão, mas não sobre procedimento

7

Em relação aos 45,66% da categoria *é sobre moderação, mas não sobre nosso objeto*, optou-se por classificar também em um segundo momento, a fim de compreender quais os outros problemas envolvendo moderação de conteúdo que geraram reclamações. Nesse sentido, a subcategorização ficou assim definida:

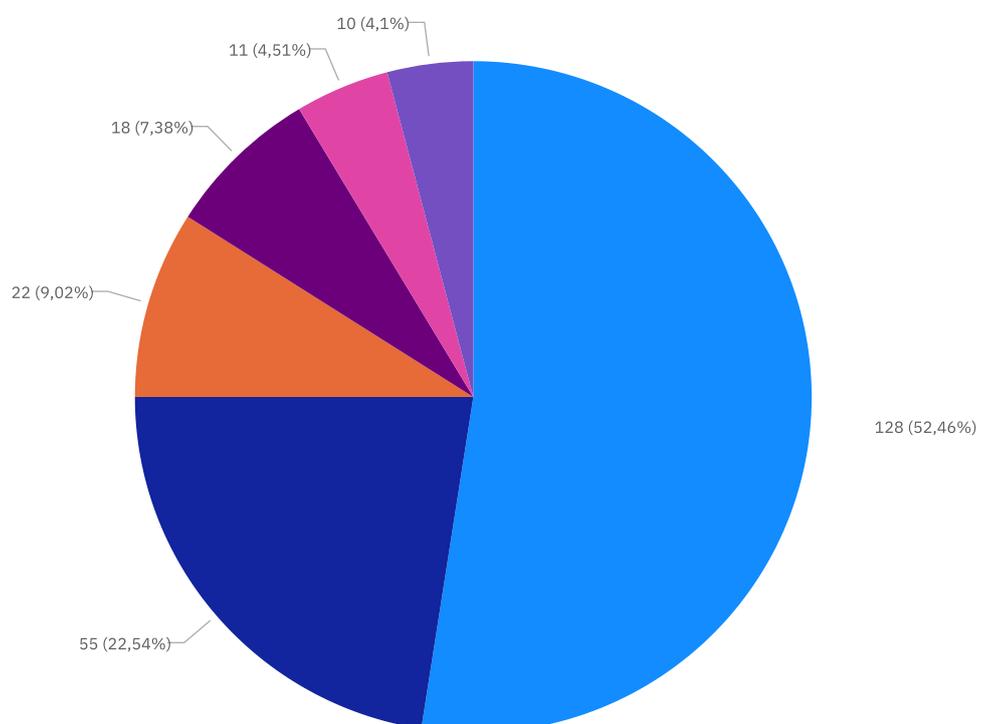
- **Suspensão por conta vinculada:** casos específicos de contas que foram moderadas por bloqueio ou suspensão por suposta conexão com uma outra conta em outra plataforma;
- **Reclamação genérica sobre a moderação da plataforma:** casos de a) reclamações genéricas sobre como a moderação funciona (ou não) na plataforma, b) meras reclamações por ter sido moderado, c) reclamação sobre recomendação de conteúdo na plataforma (considerando recomendação uma espécie do gênero moderação) e d) reclamação sobre moderação feita em perfis de terceiros;
- **Pedido de moderação de conteúdo de terceiros por conteúdo danoso:** pedidos para moderação de perfis ou de postagens específicas em razão de conteúdo danoso;
- **Pedido de moderação de conteúdo de terceiros por suposta falsa identidade:** casos de perfis acusados de se passarem por outras pessoas;
- **Pedido de moderação de conteúdo de terceiros por uso indevido de imagem:** casos de uso indevido de imagem do reclamante ou de terceiro, sem sua autorização;
- **Recomendação/alcance de conteúdo:** casos de insatisfação com a recomendação/alcance de conteúdo produzido pela própria pessoa que reclama;
- **Restrição de funcionalidades na plataforma:** casos em que a pessoa que reclama não pode mais fazer transmissão ao vivo, postar stories, fazer comentários ou algum outro tipo de restrição de funcionalidade associado a uma decisão de moderação da plataforma;
- **Monetização:** casos envolvendo monetização de perfis/contas;
- **Conta bloqueada por limitação etária da plataforma:** casos em que a conta sofreu moderação em razão da idade indicada no perfil;
- **Pedido de moderação de anúncios:** casos sobre moderação de anúncios, que se diferencia da mera recomendação de conteúdo, na medida em que anúncios são necessariamente impulsionados/pagos;

- **Outros:** engloba reclamações residuais, cujo assunto não se encaixa em nenhuma subcategoria específica.

A criação das categorias partiu de um esforço científico para buscar entender melhor os principais tipos de reclamações dos usuários sobre a moderação de conteúdo nas redes sociais. Em alguns casos, foi difícil separar claramente as categorias, já que as queixas muitas vezes envolviam mais de um problema. Nessas situações, a equipe focou em identificar o principal ponto da reclamação para classificá-la corretamente.

A fim de evitar qualquer tipo de ranqueamento entre as plataformas analisadas, optamos por apresentar os resultados de forma conjunta, sem apontar a porcentagem de reclamações por cada plataforma. Na seção 2, você acompanha os resultados da amostra de reclamações relativas ao procedimento de moderação de conteúdo, no que tange à remoção de postagens e suspensão/bloqueio de contas. Na seção 3, você acompanha a análise da categoria *é sobre moderação, mas não sobre nosso objeto*.

1.2. Quais as principais reclamações sobre o procedimento de moderação de conteúdo em relação à remoção de postagens e suspensão/bloqueio de contas?



- Plataforma informou sobre decisão de moderação de conteúdo, mas não fundamentou
- Contestação de decisão de moderação não respondida
- Usuário não foi informado sobre decisão de moderação, mas percebeu conteúdo removido
- Ausência de ferramentas de recurso/contestação/revisão
- Design inacessível da plataforma
- Outros

Considerando as categorias que envolvem nosso objeto de pesquisa, isto é, **reclamações sobre o procedimento de moderação de conteúdo em relação à remoção de postagens e suspensão/bloqueio de contas**, foi possível perceber que a maioria das reclamações versava sobre a *falta de fundamentação adequada das plataformas (52,46%)* em relação à moderação de conteúdo feita. Nesse caso, encaixam-se casos em que os usuários apontaram não terem compreendido as razões para a decisão de moderação, como se vê no exemplo abaixo:

No dia [data ocultada], tive meu [nome da plataforma ocultado] desativado e recebi a seguinte mensagem encontramos violações graves ou recorrentes das nossas diretrizes da comunidade.

No entanto, acredito que todos os vídeos publicados no canal e todas as minhas ações dentro da plataforma vão de acordo com os Termos de Uso e Diretrizes da comunidade do [nome da plataforma ocultado].

*Sendo assim, **gostaria que a plataforma explicasse quais seriam essas violações e quando elas foram cometidas** e, se possível, peço que reveja a decisão e reative o canal.*

Ou seja, existe uma série de reclamações de usuários em relação à transparência e nitidez da forma com que decisões sobre moderação de conteúdo são fundamentadas - quais conteúdos são permitidos ou não, quais as normas para os procedimentos das moderações realizadas e quais termos das políticas de comunidade efetivamente haviam sido violados. Nesta mesma categoria de fundamentação inadequada, foi possível visualizar também casos em que as pessoas reclamavam de procedimentos e respostas automáticas que não endereçavam corretamente os argumentos apresentados no recurso enviado. O trecho abaixo ilustra melhor a situação:

*Após a suspensão, busquei ativamente reativar minha conta, **mas tenho recebido apenas respostas automáticas e genéricas por parte do [nome da plataforma ocultado], que não oferecem nenhum esclarecimento concreto sobre como resolver a situação.** Esta falta de comunicação eficaz tem me deixado completamente **desamparada e sem meios adequados para contestar a decisão ou resolver o problema de forma satisfatória.***

*Desta forma, solicito urgentemente uma **revisão imparcial e transparente da situação**, com a reinstalação imediata da minha conta no [nome da plataforma ocultado] [inserir nome de usuário] ou, caso necessário, uma **explicação detalhada das razões para a suspensão e um caminho claro para resolver qualquer possível irregularidade.***

Espero uma resposta rápida e adequada para que eu possa retomar minhas atividades normais sem mais prejuízos desnecessários.

É conhecido o amplo processo de automatização do atendimento prestado por plataformas digitais, que as plataformas justificam frente à grande quantidade de conteúdo postado por minuto.¹⁷ No entanto, reclamações como essa demonstram que a falta de revisão humana pode levar a um possível cerceamento do próprio direito de recorrer do usuário. Na medida em que apresenta um recurso, expondo os argumentos pelos quais acredita que não violou as políticas de comunidade da plataforma, o usuário espera que suas razões sejam ponderadas a fim de reverter a decisão atacada. Um mero atendimento automatizado que não é capaz de compreender esses argumentos e dar uma resposta adequadamente fundamentada pode acabar gerando frustração e tornando quase inócua a possibilidade de apelação. Desta forma, um elemento central ao devido processo, o direito de apelação, fica prejudicado.

Apesar de não ser possível ignorar o fato de que pode haver uma falta de compreensão do próprio usuário em relação ao que está previsto nas políticas de comunidade da plataforma, o que impacta na sua compreensão em torno da fundamentação das decisões de moderação, isto é um problema no qual a plataforma também tem responsabilidade. Se a compreensão das regras têm sido difícil, é papel da plataforma adotar medidas capazes de simplificá-las e deixá-las mais objetivas, a fim de que as pessoas possam se apropriar delas de forma mais eficiente.

O processo de moderação de conteúdo online não envolve, neste sentido, apenas o binarismo de retirar ou não retirar determinado conteúdo online, mas todo o processo de construção de relação usuário-plataforma e das regras e normas estabelecidas para gerar uma maior segurança jurídica, assim como uma construção político-social das ações de moderação - que envolvem subjetividades, discursos e liberdades. Em outras palavras, a forma como se lida com essas questões afeta a comunicação e a interação social¹⁸.

A falta de elementos essenciais a um processo de moderação adequado e fundamentado, como transparência, notificação e consistência pode levar à percepção de censura. Essa preocupação é particularmente relevante quando se verifica que *22,54% das reclamações analisadas envolviam contestações de decisão de moderação que não foram respondidas*. Essa lacuna no processo de contestação pode alimentar a ideia de que as plataformas estão agindo de maneira arbitrária, o que acentua a necessidade de uma abordagem mais rigorosa e transparente na moderação de conteúdos.

Em um cenário onde muitas pessoas têm se utilizado de redes sociais como principal meio de comunicação para o seu trabalho, a falta de uma resposta sobre recursos de moderação pode deixá-las em completo desamparo e afetar os seus ganhos. O mesmo pode ser dito em casos em que *o usuário percebeu conteúdo removido, mas afirma não ter recebido nenhuma*

17 De acordo com pesquisa realizada pela Domo, uma empresa de serviço de computação em nuvem, foram compartilhadas 66 mil fotos no Instagram; 347,2 mil tweets no antigo Twitter – agora X; e 1,7 milhões de “pedaços de conteúdo” no Facebook por minuto em 2022. Veja mais em: DOMO. Data never sleeps 10.0. Disponível em: <https://web-assets.domo.com/miyagi/images/product/product-feature-22-data-never-sleeps-10.png>. Acesso em: 17 jun. 2024.

18 GOLDMAN, Eric. Content moderation remedies. Michigan Technology Law Review, v. 28, p. 1-xx, 2021. Disponível em: <https://repository.law.umich.edu/mtlr/vol28/iss1/2>. Acesso em: 30 ago. 2024.

notificação a respeito (9,02%). Na ausência de comunicações pela própria plataforma, fica até mesmo impossibilitada a chance de recorrer da decisão em lugar próprio, sendo necessário buscar outros mecanismos, como o próprio Reclame Aqui. O caso abaixo é exemplo de uma reclamação nesse sentido:

*(...) Recentemente, **descobri que diversos dos meus vídeos, que acumularam mais de 2 milhões de curtidas, foram apagados sem qualquer aviso prévio ou explicação adequada.** Esses vídeos representavam horas de trabalho criativo, dedicação e interação com meus seguidores, e sua **remoção abrupta** causou um impacto negativo em minha presença na plataforma.*

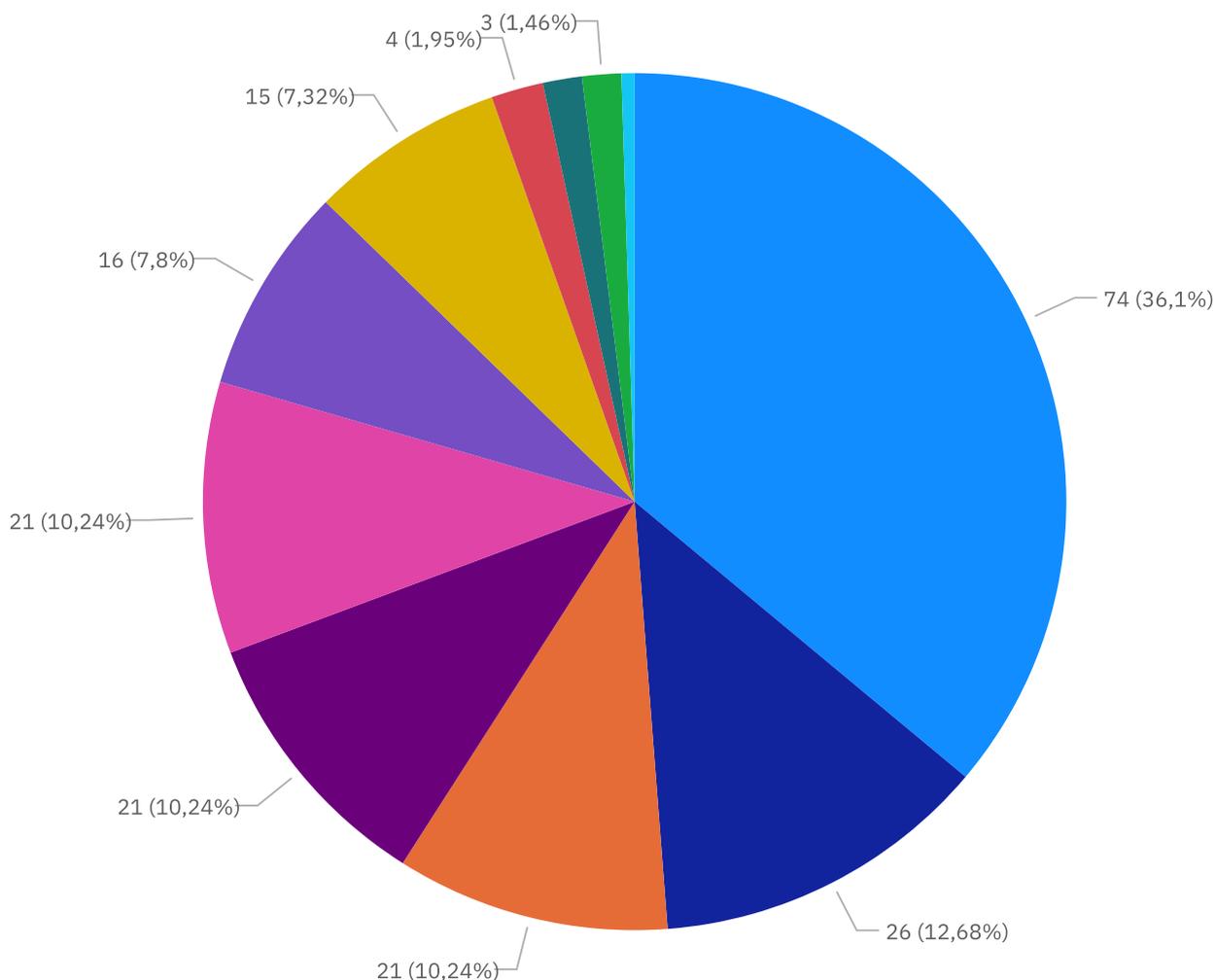
*Além disso, **é extremamente perturbador constatar que essas exclusões foram realizadas sem meu consentimento e, até o momento, não recebi nenhuma justificativa convincente por parte da equipe de suporte do [nome da plataforma ocultado]. A falta de transparência e comunicação eficaz por parte da plataforma é inaceitável e desrespeitosa** com os criadores de conteúdo que dedicam seu tempo e esforço para contribuir com a comunidade do [nome da plataforma ocultado].(...)*

Com efeito, a remoção de conteúdo já é uma medida sensível e que pode gerar frustração ao usuário, no entanto a sua realização sem qualquer tipo de informação prévia ou justificativa tende a agravar ainda mais a situação. Como já mencionado, em uma realidade onde muitas pessoas têm tirado toda ou parte de sua renda do meio digital, perceber que um conteúdo foi removido sem a devida notificação pela plataforma acaba por quebrar a relação de confiança que foi estabelecida entre as duas partes, justamente por configurar um quebra de expectativa em relação a um devido processo em que o usuário é adequadamente avisado de qualquer medida tomada.

Por fim, embora as porcentagens de reclamações nas categorias de *design inacessível da plataforma (4,51%)* e *ausência de ferramentas de recurso/contestação/revisão (7,38%)* sejam relativamente baixas, elas ainda são expressivas e indicam áreas significativas de preocupação. A revisão constante dos processos de moderação, bem como a otimização das interfaces e funcionalidades das plataformas são essenciais para garantir não apenas a manutenção da eficácia nestes aspectos, mas também para tornar as ações das plataformas cada vez mais transparentes e alinhadas com as necessidades dos usuários.

É de suma importância que as pessoas tenham acesso facilitado e intuitivo a mecanismos efetivos para solicitar revisões e manifestar discordâncias. Essa abordagem assegura a conformidade com os princípios de um devido processo na moderação, proporcionando aos usuários a possibilidade de exercer uma ampla defesa e contestar decisões de forma equitativa e transparente, sem enfrentar maiores dificuldades nesse processo.

1.3. Quais as demais reclamações sobre moderação de conteúdo em plataformas de redes sociais não relacionadas a remoção de postagens e/ou suspensão/bloqueio de contas?



- Reclamação genérica sobre a moderação da plataforma
- Pedido de moderação de conteúdo de terceiros por conteúdo danoso
- Monetização
- Recomendação/alcance de conteúdo
- Restrição de funcionalidades na plataforma
- Pedido de moderação de conteúdo de terceiros por suposta falsa identidade
- Pedido de moderação de conteúdo de terceiros por uso indevido de imagem
- Conta afetada por limitação etária da plataforma
- Outros
- Pedido de moderação de anúncios
- Suspensão por conta vinculada

Nesta seção, estão as reclamações identificadas dentro da categoria *é sobre moderação, mas não sobre nosso objeto*, que equivale a 18,96% do total da amostra coletada, em que foram enquadradas as demais reclamações envolvendo moderação de conteúdo, até mesmo sobre procedimento, **mas que não versavam sobre remoção de postagens e suspensão/bloqueio de contas**. Os casos mais recorrentes, como demonstra o gráfico, são de *reclamações genéricas sobre a moderação feita pela plataforma (36,1%)* e todas as subcategorias que, juntas, envolviam *pedidos de moderação de conteúdo de terceiros (30,24%)*.

Em relação ao primeiro, das reclamações mais gerais, foram identificadas queixas genéricas sobre como a moderação funciona (ou não) na plataforma, mero descontentamento com o fato de ter sido moderado, reclamação sobre conteúdos que eram recomendados pelo algoritmo na plataforma e reclamação sobre decisão de moderação aplicada a perfis de terceiros. Especificamente, as queixas dos usuários sobre a falta de clareza quanto à recomendação de certos conteúdos — particularmente quando eles não lembram de ter assistido material similar anteriormente —, evidenciam problemas com a opacidade dos algoritmos de recomendação. Embora a coleta e análise de dados de interação e informações de empresas parceiras sejam justificadas como uma forma de personalizar a experiência do usuário, essas queixas indicam uma falha significativa na promessa de uma personalização eficaz e transparente.

Da mesma forma, chamaram nossa atenção as reclamações sobre ações de moderação feitas sobre conteúdos de terceiros que não o reclamante. A esse respeito, identificou-se que alguns reclamantes prestaram queixas em nome de canais ou contas de influenciadores que eles acompanhavam e que haviam sido moderados, solicitando para que a plataforma revertesse a penalidade aplicada. Isto aponta para um uso do Reclame Aqui como uma espécie de espaço para ativismo em prol dos criadores de conteúdo.

Por outro lado, identificou-se que os reclamantes também utilizavam esse local para solicitar a moderação de conteúdo de terceiros. Neste caso, três situações foram mais identificadas:

12,68%

Pedido de moderação de terceiros por publicação de conteúdo danoso

Quando as pessoas alegavam encontrar conteúdo que violava as políticas de comunidade da plataforma

10,24%

Por acusação de suposta falsa identidade

Quando as pessoas denunciavam a criação de contas falsas que estariam se passando por elas; e

7,32%

Por uso indevido de imagem

Nos casos em que as pessoas apontavam que contas de terceiros estavam utilizando fotos suas, de sua família ou conhecidos, de forma não autorizada.

Este uso é curioso, na medida em que todas as plataformas analisadas dispõem de ferramentas próprias dentro de seus sites para a denúncia de conteúdos que se encaixam nos casos relatados. Em alguns casos, os próprios reclamantes apontavam já terem recorrido para a plataforma, mas nada ter sido feito a respeito. Embora seja possível que o conteúdo não tenha, de fato, violado as políticas da plataforma, a grande quantidade de pedidos de moderação de conteúdo de terceiros destaca algum tipo de insuficiência da rede social – seja de deixar seus mecanismos de denúncia mais acessíveis, seja de revisar suas políticas internas, a fim de verificar eventual lacuna que permite conteúdo que pode ser classificado como danoso, uso indevido de imagem e criação de contas falsas em nome de pessoas reais.

Em terceiro lugar, estão as reclamações tanto da categoria de *monetização* quanto de *recomendação/alcance de conteúdo*, com 10,24% cada uma. Na primeira, estão os casos relacionados à moderação de conteúdo que gerou limitação de acesso à monetização, reforçando o papel que as plataformas têm ocupado de constituir fonte de renda para uma grande quantidade de pessoas. No caso de recomendação/alcance de conteúdo, estas se diferenciam das reclamações genéricas, ao passo que dizem respeito às reclamações relacionadas à recomendação ou alcance de conteúdo produzido pelos próprios reclamantes. Em alguns relatos, os usuários mencionam expressamente a suspeita de terem sofrido com a prática de *shadowbanning*,¹⁹ que se traduz na hipótese em que a pessoa percebe uma redução no alcance de suas postagens, mas sem que a plataforma tenha notificado qualquer decisão a respeito:

Desde Outubro, conforme apontado por outros usuários, meu nome não aparece mais em nenhuma pesquisa ou hashtag. Conferindo em sites que verificam a presença desse tipo de censura, descobri que estou sim “Shadow banned” sem ter feito absolutamente nada anormal, vulgar ou de baixo calão.

Pensei então que isso fosse se resolver naturalmente com o tempo, mas até hoje ainda estou censurado nas pesquisas e o meu conteúdo (artes) simplesmente não alcançam usuários fora do meu círculo de seguidores, o que desfaz completamente o propósito da rede social.

19 Veja mais sobre o conceito em RADSCH, Courtney. Shadowban/Shadow Banning. In: BELLI, Luca; ZINGALES, Nicolo; CURZI, Yasmin (orgs.). **Glossary of platform: law and policy terms**. Rio de Janeiro: FGV Direito Rio, 2021.

*Tentei contato em inglês com @Support (o perfil de suporte do [nome da plataforma oculto]) e tudo que recebi foram sugestões de como agir corretamente em redes sociais, o que é redundante já que, até onde sei, não agi de forma a gerar represálias [sic] em nenhum momento. **Não souberam me informar o motivo da punição.***

*Solicito remoção de tal medida negativa contra minha conta, informação sobre como remove-la ou, pelo menos, **alguma satisfação sobre o porquê da conta estar sob embargo.***

A dificuldade de saber o que está acontecendo (se está acontecendo), assim como as razões para uma decisão como essa são alguns dos problemas relacionados. Há, ainda, o fato de impedir a ampla defesa e contraditório, uma vez que a pessoa sequer é informada sobre ter sido moderada.

Finalmente, de forma mais residual, estão os casos de moderação de contas por limitação etária da plataforma (1,95%), pedido de moderação de anúncios (1,46%), suspensão por conta vinculada (0,49%) e outros (1,46%). Eles demonstram problemas residuais no funcionamento das redes sociais que poderiam se beneficiar de formas mais efetivas de comunicação e recurso às plataformas digitais investigadas.



INSTITUTO
DE REFERÊNCIA
EM INTERNET
E SOCIEDADE