

**A REGULAÇÃO DO DISCURSO DE ÓDIO NAS REDES SOCIAIS: O CASO
DO OVERSIGHT BOARD DA META**

*THE REGULATION OF HATE SPEECH ON SOCIAL NETWORKS: THE CASE OF
THE META'S OVERSIGHT BOARD*

Laíse Milena Barbosa*

Resumo

O artigo tem como objetivo compreender e apresentar as tensões existentes para a regulação do discurso de ódio online através de uma análise das primeiras decisões sobre o tema do recém-criado *Oversight Board* da Meta. Questiona-se sobre os parâmetros utilizados pelo *Board* e se suas decisões respeitam os direitos humanos. Para tanto, desenvolveu-se uma pesquisa de cunho exploratório que promove uma análise qualitativa das decisões, por meio do método de análise de conteúdo. Amparando-se nas técnicas de pesquisa documental e bibliográfica, o estudo foi organizado em três itens. Primeiramente, foram expostas características do *Board* e algumas impressões da comunidade acadêmica acerca do seu formato. Em seguida, os padrões da comunidade da Meta sobre discurso de ódio e o percurso metodológico da seleção e organização dos casos foi apresentado. Por fim, foi realizada uma sistematização da pesquisa através de uma análise conjunta de todas as decisões do *Oversight Board* que envolveram as políticas ódio no período de análise. Como conclusão, observou-se que as decisões do *Board* partem de obstáculos já conhecidos sobre o tema, mas que se tornou um importante mecanismo de regulação, que fornece informações importantes e contribui para a transparência e fiscalização do processo moderação de conteúdo da Meta.

Palavras-Chave: Responsabilidade jurídica das plataformas digitais. Facebook. Instagram. Moderação de conteúdo.

Abstract

The article aims to understand and present the existing tensions for the regulation of online hate speech through an analysis of the first decisions about the newly created Oversight Board of Meta. Questions are asked about the parameters used by the Board and whether its decisions respect human rights. For that, exploratory research was developed that promotes a qualitative analysis of the decisions, through the method of content analysis. Based on documentary and bibliographical research techniques, the study was organized into three items. First, the characteristics of the Board and some impressions of the academic community about its format were exposed. Then, Meta's

* Mestre em Direito do Estado pela Universidade Federal do Paraná (UFPR), com bolsa CAPES/PROEX. Visiting Researcher na Universidade de Graz, Áustria, com bolsa Coimbra Group. Bacharela em Direito pela Universidade Estadual de Ponta Grossa (UEPG). Fundadora e Editora Científica da Revista Antinomias (ISSN 2675-9608). Research Fellow do Information Society Law Center da Università degli Studi di Milano.

community standards on hate speech and the methodological path for selecting and organizing the cases were presented. Finally, a systematization of the research was carried out through a joint analysis of all the decisions of the Oversight Board that involved hate politics in the period of analysis. In conclusion, it was observed that the Board's decisions start from obstacles already known on the subject, but that it has become an important regulatory mechanism, which provides important information and contributes to the transparency and supervision of the Meta content moderation process.

Keywords: Legal responsibility of digital platforms. Facebook. Instagram. Content moderation.

Sumário

Introdução. 1. Considerações sobre o *Oversight Board*. 2. As decisões do *Oversight Board* sobre discurso de ódio. Metodologia. 3. Insights comparativos acerca das decisões sobre discurso de ódio do *Oversight Board*. 3.1. Características dos casos. 3.2. Padrões relevantes. 3.3. Declaração do usuário. 3.4. Decisões da Meta. 3.5. Submissão de Comentários Públicos. 3.6. Análise do *Oversight Board*. 3.7. Decisões e Recomendações do *Oversight Board*. Conclusão. Referências bibliográficas.

INTRODUÇÃO

A utilização das redes sociais aumentou, consideravelmente, a capacidade dos indivíduos de exercerem a liberdade de expressão, a comunicação e a obtenção de informações. Contudo, a mesma liberdade pode resultar em discursos violentos, de estigmatização das vítimas e que perpetuam preconceitos existentes na sociedade atual, como o discurso de ódio.

Dante disso, as oportunidades oferecidas pelos serviços de redes sociais podem alcançar abusos inaceitáveis por parte de indivíduos isolados ou grupos organizados pelas redes. Nesse impasse em limitar ou não a liberdade de expressão, surge a complexidade em conceituar aquilo que se entende por *hate speech* e quando o conteúdo deve ser removido.

Como estrutura de governança¹ e não apenas empresa de mídia, tais dados demonstram que, mesmo na ausência de cobranças legais, a Meta adotou seus próprios parâmetros de regulação e restrição da liberdade de expressão na esfera virtual. Porém, a falta de transparência tem sido uma das críticas mais consistentes à moderação da Meta.

Dante da crescente discussão sobre as restrições à fala online por meio da moderação de conteúdo, o Facebook – desde 28 de outubro de 2021 e, neste estudo,

¹ KLONICK, K. The New Governors: the people, rules, and processes governing online speech. *Harvard Law Review*, v. 131, 2018.

denominado Meta – anunciou em 2018 a criação do Oversight Board, um projeto ambicioso, independente e transparente no qual os usuários poderiam recorrer de suas decisões.

A criação do *Board* foi um momento chave de maior conscientização e insatisfação do público². A forma como algumas plataformas de tecnologia decide, principalmente no que diz respeito ao discurso online, justifica a escolha do tema, pois pode trazer ramificações significativas em relação ao futuro desse discurso nas redes sociais e causar uma revolução na infraestrutura da liberdade de expressão.

Em cumprimento à proposta, em dezembro de 2020, o *Oversight Board* anunciou que começaria receber objeções de conteúdos removidos pelo Facebook ou Instagram. Em 28 de janeiro de 2021, foram publicadas suas primeiras cinco decisões, cobrindo os tópicos mais importantes e debatidos sobre discurso de ódio, nudez e organizações criminosas. Além disso, desde os primeiros relatórios de transparência desse comitê, uma das categorias com maior número de apelações recebidas, foi o *hate speech*³.

Tendo em vista o contexto relatado, a presente pesquisa busca compreender quais são os parâmetros utilizados pela Meta e pelo *Oversight Board* para a retirada de conteúdos odiosos e se os critérios e parâmetros utilizados ocorrem de forma proporcional e efetiva, incorporando valores democráticos e respeito aos direitos humanos.

A metodologia utilizada é constituída, fundamentalmente, por um estudo qualitativo, descritivo e exploratório. Para tal, foi realizado um levantamento bibliográfico através de livros e artigos científicos de autores que deram sustentação teórica à pesquisa, bem como a realização de uma pesquisa documental, com a análise das decisões e materiais complementares disponibilizados pelo *Oversight Board* e pela Meta.

Assim, o artigo progride em três partes. Inicialmente foram introduzidas algumas características especiais que envolvem a composição e organização do *Oversight Board*. Em seguida, foi apresentado o percurso metodológico da seleção e organização dos casos a partir da técnica de análise de conteúdo e uma tabela com o resumo das 11 decisões selecionadas, de janeiro de 2021 a agosto de 2022.

² DOUEK, E; KRONICK, K. Facebook Releases an Update on Its Oversight Board: Many Questions, Few Answers. 2019. *Lawfare*. Disponível em: <https://www.lawfareblog.com/facebook-releases-update-its-oversight-board-many-questions-few-answers>. Acesso em: 22 ago. 2022.

³ BALKIN, J. Free Speech is a triangle. *Columbia Law Review*, v. 118, 2018.

Finalmente, a sistematização dos resultados ocorreu por meio de uma análise conjunta de todas as decisões, sendo expostas: (1) as características dos casos, (2) os padrões relevantes utilizados, (3) as declarações dos usuários, (4) as considerações da Meta, (5) a participação do público na submissão, (6) a análise e (7) as recomendações do *Oversight Board*.

Assim, foi possível observar que o *Board* considerou os Tratados Internacionais de Direitos Humanos como fonte suprema de deliberação, realizou uma análise combinada do conteúdo com o contexto específico e a conjuntura do país envolvido e demonstrou uma preocupação com temáticas sensíveis para a seleção dos casos. Além disso, encorajou a interação com terceiros durante o processo decisório e incentivou a *accountability* através da fiscalização das políticas da plataforma e da coleta de dados e informações da Meta, exigindo liberação de relatórios constantes e, influenciando, até mesmo, na alteração das políticas da empresa.

A pesquisa conclui com a importância do *Board* como mecanismo fiscalizador do ecossistema de moderação da Meta, principalmente em termos do discurso de ódio, abrindo caminhos para futuras investigações.

1 CONSIDERAÇÕES SOBRE O OVERSIGHT BOARD

Composto por 23 membros de todo o mundo, o OB inclui pessoas de uma variedade de origens culturais e profissionais para refletir a diversidade da comunidade⁴. Essa perspectiva diversificada e o equilíbrio geográfico são importantes, pois trazem diversidade no conhecimento cultural e linguístico para a moderação de conteúdo⁵ que ocorre em nível global com várias jurisdições locais.

Seguindo seus próprios princípios de voz, autenticidade, segurança, privacidade e dignidade, os poderes do *Board* são descritos como: solicitação de informações da Meta; interpretação das normas comunitárias considerando seus valores e normas de direitos humanos; instrução para permitir ou remover conteúdo, mantendo ou revertendo indicações; e, emissão de explicações fundamentadas de suas decisões e recomendações⁶.

⁴ OVERSIGHT BOARD. Meet the Board. Disponível em: <https://www.oversightboard.com/meet-the-board/>. Acesso em: 19 ago. 2022.

⁵ DOUEK, E; KRONICK, K. Facebook Releases an Update on Its *Oversight Board*: Many Questions, Few Answers. 2019. *Lawfare*. Disponível em: <https://www.lawfareblog.com/facebook-releases-update-its-oversight-board-many-questions-few-answers>. Acesso em: 22 ago. 2022.

⁶ OVERSIGHT BOARD. Governança. Disponível em: <https://www.oversightboard.com/governance/>. Acesso em: 19 ago. 2022.

O *Board* é totalmente financiado e gerenciado pelos *Trustees*, que recebem financiamento da Meta pela prestação de seus serviços. Eles também são responsáveis pela proteção de sua independência e eficácia⁷. Além de supervisionarem o financiamento e as atividades da administração e dos membros do OB, também são responsáveis pela indicação e remoção destes⁸. Essa estrutura de independência assemelha-se aos argumentos da separação de poderes e evoca alguns usos da metáfora do “Supremo Tribunal” exposta no próximo tópico.

Em suma, cinco grandes contribuições podem ser destacadas a partir do lançamento desta experiência em governança que também serão observadas na análise das decisões nas próximas sessões: transparência para o ecossistema de moderação de conteúdo; construção de uma razão pública; trazer legitimidade; aplicar, monitorar, alterar ou preencher brechas nas políticas da Meta e trazer diversidade⁹.

O *Board* deixa claro que o seu objetivo é de garantir uma perspectiva global, por isso a escolha de membros de diversas origens do mundo e com mandato de três anos, com direito a dois mandatos de renovação. Porém, abertamente ressalta que não tem por função aplicar leis locais, restringindo-se aos padrões da comunidade da plataforma¹⁰.

Além disso, no que diz respeito à transparência e comunicação de seus atos, o OB prevê em seu Estatuto a emissão de relatórios anuais com resumos dos processos julgados, análise dos processos apresentados por região, bem como as atitudes da Meta na implementação de suas decisões¹¹.

Ainda, é reforçada a ideia da necessidade de equilibrar a liberdade de expressão e seus limites, destacando a responsabilidade das empresas de definir os padrões e valores de proteção¹². Desta forma, as decisões do *Board* podem ser aplicadas pela Meta em casos de conteúdos idênticos e contextos paralelos¹³.

⁷ OVERSIGHT BOARD. Governança. Disponível em: <https://www.oversightboard.com/governance/>. Acesso em: 19 ago. 2022.

⁸ *Ibid.*

⁹ DOUEK, E. Facebook’s *Oversight Board*: move fast with stable infrastructure and humility. *North Carolina Journal of Law and Technology*, v. 21, 2019.

¹⁰ *Ibid.*

¹¹ OVERSIGHT BOARD. Oversight Board Bylaws. Disponível em: <https://www.oversightboard.com/sr/governance/bylaws>. Acesso em: 19 ago. 2022.

¹² OVERSIGHT BOARD. Governança. Disponível em: <https://www.oversightboard.com/governance/>. Acesso em: 19 ago. 2022.

¹³ *Ibid.*

São diversos os tipos de conteúdo qualificados para serem analisados pelo Board, sendo alguns deles: publicações/status, fotos, vídeos, comentários e compartilhamentos¹⁴.

Ainda, as informações que os membros do *Board* poderão ter acesso ao conteúdo alvo de apelação são: o conteúdo em questão, as informações sobre onde o conteúdo foi publicado, as informações básicas sobre a pessoa que postou o conteúdo, o histórico das revisões do conteúdo, o raciocínio adotado pelo Facebook sobre sua decisão e as políticas envolvidas no caso. Para facilitar na compreensão do contexto, o OB pode solicitar, ainda, informações adicionais como o envolvimento e o alcance do conteúdo e os detalhes das políticas adotadas¹⁵.

Não só os usuários podem levar seu conteúdo ao Board, caso discordem da decisão e do processo de apelação da Meta. A própria Meta pode encaminhar casos considerados “significativos e difíceis”¹⁶ diretamente ao OB e, a pessoa que postou o conteúdo será notificada e terá a oportunidade de fornecer mais informações e o contexto do que postou¹⁷.

2 AS DECISÕES DO *OVERSIGHT BOARD* SOBRE DISCURSO DE ÓDIO. METODOLOGIA

O primeiro relatório anual do *Board* revelou que mais de 8 em cada 10 usuários apelam para restaurar o conteúdo relacionado às regras da Meta sobre bullying, discurso de ódio ou violência¹⁸. Entre outubro de 2020 e dezembro de 2021, a segunda categoria

¹⁴ ROSEN, G. Usuários agora podem apelar sobre conteúdos mantidos no Facebook ou Instagram para o Comitê de Supervisão. 13 de abril de 2021. *Facebook*. Disponível em: <https://about.fb.com.br/news/2021/04/usuarios-agora-podem-apelar-sobre-conteudos-mantidos-no-facebook-ou-instagram-para-o-comite-de-supervisao/>. Acesso em: 25 ago. 2022.

¹⁵ JOHNSON, F. A Privacy-Protected Way to Refer Cases to the *Oversight Board*. 22 de outubro de 2022. *Facebook*. Disponível em: <https://about.fb.com/news/2020/10/privacy-protected-way-to-refer-cases-to-oversight-board/>. Acesso em: 25 ago. 2022.

¹⁶ HARRIS, B. Establishing Structure and Governance for an Independent *Oversight Board*. 17 de setembro de 2019. *Facebook*. Disponível em: <https://about.fb.com/news/2019/09/oversight-board-structure/>. Acesso em: 25 ago. 2022.

¹⁷ JOHNSON, 2022.

¹⁸ OVERSIGHT BOARD. Annual Report 2021. Junho de 2022. Disponível em: <https://www.oversightboard.com/news/322324590080612-oversight-board-publishes-first-annual-report/>. Acesso em: 21 ago. 2022.

com mais casos encaminhados ao OB foi o discurso de ódio (28,9%)¹⁹, o qual persistiu no primeiro trimestre de 2022²⁰.

Por mais que os números chamem a atenção, a dificuldade das plataformas em seguir padrões rotineiros para a verificação do discurso de ódio não surpreende, especialmente pelas particularidades envolvidas em sua identificação. Dentre as particularidades estão: variações linguísticas, intenção do usuário, contextos culturais e normas locais, bem como a complexidade para detectá-los por meio de inteligência artificial²¹.

O próprio Zuckerberg reconhece a necessidade de entender as “variedades linguísticas” envolvidas na identificação do discurso de ódio²². Assim, cabe ao *Oversight Board* assumir o papel de considerar as particularidades dos contextos e corrigir os erros, principalmente neste tipo de discurso.

A pesquisa empregou o método de abordagem de análise de conteúdo, o qual é descrito como “um conjunto de técnicas de análise das comunicações que utiliza procedimentos sistemáticos e objetivos de descrição do conteúdo das mensagens”²³.

Ainda, o processo metodológico desta pesquisa foi desenvolvido através de três fases. Primeiramente, a escolha dos casos se deu por meio do levantamento de todas as decisões envolvendo os padrões da comunidade da Meta sobre o tema desde o início de suas atividades até o desenvolvimento desta pesquisa, ou seja, de janeiro de 2021 a agosto de 2022.

O número de decisões coletadas diretamente pelo site do *Oversight Board* nesse período totalizou 12. Contudo, uma decisão específica sobre o ex-primeiro-ministro da Malásia, datada de 28 de janeiro de 2021, foi excluída do *corpus*, pois a postagem havia sido removida posteriormente pelo usuário, o que foi considerado pelo *Board* como uma “decisão de erro”, um caso de indisponibilidade.

¹⁹ OVERSIGHT BOARD. Annual Report 2021. Junho de 2022. Disponível em: <https://www.oversightboard.com/news/322324590080612-oversight-board-publishes-first-annual-report/>. Acesso em: 21 ago. 2022..

²⁰ *Idem*. Report for the First Quarter of 2022. Agosto 2022. Disponível em: <https://oversightboard.com/news/572895201133203-oversight-board-publishes-transparency-report-for-first-quarter-of-2022/>. Acesso em: 01 set. 2022.

²¹ DOUEK, E. Facebook's *Oversight Board*: move fast with stable infrastructure and humility. *North Carolina Journal of Law and Technology*, v. 21, 2019.

²² *Idem*. Zuckerberg's New Hate Speech Plan: Out With the Court and In With the Code. *Lawfare*. Disponível em: <https://www.lawfareblog.com/zuckerbergs-new-hate-speech-plan-out-court-and-code>. Acesso em: 25 ago. 2022.

²³ BARDIN, L. *Análise de Conteúdo*. São Paulo: Edições 70, 2016, p. 38.

Em seguida, todas as 11 decisões selecionadas foram organizadas na tabela abaixo pelo nome do caso, o país envolvido, um breve resumo dos fatos que envolveram o caso e a decisão final do OB.

CASO	PAÍS	RESUMO DO CASO	DECISÃO FINAL
Publicação de Myanmar	Myanmar, França, China	Um usuário em Myanmar postou em um grupo do Facebook um texto afirmando que há algo de errado com os muçulmanos (ou homens muçulmanos), psicologicamente ou com sua mentalidade.	Anula a decisão da Meta de remover
Armênios no Azerbaijão	Armênia, Azerbaijão	O usuário usou o termo “taziks” para descrever os azerbaijanos, que o usuário alegou serem nômades e não terem histórico comparado aos armênios.	Mantém a decisão da Meta de remover
Zwarte Piet	Holanda	Um usuário do Facebook na Holanda compartilhou um vídeo que mostrava uma criança encontrando três adultos, dois retratando “Zwarte Piet”, com os rostos pintados de preto e usando perucas afro.	Mantém a decisão da Meta de remover
Meme “Dois Botões”	Armênia, Peru, Estados Unidos	Uma adaptação do meme 'dois botões'. A bandeira turca substituiu o rosto do personagem de desenho animado e os dois botões vermelhos com “O Genocídio Armênio é uma mentira” e “Os armênios eram terroristas que mereciam”.	Anula a decisão da Meta de remover
Bot de Myanmar	Myanmar	O post discutiu maneiras de limitar o financiamento aos militares de Myanmar. “O povo de Hong Kong, porque os porras dos chineses torturaram eles, mudaram seus bancos para o Reino Unido e agora eles [os chineses] não podem mais encostar neles”.	Anula a decisão da Meta de remover
Protestos na Colômbia	Colômbia	Vídeo de manifestantes na Colômbia criticando o presidente colombiano após recente reforma tributária. Eles chamam o presidente em espanhol de “filho da puta” e “pare de ser bicha na TV”.	Anula a decisão da Meta de remover
Calúnias na África do Sul	África do Sul	Um usuário postou que a população branca possui e controla a maior parte da riqueza e que se um negro que mora em um bairro branco garante o mesmo tratamento, e que essa pessoa “não bate bem da cabeça”. Ele concluiu que ela seria um “escravo sofisticado”, “um preto esperto”.	Mantém a decisão da Meta de remover
Cinto Wampum	Estados Unidos, Canadá	Um artista indígena postou uma foto de um cinto de wampum (uma forma de arte que utiliza conchas para formar imagens). O objeto feito por ele mesmo e tinha como objetivo aumentar a conscientização	Anula a decisão da

		sobre o massacre de Kamloops. O título de seu trabalho era ' <i>Kill the Indian/Save the Man</i> ' ('Mate o índio/Salve o homem').)	Meta de remover
Raya Kobo	Etiópia	Um usuário etíope postou uma publicação em amárico que incluía alegações de que a Frente de Libertação do Povo Tigray (TPLF) matou e estuprou mulheres e crianças e saqueou propriedades civis em Raya Kobo na Etiópia. Ainda, alegou que civis da etnia Tigray ajudaram o TPLF com essas atrocidades.	Mantém a decisão da Meta de remover
Palavras árabes	Marrocos, Egito, Líbano	Uma conta pública do Instagram descrita como um espaço para discutir narrativas <i>queer</i> na cultura árabe postou uma série de fotos em um carrossel. Cada foto explicava uma palavra diferente que pode ser usada de forma pejorativa para se referir a homens com “trejeitos afeminados” no mundo árabe.	Anula a decisão da Meta de remover
Animação Knin	Croácia	Uma página do Facebook publicou uma versão editada da animação da Disney “O Flautista Encantado” com uma legenda “O músico de Čavoglave e os ratos de Knin”. O vídeo retrata uma cidade invadida por ratos que foi salva por um gaiteiro – padre. A letra da música homenageia um padre ortodoxo sérvio que foi líder das forças de resistência sérvias durante a Segunda Guerra Mundial.	Anula a decisão da Meta de remover

Finalmente, a interpretação dos documentos partiu da premissa de que a análise das primeiras decisões do *Oversight Board* sobre o discurso de ódio pode traçar algumas conclusões acerca das suas atividades iniciais de julgamento e apontar pistas sobre o processo de moderação da Meta no tema. Portanto, buscou-se compreender como este primeiro modelo de apelação de moderação de conteúdo privado está decidindo, quais recursos está utilizando e as características dos casos escolhidos.

3 INSIGHTS COMPARATIVOS ACERCA DAS DECISÕES SOBRE DISCURSO DE ÓDIO DO *OVERSIGHT BOARD*

3.1 CARACTERÍSTICAS DOS CASOS

Conforme ressaltado no item anterior, o destaque da categoria discurso de ódio na apresentação de recursos ao *Board* não é surpreendente. Variações linguísticas

(Publicação de Myanmar, Armênios no Azerbaijão, Bot de Myanmar, Protestos na Colômbia e Calúnias na África do Sul) e dificuldades de compreensão de termos específicos do contexto que necessitam de uma análise combinada de conteúdo com fotos (Dois botões e Palavras Árabes), vídeos (Animação Knin) e legendas (Cinto Wampun), foram controvérsias apresentadas nos casos.

O OB repetidamente reforça a magnitude do contexto local e a necessidade de interpretações individuais. A compreensão do conteúdo e da utilização da retórica do ódio deve considerar as características do cenário atual ou da conjuntura histórica para a qual foi direcionado, como o conflito armado em Armênios no Azerbaijão e Raya Kobo²⁴, o histórico de discriminação contra negros na Holanda em Zwart Piet, o contexto histórico do *Apartheid* na África do Sul e a violência étnica sérvia em Animação Knin.

Além disso, os obstáculos à identificação do discurso de ódio envolvem os limites da expressão que podem ser enquadrados como opinião política (Bot de Myanmar, Protestos na Colômbia), sarcasmo (Publicação de Myanmar) e humor ou sátira (Dois Botões, Palavras Árabes e Animação Knin).

Por fim, valorizando espaços online seguros, o *Board* também considerou temáticas sensíveis para a seleção dos casos. Contextos que envolviam comunidades marginalizadas e ameaçadas, como os povos indígenas no caso do Cinto Wampun e os falantes de árabe LGBTQIA+ no caso Palavras Árabes, são exemplos.

3.2 PADRÕES RELEVANTES

Os padrões relevantes utilizados nas decisões do OB e que mais tarde passaram a ser adotados pelas deliberações da Meta, foram um “mix de autoridades” organizados a partir de duas perspectivas. Primeiramente, o OB avaliou as decisões a partir dos padrões da comunidade e valores da própria empresa Meta. Posteriormente, elencou os casos através de Tratados do Direito Internacional dos Direitos Humanos.

Na análise dos casos, os padrões da Meta foram usados apenas preliminarmente. Os valores de voz, dignidade e segurança foram brevemente mencionados e os padrões da

²⁴ Na primeira decisão, “Publicação de Myanmar”, os contextos de conflitos regionais, as eleições gerais do país em novembro de 2020 e a violência antimuçulmana em geral foram amplamente discutidos pela Meta e pelo *Board*. No entanto, os fatores não foram considerados suficientes para “incitar qualquer forma de dano iminente”. OVERSIGHT BOARD. Decisão Publicação de Myanmar sobre muçulmanos. Disponível em: <https://www.oversightboard.com/decision/FB-I2T6526K/>. Acesso em: 27 ago. 2022, p 2.

comunidade, na maior parte das decisões, foram usados para categorizar o nível de discurso de ódio.

Em todos esses casos, o OB baseia-se principalmente nos Princípios Orientadores da ONU sobre Empresas e Direitos Humanos, no Pacto Internacional sobre Direitos Civis e Políticos (PIDCP), no Comentário Geral n.º 34, no Plano de Ação de Rabat e no Relatório do Relator Especial sobre a Promoção e Proteção do Direito à Liberdade de Opinião e Expressão (2019).

Ao interpretar o significado de liberdade de expressão, o OB utilizou o teste de três partes presente no artigo 19 do PIDCP para restrição de expressão, o qual se subdivide em: legalidade, legitimidade e necessidade e proporcionalidade.

Em relação à legalidade, o *Board* ressalta a necessidade de que a regra aplicada ao usuário seja clara e acessível, ou seja, todas as informações sobre a limitação do discurso devem ser apresentadas. A utilização dos pontos entre as letras do termo “т.а.з.и.к.и”, foi considerada como ciência do usuário de que estava reproduzindo um termo ilegal para os padrões do Facebook.

Já a segunda parte, a legitimidade ou objetivo legítimo, visa proteger o direito de terceiros e dar segurança contra danos previsíveis e intencionais. A metodologia de fundamentação utilizada foi de compreender o alvo do discurso, não sendo possível restringir a expressão com o único objetivo de proteger indivíduos de maneira geral.

Por fim, a proporcionalidade e a necessidade dizem respeito à restrição necessária para enfrentar a ameaça ao grupo-alvo, sendo dispensável a incitação direta à violência ou à discriminação. Aqui pode ser discutido se as intervenções menos invasivas, como rótulos ou outras medidas para reduzir a divulgação, não teriam proporcionado proteção adequada. Além disso, se o contexto para a remoção do conteúdo é proporcional, por exemplo, ao fato de alguns casos se encaixarem na crítica e expressão política²⁵ ou de estarem no contexto de conflito armado²⁶.

²⁵ OVERSIGHT BOARD. Decisão do Bot de Myanmar. Disponível em: <https://www.oversightboard.com/decision/FB-ZWQUPZLZ/>. Acesso em: 27 ago. 2022, p 12.

²⁶ *Idem*. Decisão dos armênios no Azerbaijão. Disponível em: <https://www.oversightboard.com/decision/FB-QBJDASCV/>. Acesso em: 27 ago. 2022.

3.3 DECLARAÇÃO DO USUÁRIO

A declaração do usuário pode ser considerada como um “devido processo legal para usuários”, sugere mais transparência e os ajuda a entender as regras da plataforma. Essa categoria permite que seja exposto o raciocínio do usuário que publicou o conteúdo ou daquele que o denunciou e é de grande importância na categoria de discurso de ódio, pois a Meta reconhece em sua política a relevância dos usuários indicarem sua real intenção diante das ambiguidades no verdadeiro significado de alguns termos.

Em todas as decisões, a declaração do usuário foi exposta. Para conhecer a real intenção do usuário e evitar a remoção do conteúdo, a declaração foi usada como argumento de raciocínio em decisões que envolviam sarcasmo (Publicação de Myanmar) e sátira (Dois Botões). Ademais, também foi empregada como autorreferencial ou de uma forma empoderadora (Palavras Árabes).

Na decisão Animação Knin, um usuário externo recorreu da postagem, mas o OB convidou o indivíduo que postou o conteúdo a fornecer uma declaração. No entanto, por tratar-se de um grupo, o administrador respondeu que “eles faziam parte da página apenas como associados de negócios”.

3.4 DECISÕES DA META

Um dos benefícios mais relevantes da análise das decisões do *Oversight Board* é poder compreender melhor o processo de moderação de conteúdo da Meta.

Primeiramente, de modo geral, foi possível observar que, dado o formato adotado nas decisões do OB, a plataforma passou a empregar parâmetros de análise semelhantes (como o teste de três partes para restrição da liberdade de expressão), citando, até mesmo, decisões anteriores do *Board* como precedente na sua fundamentação.

A empresa destacou que possui uma lista interna de termos caluniosos que é compilada após consulta à especialistas regionais e organizações da sociedade civil e que não há exceção para o uso de calúnias contra líderes políticos ou figuras públicas²⁷ e

²⁷ OVERSIGHT BOARD. Decisão Protestos na Colômbia. Disponível em: <https://www.oversightboard.com/decision/FB-E5M6QZGA/>. Acesso em: 29 ago. 2022, p 7.

confirmou a necessidade da exceção que permite conteúdos odiosos quando usados de forma autorreferencial ou de empoderamento²⁸.

Ainda, na decisão Raya Kobo, a Meta também anunciou que cataloga vários países como “países em risco” e que a Etiópia é designada como “um país em risco de Nível 1” e em “crise de alto nível”. No mais, que estabeleceu uma equipe multifuncional com foco específico no país que se reúne semanalmente para compreender e reduzir os riscos contínuos²⁹.

Referente ao sistema de revisão humana, a plataforma informou que muitas vezes ele acontece por empresas terceirizadas³⁰, as quais costumam usar o “processo de escalonamento” para enviar um caso controverso para as equipes especializadas internas da Meta que, posteriormente, reavaliam o conteúdo³¹. Normalmente, por meio desse processo, a empresa pode reverter ou não a decisão inicial de remoção³².

Diante disso, o processo de escalonamento visaria reduzir, especialmente, os casos de remoção de falsos positivos. Entretanto, na decisão Animação Knin, esse sistema se mostrou ineficaz. 362 usuários denunciaram o conteúdo como violador das políticas de discurso de ódio. 40 revisores humanos avaliaram a publicação como "não violadora". Como esse conteúdo foi considerado “inviolável” pelos revisores humanos da Meta um certo número de vezes, as denúncias dos usuários foram automaticamente rejeitadas por um sistema automático. Quando o caso foi designado ao *Oversight Board* para que fosse feita uma revisão completa, a equipe de política de conteúdo da Meta deu “outra olhada” nas suas últimas decisões sobre essa publicação e confirmou que, na verdade, ela violava os padrões da comunidade de discurso de ódio³³.

Mesmo com as imprecisões apresentados pelo caso, a Meta informou ao *Board* que incentiva os revisores a encaminharem “conteúdo em alta” e a transferirem quando estiverem em dúvida³⁴. Todavia, o exemplo demonstra que os moderadores ou analisaram

²⁸ *Idem*. Decisão Calúnias na África do Sul. Disponível em: <https://www.oversightboard.com/decision/FB-TYE2766G/>. Acesso em: 29 ago. 2022, p 10.

²⁹ *Idem*. Decisão Raya Kobo. Disponível em: <https://www.oversightboard.com/decision/FB-MP4ZC4CC/>. Acesso em: 29 ago. 2022, p 17.

³⁰ OVERSIGHT BOARD. Decisão ressignificação de palavras em árabe. Disponível em: <https://www.oversightboard.com/decision/IG-2PJ00L4T/>. Acesso em: 28 ago. 2022, p 11.

³¹ *Idem*. Decisão Animação Knin. Disponível em: <https://www.oversightboard.com/decision/FB-JRQ1XP2M/>. Acesso em: 30 ago. 2022, p. 15.

³² O processo de escalonamento foi citado no caso Animação Knin, porém mesmo após avaliação interna da equipe da Meta, o conteúdo não foi removido. OVERSIGHT BOARD. Decisão Animação Knin. Disponível em: <https://www.oversightboard.com/decision/FB-JRQ1XP2M/>. Acesso em: 30 ago. 2022.

³³ OVERSIGHT BOARD. Decisão Animação Knin. Disponível em: <https://www.oversightboard.com/decision/FB-JRQ1XP2M/>. Acesso em: 30 ago. 2022, p 7.

³⁴ *Ibid*, p. 15.

a publicação com suas próprias inclinações e preconceitos culturais ou possuem curto espaço de tempo de análise e seguem a tendência das demais avaliações.

3.5 SUBMISSÃO DE COMENTÁRIOS PÚBLICOS

A submissão de comentários públicos é a parte mais interessante para indivíduos e organizações, que contribuem com perspectivas valiosas e ajudam na tomada de decisões. Esse processo auxilia no envolvimento do público e das partes interessadas no caso de forma interativa em cada revisão³⁵.

Dentre as decisões selecionadas, a gama de envios de comentários públicos ficou entre 2 e 35. Em geral, os comentários sempre foram bastante diversificados, com opiniões variadas sobre os casos (as vezes de partes diretamente ligadas ao tópico em discussão³⁶), vários comentários sobre a postura da própria Meta em suas deliberações³⁷, bem como algumas sugestões sobre o processo de moderação.

Os comentários públicos foram abordados de diferentes maneiras pelo *Oversight Board*, como fonte argumentativa para a descoberta de intenção do usuário, auxílio de interpretação de algum termo local, ou apenas breve menção. No caso Dois Botões, um comentário foi mencionado e considerado como argumento no corpo da decisão (o que sugeriu que o meme de conteúdo removido não estava zombando das vítimas, mas demonstrando o negacionismo comum na Turquia³⁸).

Já no caso Protestos na Colômbia, os especialistas do OB e os comentários do público confirmaram que a palavra “marica” porta múltiplos significados. Por fim, no caso do Bot de Myanmar, a grande parte dos comentários públicos apontou o conteúdo do post como uma discussão política, o que definiu a linha argumentativa do Board.

Balkin afirma que a possibilidade de usuários rotularem ou identificarem conteúdos impróprios em plataformas online constitui a comunidade online e suas normas e que os

³⁵ DOUEK, E. Content Moderation as Administration. *Harvard Law Review*, v 136, 2022, p 74.

³⁶ Decisão Dois Botões (descendente de vítimas do genocídio armênio, organizações que estudam a natureza, causas e consequências do genocídio), decisão Bot de Myanmar (defensores de direitos humanos e organizações da sociedade civil com foco na liberdade de expressão e discurso de ódio em Myanmar explicaram o termo), decisão Calúnias na África do Sul (um comentário da *Association on American Indian Affairs* que explicou o título da obra de arte).

³⁷ Zwart Piet e Bot de Myanmar (“o ‘bot de ação administrativa’ está mal programado”, “Reconheçam que a utilização da IA resultam em efeitos desnecessários e desproporcionais”).

³⁸ OVERSIGHT BOARD. Decisão Dois Botões. Disponível em: <https://www.oversightboard.com/decision/FB-RZL57QHJ/>. Acesso em: 27 ago. 2022, p 11.

sistemas de moderação das empresas online dependem desse tipo de voz³⁹. Da mesma forma, conforme observado, as submissões de terceiros tem sido uma parte fundamental para o *Board* compreender o sentido das expressões locais e os diferentes contextos.

3.6 ANÁLISE DO OVERSIGHT BOARD

O *Board* utilizou tradução independente por especialistas linguísticos para complementar a interpretação dos casos, os quais procuraram compreender os contextos sociopolíticos e culturais relevantes⁴⁰. Além disso, também foi indicada documentação extra local⁴¹ para garantir, por exemplo, no caso Zwart Piet, a dignidade, a segurança e a voz dos negros na plataforma.

Em termos de viralidade do conteúdo, o mero número de visualizações não foi um fator considerado pelo OB para a remoção. Por exemplo, na decisão de Bot de Myanmar, houve cerca de meio milhão de visualizações e o conteúdo ainda foi mantido. Por outro lado, o conteúdo referente a Zwart Piet teve cerca de 1.000 visualizações antes de sua remoção. Isso reforça o fato de que os conteúdos estão, necessariamente, vinculados ao seu contexto e, consequentemente, o elemento humano no processo de moderação é indispensável.

Até mesmo a língua em que foi feita a publicação do conteúdo se tornou fonte de argumentação do OB. No caso Dois Botões, o conteúdo foi publicado em inglês e, além de não causar divergências de tradução por ser compreendida como uma língua de “fácil entendimento” para os moderadores, considerou-se que a postagem era dirigida a usuários em diferentes regiões globais e, portanto, não representaria risco específico para o país.

Com relação às políticas de discurso de ódio da Meta e suas exceções, o *Board* reforçou a importância da exceção para conteúdo que compartilhe discurso de ódio com

³⁹ BALKIN, J. Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation, *UC Davis Law Review*, Public Law Research Paper, n. 615, 2018, p. 57.

⁴⁰ Decisão Publicação de Myanmar (os tradutores do Board encontraram significados diferentes para a mesma expressão), decisão do Bot de Myanmar (para entender o significado pretendido, o Board usou um falante de birmanês localizado em Myanmar e outro falante de birmanês localizado fora de Myanmar), decisão de calúnias da África do Sul (para identificar se a palavra “k***ir” é uma palavra particularmente odiosa e prejudicial no contexto sul-africano).

⁴¹ Decisão de Publicação de Myanmar (citado os documentos de direitos humanos sobre o contexto relevante nas eleições de Myanmar), Decisão Zwart Piet (para reafirmar a representação de racismo estrutural de Black Piet, o Board citou a declaração de um Defensor Público Holandês das Crianças e o trabalho do Instituto Holandês de Direitos Humanos, além de vários estudos acadêmicos).

o objetivo de “condená-lo e conscientizá-lo”⁴². Isso se justifica pelos entraves no monitoramento em larga escala realizado pela plataforma com auxílio da inteligência artificial. No mesmo sentido, outro grande desafio enfrentado é para a identificação da sátira⁴³. Em relação à decisão Dois Botões, o *Board* notou que a Meta removeu uma exceção para humor de sua política de discurso de ódio após uma Auditoria de Direitos Civis concluída em julho de 2020.

A criação de uma exceção para conteúdo de valor jornalístico também foi mencionada pelo OB, mas o conselho afirma que ela não deve ser interpretada como uma ampla permissão para que o discurso de ódio permaneça disponível⁴⁴. Contudo, até a data de elaboração da pesquisa, a Meta não se manifestou em seus padrões acerca dessa exceção.

O OB também se preocupou com a postura dos moderadores de conteúdo da Meta em diferentes casos. Inicialmente, com a discrepância nas traduções de um mesmo conteúdo pela Meta e pelo Board⁴⁵ e a capacidade dos moderadores de analisarem um meme, bem como peças semelhantes de conteúdo satírico⁴⁶. Além disso, de acordo com o Board, a Meta deve fornecer ao público mais informações sobre sua lista de calúnias para permitir que os usuários regulem sua conduta⁴⁷ e garantir que seus moderadores tenham recursos e suporte suficientes para avaliar adequadamente os contextos relevantes⁴⁸. Um exemplo de tal recurso é que os padrões devem estar disponíveis no idioma em que os moderadores revisam o conteúdo⁴⁹. Não obstante, a Meta não tomou

⁴² Interessante mencionar que o reconhecimento da exceção de que as pessoas podem compartilhar discurso de ódio com o intuito de conscientização ou condenação foi acrescentado nos padrões da comunidade sobre discurso de ódio da Meta apenas na versão de 30 de junho de 2022, ou seja, após as deliberações e recomendações do Board. Cf. OVERSIGHT BOARD. Decisão Cinto Wampum. Disponível em: <https://www.oversightboard.com/decision/FB-L1LANIA7/>. Acesso em: 29 ago. 2022, p 10 e META. Discurso de ódio. Disponível em: <<https://transparency.fb.com/policies/community-standards/hate-speech/>> Acesso em: 25 nov. 2022.

⁴³ OVERSIGHT BOARD. Decisão Dois Botões. Disponível em: <https://www.oversightboard.com/decision/FB-RZL57QHJ/>. Acesso em: 27 ago. 2022.

⁴⁴ OVERSIGHT BOARD. Decisão Protestos na Colômbia. Disponível em: <https://www.oversightboard.com/decision/FB-E5M6QZGA/>. Acesso em: 29 ago. 2022.

⁴⁵ OVERSIGHT BOARD. Decisão Publicação de Myanmar sobre muçulmanos. Disponível em: <https://www.oversightboard.com/decision/FB-I2T6526K/>. Acesso em: 27 ago. 2022.

⁴⁶ *Idem*. Decisão Dois Botões. Disponível em: <https://www.oversightboard.com/decision/FB-RZL57QHJ/>. Acesso em: 27 ago. 2022.

⁴⁷ Decisão Protestos na Colômbia, reiterado na decisão ressignificação de palavras em árabe.

⁴⁸ OVERSIGHT BOARD. Decisão ressignificação de palavras em árabe. Disponível em: <https://www.oversightboard.com/decision/IG-2PJ00L4T/>. Acesso em: 28 ago. 2022

⁴⁹ Decisões de Bot de Myanmar e Ressignificação de palavras em árabe.

nenhuma providência e, sem demora, disse como resposta que os seus moderadores de conteúdo eram fluentes em inglês⁵⁰.

A preocupação persistente do OB com grupos marginalizados foi destacada no caso Palavras Árabes. O *Board* considerou apropriado investigar o raciocínio por trás das decisões de aplicação ou aumentar o treinamento ou a supervisão com relação às regras existentes para evitar a aplicação excessiva de restrição em discursos de membros de grupos cujos direitos já estão em risco⁵¹.

3.7 DECISÕES E RECOMENDAÇÕES DO BOARD

Com referência à política de discurso de ódio, o *Board* instou a Meta a aumentar a argumentação de danos e o que ela busca evitar ao banir o *blackface*⁵²; desenvolver ferramentas e exceções apropriadas para tratar de questões satíricas, como equipes de operações locais e dar-lhes tempo suficiente⁵³. Além disso, o OB sugeriu à Meta que incluísse uma breve explicação sobre a exceção de conteúdo com valor jornalístico, com exemplos ilustrativos de uma variedade de contextos, incluindo reportagens sobre protestos em grande escala⁵⁴. Nesse caso especial, por exemplo, a Meta se deparou com a mesma questão e controvérsias que outros tribunais possuem ao definir os limites do valor jornalístico de um conteúdo⁵⁵.

Além disso, o OB também reiterou à Meta a necessidade de realizar avaliações de precisão focadas na política de discurso de ódio que abranjam expressão artística e expressão sobre violações de direitos humanos⁵⁶.

Quanto às indicações processuais, a primeira diz respeito à transparência e ao cumprimento da legalidade na restrição à liberdade de expressão, informando os usuários

⁵⁰ OVERSIGHT BOARD. Decisão ressignificação de palavras em árabe. Disponível em: <https://www.oversightboard.com/decision/IG-2PJ00L4T/>. Acesso em: 28 ago. 2022, p 10.

⁵¹ OVERSIGHT BOARD. Decisão ressignificação de palavras em árabe. Disponível em: <https://www.oversightboard.com/decision/IG-2PJ00L4T/>. Acesso em: 28 ago. 2022, p 11.

⁵² OVERSIGHT BOARD. Decisão Zwart Piet. Disponível em: <https://www.oversightboard.com/decision/FB-S6NRTDAJ/>. Acesso em: 29 ago. 2022.

⁵³ *Idem*. Decisão Dois Botões. Disponível em: <https://www.oversightboard.com/decision/FB-RZL57QHJ/>. Acesso em: 27 ago. 2022, p. 17.

⁵⁴ *Idem*. Decisão Protestos na Colômbia. Disponível em: <https://www.oversightboard.com/decision/FB-E5M6QZGA/>. Acesso em: 29 ago. 2022, p 14.

⁵⁵ KRONICK, K; KADRI, T. Facebook v. Sullivan: Public Figures and Newsworthiness in Online Speech. *St. John's Legal Studies Research Paper*, n. 19-0020, 2019, p. 63.

⁵⁶ OVERSIGHT BOARD. Decisão Cinto Wampum. Disponível em: <https://www.oversightboard.com/decision/FB-L1LANIA7/>. Acesso em: 29 ago. 2022, p 18.

de forma clara e precisa sobre as decisões ou quando alterar sua determinação inicial⁵⁷, elencando as categorias e motivos, não apenas que a publicação analisada violou sua política de discurso de ódio⁵⁸.

No tocante aos processos de apelação, o *Board* destacou que a Meta deve aconselhar seus usuários antes de recorrer a ele, abrindo espaço que os permitam indicar se seu conteúdo se enquadra em uma das exceções à política de discurso de ódio. Além disso, como dito anteriormente, ressalta a importância da Meta priorizar os casos com riscos aos direitos humanos em uma revisão humana⁵⁹.

Com relação às listas internas de calúnias, o *Board* fez algumas recomendações específicas no caso Palavras Árabes. A Meta deve explicar os processos e critérios para designar quais calúnias e países são atribuídos cada lista específica e como eles são aplicados e auditados. Tal explicação deve incluir os processos e critérios para avaliar, precisamente, quando e onde a proibição de calúnias deve ser aplicada, bem como os critérios para remover/manter as calúnias nas listas específicas⁶⁰.

Com relação às indicações de procedimento para os moderadores, a Meta deve garantir que seus padrões internos de implementação estejam disponíveis no idioma em que os moderadores de conteúdo revisam a publicação⁶¹. Além disso, critérios claros devem ser desenvolvidos e divulgados, para que os revisores de conteúdo possam enviar o conteúdo de interesse público que seja potencialmente violador para revisão posterior⁶². Por fim, também deve ser investigado como a localização de um revisor afeta a capacidade dos moderadores de avaliarem com precisão o discurso de ódio e o contradiscurso da mesma ou de regiões diferentes⁶³.

Em resumo, das primeiras decisões de discurso de ódio do *Oversight Board* até agosto de 2022, houve sete decisões que anularam a decisão da Meta de remover o conteúdo e quatro que confirmaram a decisão de removê-lo. Além disso, três dessas

⁵⁷ *Idem*. Decisão Animação Knin. Disponível em: <https://www.oversightboard.com/decision/FB-JRQ1XP2M/>. Acesso em: 30 ago. 2022, p. 20.

⁵⁸ Decisões Armênio no Azerbaijão, Zwarie Piet, Calúnias na África do Sul e Cinto Wampum.

⁵⁹ OVERSIGHT BOARD. Decisão Dois Botões. Disponível em: <https://www.oversightboard.com/decision/FB-RZL57QHJ/>. Acesso em: 27 ago. 2022.

⁶⁰ OVERSIGHT BOARD. Decisão Ressignificação de palavras em árabe. Disponível em: <https://www.oversightboard.com/decision/IG-2PJ00L4T/>. Acesso em: 28 ago. 2022.

⁶¹ *Idem*. Decisão do Bot de Myanmar. Disponível em: <https://www.oversightboard.com/decision/FB-ZWQUPZLZ/>. Acesso em: 27 ago. 2022, p 13.

⁶² *Idem*. Decisão Protestos na Colômbia. Disponível em: <https://www.oversightboard.com/decision/FB-E5M6QZGA/>. Acesso em: 29 ago. 2022, p 15.

⁶³ *Idem*. Decisão Cinto Wampum. Disponível em: <https://www.oversightboard.com/decision/FB-L1LANIA7/>. Acesso em: 29 ago. 2022.

decisões analisadas pelo Board⁶⁴ foram entregues fora do prazo regimental previsto, o que sugere que o comitê, mesmo no início da sua atividade, já pode estar “sobrecarregado por carga de trabalho”⁶⁵.

CONCLUSÕES

Por meio da análise das decisões do *Oversight Board*, este estudo teve como objetivo compreender e apresentar as tensões existentes para a regulação do discurso de ódio online da Meta e se o seu processo de moderação é compatível com os valores democráticos e os direitos humanos.

Diante da análise foi possível observar que as características específicas que compõe o discurso de ódio, como humor, crítica política e as dificuldades de compreender o contexto e intenção do autor, são obstáculos intensificados pelo ambiente virtual e detecção automatizada. Diante disso, surge a importância do elemento humano nas disputas online, momento em que a criação do *Oversight Board* se mostrou uma grande vantagem.

Além disso, como estrutura de governança⁶⁶, espera-se que as redes cumpram obrigações básicas daqueles que governam populações em sociedades democráticas como transparência, procedimentos justos, explicações fundamentadas para decisões ou mudanças de política e a possibilidade de os usuários reclamarem da conduta e participarem da governança da instituição⁶⁷.

Nesse sentido, é possível argumentar que a iniciativa da Meta de fundação do *Board* aperfeiçoa esses procedimentos democráticos, com foco nas duas principais funções descritas por Douek: evidenciar fragilidades no processo de formação de políticas na Meta e tornar-se um fórum importante para o processo de raciocínio público⁶⁸.

Através das decisões, isso pôde ser constatado de diversas formas. A possibilidade de emissão de comentários públicos tem permitido a participação ativa de terceiros no

⁶⁴ Decisões Cinto Wampum, Raya Kobo e Animação Knin.

⁶⁵ DOUEK, E. SEWELL, T. Meta’s Oversight Board Often Turns in Its Homework Late. Does it Matter? *Lawfare*. 2022. Disponível em: <https://www.lawfareblog.com/metas-oversight-board-often-turns-its-homework-late-does-it-matter>. Acesso em 04 set. 2022.

⁶⁶ KLONICK, K. The New Governors: the people, rules, and processes governing online speech. *Harvard Law Review*, v. 131, 2018.

⁶⁷ BALKIN, J. Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation, *UC Davis Law Review*, Public Law Research Paper, n. 615, 2018, p. 54.

⁶⁸ DOUEK, E. Facebook’s *Oversight Board*: move fast with stable infrastructure and humility. *North Carolina Journal of Law and Technology*, v. 21, 2019, p. 7.

ambiente virtual de moderação e que, conforme exposto no item 3.3, pode apresentar respostas poderosas e interpretações basilares sobre os contextos locais.

Ademais, o sistema geral de moderação de conteúdo exige mudança constante e adaptação. Para isso, é imprescindível que suas decisões de restrição ou não de publicações sejam fundamentadas em argumentos e publicizadas através de relatórios constantes. Estes relatórios de atividades, liberação de dados aos pesquisadores e à sociedade e a formulação de regras claras, podem garantir a *accountability* na moderação de conteúdo por parte das plataformas⁶⁹.

À vista disso, outro ponto positivo a ser ressaltado que foi impulsionado pela criação do *Board* diz respeito à disponibilização de relatórios trimestrais pela Meta para dar visibilidade à comunidade das respostas da empresa ao comitê, com informações sobre o encaminhamento dos casos e seu progresso na implementação das recomendações.

Através de uma análise desses relatórios, foi possível constatar que a empresa adicionou expressamente a proibição de *blackface* nos seus padrões da comunidade e a exceção para conteúdos satíricos, conforme recomendação do *Board*. Outras recomendações ainda estão em fase de processamento de acordo com a empresa.

Ainda, uma outra característica importante do *Board* é o seu objetivo de prevenir discursos inflamados e não apenas puni-lo, dando ao público a chance de acompanhar as respostas da Meta, abrindo um diálogo sobre as possibilidades, restrições e incentivos na moderação de conteúdo. Na mesma lógica, ao utilizar parâmetros internacionais de direitos humanos, o comitê incentiva a superação da visão de que as normas de direitos humanos são unicamente destinadas ao Estado, encorajando a Meta a repensar o seu papel central no contexto social e sua responsabilidade coletiva.

Contudo, por mais que diversos pontos positivos da atuação do *Board* sejam destacados, é necessário reconhecer e reforçar o modelo de negócios das empresas de mídia social e que pelo órgão constituir uma recente iniciativa privada, ele precisa de amadurecimento. Ao estabelecer critérios de restrição ou não da liberdade de expressão, o OB pode abrir precedentes perigosos que podem ditar o futuro do discurso na rede e,

⁶⁹ BOWERS, J. ZITTRAIN, J. Answering impossible questions: Content governance in an age of disinformation. *Misinformation Review*. 14 de janeiro de 2020. Disponível em: <<https://misinforeview.hks.harvard.edu/article/content-governance-in-an-age-of-disinformation/>>. Acesso em: 20 set. 2021.

por tratar-se da empresa com maior número de usuários do mundo, seu impacto pode alcançar toda a esfera virtual.

REFERÊNCIAS BIBLIOGRÁFICAS

- BALKIN, J. Free Speech is a triangle. *Columbia Law Review*, v. 118, n. 07, 2018. Disponível em: https://columbialawreview.org/wp-content/uploads/2018/11/Balkin-FREE_SPEECH_IS_A_TRIANGLE.pdf. Acesso em: 28 ago. 2022.
- BALKIN, J. Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation, *UC Davis Law Review*, Public Law Research Paper, n. 615, 2018b. Disponível em: https://lawreview.law.ucdavis.edu/issues/51/3/Essays/51-3_Balkin.pdf. Acesso em: 30 ago. 2022.
- BALKIN, J. Old-School/New-School Speech Regulation. *Harvard Law Review*, v. 127, 2014. Disponível em: https://harvardlawreview.org/wp-content/uploads/2014/06/vol127_balkin.pdf. Acesso em: 30 ago. 2022.
- BARDIN, L. *Análise de Conteúdo*. São Paulo: Edições 70, 2016.
- BOWERS, J. ZITTRAIN, J. Answering impossible questions: Content governance in an age of disinformation. *Misinformation Review*. 14 de janeiro de 2020. Disponível em: <<https://misinforeview.hks.harvard.edu/article/content-governance-in-an-age-of-disinformation/>>. Acesso em: 20 set. 2021.
- DOUEK, E. Facebook's Oversight Board: move fast with stable infrastructure and humility. *North Carolina Journal of Law and Technology*, v. 21, 2019. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3365358. Acesso em: 12 ago. 2022.
- DOUEK, E. Zuckerberg's New Hate Speech Plan: Out with the Court and In With the Code. *Lawfare*. 14 de abril de 2018. Disponível em: <https://www.lawfareblog.com/zuckerbergs-new-hate-speech-plan-out-court-and-code>. Acesso em: 27 set. 2021.
- DOUEK, E.; KRONICK, K. Facebook Releases an Update on Its Oversight Board: Many Questions, Few Answers. 2019. *Lawfare*. Disponível em: <https://www.lawfareblog.com/facebook-releases-update-its-oversight-board-many-questions-few-answers>. Acesso em: 22 ago. 2022.
- DOUEK, E. The Facebook Oversight Board's First Decisions: Ambitious, and Perhaps Impractical. *Lawfare*. 2021. Disponível em: <https://www.lawfareblog.com/facebook-oversight-boards-first-decisions-ambitious-and-perhaps-impractical>. Acesso em: 20 set. 2022.
- DOUEK, E.; SEWELL, T. Meta's Oversight Board Often Turns in Its Homework Late. Does it Matter? *Lawfare*. 2022. Disponível em: <https://www.lawfareblog.com/metas-oversight-board-often-turns-its-homework-late-does-it-matter>. Acesso em 04 set. 2022.
- HARRIS, B. Establishing Structure and Governance for an Independent Oversight Board. 17 de setembro de 2019. *Facebook*. Disponível em: <https://about.fb.com/news/2019/09/oversight-board-structure/>. Acesso em: 25 ago. 2022.

JOHNSON, F. A Privacy-Protected Way to Refer Cases to the Oversight Board. 22 de outubro de 2022. *Facebook*. Disponível em:

<https://about.fb.com/news/2020/10/privacy-protected-way-to-refer-cases-to-oversight-board/>. Acesso em: 25 ago. 2022.

KLONICK, K. The New Governors: the people, rules, and processes governing online speech. *Harvard Law Review*, v. 131, 2018. Disponível em:
https://harvardlawreview.org/wp-content/uploads/2018/04/1598-1670_Online.pdf.
Acesso em: 15 jun. 2022.

KLONICK, K. The Facebook Oversight Board: Creating an Independent Institution to Adjudicate Online Free Expression. *The Yale Law Journal*, 2020. Disponível em:
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3639234. Acesso em: 18 jun. 2022.

KLONICK, K.; KADRI, T. Facebook v. Sullivan: Public Figures and Newsworthiness in Online Speech. *St. John's Legal Studies Research Paper*, n. 19-0020, 2019.
Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3332530. Acesso em: 16 nov. 2022.

META. *Discurso de ódio*. Disponível em:
<<https://transparency.fb.com/policies/community-standards/hate-speech/>> Acesso em:
25 nov. 2022.

META. *Padrões da Comunidade*. Disponível em: <https://transparency.fb.com/pt-br/policies/community-standards/hate-speech/>. Acesso em: 20 nov. 2022.

META. *Atualizações trimestrais da Meta sobre o Comitê de Supervisão*. Disponível em:
<https://transparency.fb.com/pt-br/oversight/meta-quarterly-updates-on-the-oversight-board>. Acesso em: 08 dez. 2022.

META. *Nossa abordagem para conteúdo interessante*. 25 de agosto de 2022.
Disponível em: <https://transparency.fb.com/pt-br/features/approach-to-newsworthy-content/>. Acesso em: 09 dez. 2022.

OVERSIGHT BOARD. *Annual Report 2021*. Junho de 2022. Disponível em:
<https://www.oversightboard.com/news/322324590080612-oversight-board-publishes-first-annual-report/>. Acesso em: 21 ago. 2022.

OVERSIGHT BOARD. *Conheça o Comitê*. 2022a. Disponível em:
<<https://oversightboard.com/meet-the-board>>. Acesso em: 01 set. 2021.

OVERSIGHT BOARD. *Rulebook for Case Review and Policy Guidance*. Novembro de 2020. disponível em: <https://oversightboard.com/sr/rulebook-for-case-review-and-policy-guidance>. Acesso em: 19 ago. 2022.

OVERSIGHT BOARD. *Governança*. 2022b. Disponível em:
<https://www.oversightboard.com/governance/>. Acesso em: 19 ago. 2022.

OVERSIGHT BOARD. *Oversight Board Bylaws*. 2022c. Disponível em:
<<https://www.oversightboard.com/sr/governance/bylaws>>. Acesso em: 19 ago. 2022.

OVERSIGHT BOARD. *Report for the First Quarter of 2022*. Agosto 2022d.
Disponível em: <https://oversightboard.com/news/572895201133203-oversight-board-publishes-transparency-report-for-first-quarter-of-2022/>. Acesso em: 01 set. 2022.

OVERSIGHT BOARD. *Decisão Publicação de Myanmar sobre muçulmanos*. 2022l.
Disponível em: <https://www.oversightboard.com/decision/FB-I2T6526K/>. Acesso em:
27 ago. 2022.

OVERSIGHT BOARD. *Decisão ressignificação de palavras em árabe.* 2022e.
Disponível em: <https://www.oversightboard.com/decision/IG-2PJ00L4T/>. Acesso em:
28 ago. 2022.

OVERSIGHT BOARD. *Decisão do Bot de Myanmar.* 2022f. Disponível em:
<https://www.oversightboard.com/decision/FB-ZWQUPZLZ/>. Acesso em: 27 ago. 2022.

OVERSIGHT BOARD. *Decisão dos Armênios no Azerbaijão.* 2022g. Disponível em:
<https://www.oversightboard.com/decision/FB-QBJDASCV/>. Acesso em: 27 ago. 2022.

OVERSIGHT BOARD. *Decisão Protestos na Colômbia.* 2022h. Disponível em:
<https://www.oversightboard.com/decision/FB-E5M6QZGA/>. Acesso em: 29 ago. 2022.

OVERSIGHT BOARD. *Decisão Calúnias na África do Sul.* 2022o. Disponível em:
<https://www.oversightboard.com/decision/FB-TYE2766G/>. Acesso em: 29 ago. 2022.

OVERSIGHT BOARD. *Decisão Raya Kobo.* 2022i. Disponível em:
<https://www.oversightboard.com/decision/FB-MP4ZC4CC/>. Acesso em: 29 ago. 2022.

OVERSIGHT BOARD. *Decisão Animação Knin.* 2022j. Disponível em:
<https://www.oversightboard.com/decision/FB-JRQ1XP2M/>. Acesso em: 30 ago. 2022.

OVERSIGHT BOARD. *Decisão Dois Botões.* 2022k. Disponível em:
<https://www.oversightboard.com/decision/FB-RZL57QHJ/>. Acesso em: 27 ago. 2022.

OVERSIGHT BOARD. *Decisão Cinto Wampum.* 2022m. Disponível em:
<https://www.oversightboard.com/decision/FB-L1LANIA7/>. Acesso em: 29 ago. 2022.

OVERSIGHT BOARD. *Decisão Zwarte Piet.* 2022n. Disponível em:
<https://www.oversightboard.com/decision/FB-S6NRTDAJ/>. Acesso em: 29 ago. 2022.

ROSEN, G. Usuários agora podem apelar sobre conteúdos mantidos no Facebook ou Instagram para o Comitê de Supervisão. 13 de abril de 2021. *Facebook*. Disponível em: <https://about.fb.com.br/news/2021/04/usuarios-agora-podem-apelar-sobre-conteudos-mantidos-no-facebook-ou-instagram-para-o-comite-de-supervisao/>. Acesso em: 25 ago. 2022.

SUZOR, N. Digital Constitutionalism: Using the Rule of Law to Evaluate the Legitimacy of Governance by Platforms. *Social Media + Society*, 2018.

Submetido em 31 de janeiro de 2024.

Aprovado para publicação em 22 de dezembro de 2025.

