



Anais do Encontro Anual
da Rede de Pesquisa em
Governança da Internet
Maio de 2023

VI ENCONTRO DA REDE DE PESQUISA EM GOVERNANÇA DA INTERNET - REDE 2023

FICHA TÉCNICA

ANAIS DA REDE DE PESQUISA EM GOVERNANÇA DA INTERNET-VOL. 6, ISSN 2675-1690

Editores

Maria Vitoria Pereira de Jesus
Rodolfo Avelino

Autores

Christiana Soares de Freitas
Isabelle Brito Bezerra Mendes
Jaqueline Trevisan Pigatto
João Araújo Monteiro Neto
Luis Henrique de Mendes Acioly
Luize Pereira Ribeiro
Maria Fernanda Amorim Fossaluza
Matheus Fernandes da Silva
Renata de Oliveira Miranda Gomes
Rhaiana Caminha Valois

COMITÊ ORGANIZADOR

Núcleo de Coordenação da Rede de Pesquisa em Governança da Internet

Alexandre Arns Gonzales
Carolina Batista Israel
Diego Vicentin
Fernanda R. Rosa
Gustavo Ramos Rodrigues
Hemanuel Jhosé Alves Veras
Kimberly de Aguiar Anastácio
Maria Vitoria Pereira de Jesus
Nathan Paschoalini Ribeiro Batista
Rodolfo Avelino

Comitê Científico

Alexandre Arns Gonzales
Carolina Batista Israel
Diego Vicentin
Edison Spina
Fernanda R. Rosa
Flávio Rech Wagner

Jean Santos
Leonardo Ribeiro da Cruz
Maria Mônica Arroyo
Rafael Evangelista
Raquel Gatto
Rodolfo Avelino
Sergio Negri

Moderadores

Carolina Batista Israel
Flávio Rech Wagner
Gustavo Ramos Rodrigues

Debatedores

Jaqueline Trevisan Pigatto
Luis Henrique de Mendes Acioly
Luize Pereira Ribeiro
Maria Fernanda Amorim Fossaluza
Renata de Oliveira Miranda Gomes
Rhaiana Caminha Valois

O VI Encontro da Rede de Pesquisa em Governança da Internet e a publicação dos Anais resultantes deste encontro receberam apoio de: Comitê Gestor da Internet no Brasil - CGI.br, Internet Society - Capítulo Brasil e Programa de Pós-graduação em Ciências Sociais, da Universidade Estadual de Campinas (Unicamp), através de recursos da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)





CONSTITUCIONALISMO DIGITAL E O *OVERSIGHT BOARD*: UM POSSÍVEL FUTURO PARA A MODERAÇÃO DE CONTEÚDO

LUIZE PEREIRA RIBEIRO

SUMÁRIO

INTRODUÇÃO	5
METODOLOGIA.....	7
CONSTITUCIONALISMO DIGITAL: UMA BREVE ANÁLISE	8
O “OVERSIGHT BOARD” DA META.....	12
ESTUDO SOBRE AS DECISÕES DO COMITÊ	15
RESULTADOS E DISCUSSÃO	23
CONSIDERAÇÕES FINAIS	27
REFERÊNCIAS	28



CONSTITUCIONALISMO DIGITAL E O *OVERSIGHT BOARD*: UM POSSÍVEL FUTURO PARA A MODERAÇÃO DE CONTEÚDO

Luize Pereira Ribeiro¹

RESUMO

Diante do crescimento desenfreado das mídias sociais e, por conseguinte, o conteúdo compartilhado nessas plataformas, surge a urgência de moderar esses dados a fim de evitar alastramento do ódio contra indivíduos de grupos vulneráveis, o cyberbullying, a pornografia infantil e a desinformação descontrolada. Dessa maneira, o Constitucionalismo Digital surge como fenômeno relevante para construção de regulações das plataformas e da moderação de conteúdo, influenciando a origem de entidades como o Oversight Board, da Meta. Sendo assim, pretende-se investigar a proposta do Comitê, com o propósito de compreender a disponibilidade do modelo para uma moderação de conteúdo mais transparente e positiva para o futuro da governança de plataformas. Destarte, visa assinalar a influência do Constitucionalismo Digital no formato da entidade e no possível futuro da moderação de conteúdo digital, avaliar qual vertente direcionada pelas principais decisões publicadas pelo Comitê nos últimos dois anos de atividade e checar a estrutura do Comitê de Supervisão a fim de compreender o modelo. Portanto, pretende-se validar a relevância e impacto do modelo criado no Oversight Board para o futuro da autorregulamentação e moderação nas plataformas digitais. O órgão apresenta proposta até então inédita, influenciada pelos princípios do Constitucionalismo Digital, e pode contribuir para a construção de ferramentas ainda mais transparentes, viáveis e eficientes para a estrutura descentralizada de valores fundamentais.

PALAVRAS-CHAVE

Constitucionalismo Digital; Moderação de conteúdo; Regulação; Oversight Board.

ABSTRACT

Given the rapid growth of social media, and therefore the content shared on these platforms, there is an urgent need to moderate this data in order to prevent the spread of hatred against individuals from vulnerable groups, the cyberbullying, the child porno-

¹ Bacharel em Humanidades pela Universidade Federal da Bahia (UFBA). Graduanda em Direito pela UFBA. Pesquisadora. E-mail: luize.pereira@ufba.br.

graphy and the uncontrolled misinformation. Hence, the Digital Constitutionalism emerges as a relevant phenomenon for the construction of platform regulations and content moderation, influencing the origin of entities such as Meta's Oversight Board. Thus, it is aimed to investigate the Committee's proposal, in order to understand the viability of the model for a more transparent and positive content moderation for the future of platform governance. Therefore, it aims to point out the influence of Digital Constitutionalism in the structure of the entity and in the possible future of digital content moderation, evaluate which direction the main decisions published by the Committee in the last two years of activity have taken, check the structure of the Oversight Board in order to understand the model. Therefore, it is intended to validate the relevance and impact of the model created in the Oversight Board for the future of self-regulation and moderation in digital platforms. The body presents a hitherto unprecedented proposal, influenced by the principles of Digital Constitutionalism and may contribute to the construction of even more transparent, feasible and efficient tools for the decentralized structure of fundamental values.

KEY WORDS

Digital Constitutionalism; Content Moderation; Regulation; Oversight Board.

INTRODUÇÃO

Com o desenvolvimento da Internet como centro de comunicação e informação, as mídias sociais se tornaram protagonistas dado o crescente número de usuários. Como resultado, as plataformas que administram esses espaços têm acesso constante a vasto volume de dados e a quantidade, a velocidade e a variedade de conteúdos são surpreendentes (GILLESPIE, 2020). Ademais, o suposto anonimato garantido pelo véu da Internet criou ambiente confortável para disseminação do ódio contra indivíduos de grupos vulneráveis, cyberbullying, pornografia infantil e desinformação descontrolada, principalmente através de notícias falsas (fake news).

Não obstante, como a mídia social tem se expandido, em especial, nas redes de interação online, também intensificou-se o desafio de moderá-la. As consequências do dano online se estendem além da plataforma onde ocorre os casos concretos, assim como as críticas as plataformas e suas falhas de moderação explodiram, após casos como Gamergate, Myanmar, a crescente cultura da pornografia de vingança (revenge porn), as eleições presidenciais americanas de 2016, Alex Jones e Christchurch (GILLESPIE, 2020).

Atualmente, com o grande volume de conteúdo não há capacidade suficiente para moderar todos os dados compartilhados. Como resultado, as plataformas são reguladas pelas diretrizes de seus termos de serviço. Além desta autorregulação, a

moderação ocorre através de dois tipos principais de moderação: (a) triagem ativa, geralmente, empregada pela inteligência artificial, e (b) análise humana das reclamações dos usuários (flagging) (HATZIPANAGOS, 2018).

Todavia, os sistemas de moderação apresentam diversos problemas, incluindo (a) opacidade adicional, tornando conjunto de práticas pouco transparente, difícil de entender ou auditar; (b) complicar ainda mais as questões pendentes de justiça e equidade em sistemas de grande escala, e (c) re-obscurar a natureza fundamentalmente política das decisões de fala sendo executadas em escala (LLANSO, 2020).

Nessa conjuntura, um dos desafios significativos quando se trata de moderação de conteúdo online é a garantia da liberdade de expressão. A abordagem baseada nos direitos humanos na moderação de plataformas é necessária, alinhando as regras de moderação de conteúdo com a legislação internacional de direitos humanos.

Diante desse prisma, a crescente do fenômeno do Constitucionalismo Digital parece influenciar o nascimento de novo modelo de autorregulação das plataformas, como será possível aprofundar no presente estudo. Esse fenômeno dispõe as grandes empresas ferramentas de organização que favorecem o fortalecimento de autorregulação, ainda que com os vieses imbricados desse setor. Sendo assim, o movimento vai além da dimensão estatal, pois se aplica às regras internas das poderosas empresas multinacionais de tecnologia, organizada com base em prescrições normativas comuns para o reconhecimento, afirmação e proteção dos direitos fundamentais no ciberespaço (MENDES; OLIVEIRA FERNANDES, 2020), com foco na proteção dos direitos fundamentais e no reequilíbrio do poder na governança (GILL; REDEKER; GASSER, 2015).

Como intermediários na gestão de conteúdo e dados, os crescentes poderes das plataformas digitais têm levantado preocupações sobre o impacto de suas ações nas garantias individuais relacionadas à privacidade, liberdade de expressão, censura, autodeterminação e acesso à informação (CELESTE, 2019). Esta mudança deslocou o centro de aplicação dos direitos fundamentais da esfera pública para a esfera privada (MENDES; OLIVEIRA FERNANDES, 2020).

Esse fenômeno é evidenciado, por exemplo, na construção dos Princípios de Santa Clara, os quais se propõem obter uma maior transparência e responsabilização da moderação de conteúdo em plataformas digitais. Os cinco princípios fundamentais caminham adjacentes ao Constitucionalismo Digital, pregando por uma gestão, em

especial, coerente e nítida. Dado o crescimento exponencial das mídias sociais e dos dados compartilhados nesses espaços, ainda é um desafio diário construir soluções em compasso ao engrandecimento. Remanescem obstáculos a serem ultrapassados nessa corrida da moderação de conteúdo e o fenômeno constitucionalista aparece como uma das possíveis soluções.

Influenciada por esta corrente, em 2020, Meta fundou oficialmente o Oversight Board ou Comitê de Supervisão, conselho independente responsável pela revisão de casos concretos e significativos sobre o conteúdo do Facebook e do Instagram. O Oversight Board deixa claro que visa garantir o respeito à liberdade de expressão e os direitos fundamentais por meio do julgamento independente (META, 2021).

O órgão global composto por especialistas na área é organizado e estabelecido por um regulamento interno, apresentando modelo de governança diferente, com códigos internos privados, organização institucional com membros e apoio financeiro (BARBOSA, 2023). A primeira decisão foi publicada pela entidade em janeiro de 2021 e, até hoje (junho de 2023), a autoridade já tomou mais de 40 decisões, incluindo quatro casos brasileiros (META, 2023).

METODOLOGIA

Assim sendo, o presente estudo tem como objetivo investigar a proposta do Oversight Board, a fim de compreender a disponibilidade do modelo para uma moderação de conteúdo mais transparente e positiva para o futuro da governança de plataformas. Entre os objetivos específicos, visa assinalar a influência do Constitucionalismo Digital no formato da entidade e no possível futuro da moderação de conteúdo digital, checar a estrutura do Comitê de Supervisão a fim de compreender o modelo e, por fim, avaliar a atuação do Comitê através das principais decisões publicadas nos últimos dois anos de atividade. A presente pesquisa será conduzida por abordagem interdisciplinar, apelando às diversas disciplinas das ciências humanas, operando outras perspectivas e multiplicando os ângulos de questionamento e de visão (LAVILLE, 1999). O estudo pretende seguir o estilo de pesquisa exploratória e explicativa, posta a necessidade de investigar os aspectos estruturantes do objeto, com o intuito de contribuir na percepção e debate do problema.

Sendo assim, a natureza é de natureza qualitativa, tendo em vista o uso de técnicas como a revisão bibliográfica narrativa, a pesquisa teórico-conceitual do Cons-

titucionalismo Digital e a análise documental dos divulgados pelo Comitê. Propõe-se o levantamento das informações disponíveis sobre a estrutura do Oversight Board no site oficial do Comitê e da Meta, além do agrupamento das principais decisões, especialmente aquelas que envolvem matéria brasileira.

Faz-se relevante sinalizar que a revisão narrativa foi escolhida no presente estudo visando garantir uma abrangência na coleta documental que possa ser relevante para a construção de uma contextualização do contexto em questão, dado que a revisão da literatura narrativa apresenta uma temática mais aberta, não exigindo um protocolo rígido para sua confecção (UNESP, 2015, p. 4).

Portanto, pretende-se validar a relevância e impacto do modelo criado no Oversight Board para o futuro da autorregulação e moderação nas plataformas digitais. O órgão apresenta proposta até então inédita, influenciada pelos princípios do Constitucionalismo Digital e pode contribuir para a construção de ferramentas ainda mais transparentes, viáveis e eficientes para a estrutura descentralizada de valores fundamentais. A partir do destrinchamento do modelo, pretende-se compreender como são realizadas as decisões, qual tem sido o posicionamento de maneira geral do Comitê e qual o efeito dessa estrutura na moderação de conteúdo. Destarte, acredita-se que o presente estudo pode colaborar no debate e na produção científica acerca do tema, principalmente tratando da importância e novidade do objeto.

Diante desse panorama, a pesquisa apresenta campo ainda pouco explorado pela comunidade acadêmica brasileira, desenvolvendo matéria significativa para constituição dos estudos disponíveis na literatura.

O estudo trata em primeiro momento sobre o fenômeno do constitucionalismo digital, para, em seguida, apresentar o Oversight Board, expondo sua estrutura e comportamento nas decisões, a fim de verificar, nas discussões e resultados, a possibilidade dessa configuração ser o futuro para a moderação de conteúdo ou até mesmo, trazer novas questões ao debate.

CONSTITUCIONALISMO DIGITAL: UMA BREVE ANÁLISE

Antes de compreender a influência do constitucionalismo digital na origem e estrutura do Oversight Board da Meta, faz-se necessário compreender qual definição o presente estudo adotou, posto que o conceito desse fenômeno ainda está em debate e construção. Em verdade, autores como Redeker, Gill e Gasser definem o constitu-

cionalismo digital como uma “constelação de iniciativas que procuraram articular um conjunto de direitos políticos, normas e limites de governança sobre o exercício de poder no âmbito da Internet” (REDEKER; GILL; GASSER, 2018, p. 2).

Essa característica é motivo de crítica de alguns pesquisadores que acreditam que o conceito é excessivamente dilatado, pois designa fenômenos muito distintos entre si (PEREIRA; KELLER, 2022, p. 2679). Em contrapartida, Suzor defende que a solução para uma definição mais específica seja figurá-la como projeto que visa dar parâmetros e legitimidade para governança na era digital para articular limites a serem impostos à autonomia das plataformas. Sob esse prisma, o consentimento, a previsibilidade e a equidade processual são os princípios procedimentais do Estado de Direito que materializam a ideia de “boa governança” (SUZOR, 2018).

Nessa mesma linha, Celeste admite que “este é um novo momento constitucional” (2023, p. 24), que oportuniza a origem de um constitucionalismo digital pautado em prescrições normativas comuns de reconhecimento, afirmação e proteção de direitos fundamentais no ciberespaço. Essa ideologia constitucional se estrutura em um quadro normativo de proteção dos direitos fundamentais e de reequilíbrio de poderes na governança do ambiente digital (FERREIRA; OLIVEIRA, 2020).

Destarte, Celeste (2023) nota que esse é o fenômeno consequente da globalização e revolução digital, favorece o surgimento de novos atores poderosos para além do Estado, órgão até então central sob égide constitucional tradicional. Nessa nova era, as plataformas digitais podem influenciar a vida dos indivíduos em escala global, de maneira muito semelhante aos Estados (CELESTE, 2023, p. 12-15). Essas empresas privadas, ao determinarem como os indivíduos podem usar esses instrumentos digitais sob sua administração, interferem, inevitavelmente, no exercício de seus direitos fundamentais (CELESTE, 2023).

Ainda em 2018, o Relator Especial da ONU sobre Liberdade de Expressão, David Kaye, publicou um relatório sobre moderação de conteúdo on-line, a atividade de monitoramento e gerenciamento de conteúdo gerado pelo usuário realizada por plataformas on-line. O Relator Especial incentiva as empresas privadas a garantir a “legalidade” de suas decisões, ou seja, exige que esses atores definam a priori seus princípios de moderação de conteúdo e evitem escolhas arbitrárias. Kaye ainda se refere às práticas normativas das empresas de mídia social como a “lei da plataforma” (2018, p. 17).

O relatório também enfatiza a necessidade de oferecer a possibilidade de recorrer das decisões, propondo a ideia de instituir um “conselho de mídia social” independente (2018, p. 18). Como será possível compreender na próxima seção, esse é o formato utilizado pela Meta na estruturação do Oversight Board, a partir de 2019. Entretanto, a origem do Comitê não é o primeiro ato da Meta em concordância com os parâmetros do constitucionalismo digital.

Celeste (2023) destaca três principais aspectos objetivos que as plataformas influenciadas pelo fenômeno constitucional comumente utilizam: (i) o uso de termos que claramente evocam uma dimensão constitucional nos títulos dos documentos regulatórios, como “declaração”, “carta” ou “estatuto”; (ii) o jargão característico das constituições; (iii) articulam direitos individuais fundamentais e estabelecem obrigações (CELESTE, 2023, p. 44-45). No caso da Meta, em especial, o autor expõe ainda que, até maio de 2018, os termos de uso do Facebook eram chamados de “Declaração de direitos e responsabilidades”, descritor que ecoa explicitamente o conteúdo de textos constitucionais. Pettrachin (2018) explica que essa é uma das características mais influentes no movimento constitucional na Internet:

O progresso na definição dos direitos fundamentais no ecossistema de governança da Internet, em vez disso, precisa ser definido pelas “instâncias menores em que os atores contribuem com princípios para o ecossistema” (Cerf et al., 2014: 60). Sob essa perspectiva, os atores que redigem novos documentos proclamando direitos humanos relacionados à Internet estão todos participando de um processo de formação de discurso e de criação de um regime constitucional (Jakab, 2016: 3). Avaliando até que ponto o “vocabulário constitucional” e as conceitualizações usadas por esses atores são compartilhados ou amplamente aceito é, portanto, importante para identificar tensões, dimensões e possíveis desenvolvimentos desse processo.” (PETTRACHIN, 2018, p. 3).

Retira-se dessa análise que, entre as particularidades influentes do constitucionalismo digital, as grandes empresas multinacionais de tecnologia na formação de regras e governança internas está, especialmente, o uso dos jargões constitucionais para dar formalidade ou, em outras palavras, “uma mera estratégia de marketing, para construir uma espécie de fachada constitucional” (CELESTE, 2023, p. 52), criando um panorama (distante da realidade) para o usuário de zelo das plataformas para com os seus direitos fundamentais. Esse é um movimento similar ao “ethics washing” (BIETTI, 2021, p. 210), já notado nos estudos sobre o uso de Inteligência Artificial das Big Techs,

na qual a crescente instrumentalização da linguagem ética pelas empresas tecnológicas é chamada de “lavagem ética” (“ethics washing”).

Para além dessa característica, observa-se uma concentração vultosa de poder nessas estruturas de autorregulação das plataformas digitais, posto que essas são ao mesmo tempo legisladoras das diretrizes e funcionamento, executoras ao administrar o ciberespaço de sua responsabilidade e, como será possível aprofundar com o exemplo do Oversight Board, arbitram sob suas próprias regras. Ferreira e Oliveira (2020) destacam:

De fato, as plataformas digitais exercem uma função normativa importante ao estabelecer regulamentos e termos de uso dos seus serviços. Embora representem simples contratos entre as partes, em muitos casos, esses documentos adotam jargões típicos de textos constitucionais que projetam, na relação privada, direitos como o de livre acesso e compartilhamento de informações e o direito de estabelecer controle de privacidade sobre os seus dados. (...) Ao fazer cumprir os regulamentos e termos de uso pactuados, as empresas se engajam em uma verdadeira função adjudicatória de direitos (FERREIRA; OLIVEIRA, 2020, p.14-15)

Para além da vinculação normativa, dotada pelo constitucionalismo digital, o fenômeno destaca a relevância de uma autorregulação em conjunto com os princípios fundamentais. Como fora sinalizado na introdução, atualmente, a moderação de conteúdo é desafio das plataformas digitais, matéria que tem efeitos diretos sobre as garantias individuais, como privacidade, liberdade de expressão, censura, autodeterminação e acesso à informação.

A governança das plataformas levanta preocupações constitucionais fundamentais, no sentido de observar com zelo as responsabilidades legais e sociais sobre como esses espaços sociais são constituídos e como o exercício do poder deve ser limitado (SUZOR, 2018), aspectos até o momento não endereçados em definitivo pelas empresas *tech* do ramo. Em especial, observa-se no campo da moderação de conteúdo a carência do tratamento constitucional descrito nos termos de uso e diretrizes das mídias sociais, em geral. Devido a massificação no volume de dados, a moderação se resume aos modelos arbitrados pela decisão automatizada ou o flagging (HATZIPANAGOS, 2018), os quais já demonstraram não serem tão eficientes na proteção dos direitos fundamentais ao preservarem padrões repetidos no ambiente offline.

O compartilhamento da esfera pública para a privada na aplicação dos direitos fundamentais destacada pelo constitucionalismo digital ainda não conferiu, na rea-

lidade, uma gestão mais transparente nas decisões de moderação. Por essa razão, chama atenção o formato proposto pelo Oversight Board como possível futuro para a moderação de conteúdo. Esse modelo destaca princípios procedimentais relevantes para o constitucionalismo digital como: o consentimento, a previsibilidade e a equidade processual (SUZOR, 2018).

O “OVERSIGHT BOARD” DA META

Em novembro de 2018, o CEO do então Facebook, Mark Zuckerberg, anunciou o projeto para governança e aplicação de conteúdo da plataforma, no qual listou novas ferramentas que poderiam garantir responsabilidade à governança e legitimidade às regras que regem grandes comunidades na internet. Entre as propostas, Zuckerberg cita o planejamento na criação de um órgão independente recursal para as decisões de moderação de conteúdo. Para ele, o Facebook não deve tomar sozinho tantas decisões importantes sobre liberdade de expressão e segurança (ZUCKERBERG, 2018).

Faz-se relevante destacar os três aspectos motivadores da origem do modelo, segundo o projeto. Em primeiro lugar, o Comitê evitará a concentração excessiva de decisões em suas equipes. Seguida da promoção da prestação de contas e a supervisão, permitindo maior transparência. E por fim, em terceiro caso, a garantia que essas decisões sejam tomadas no melhor interesse da comunidade e não por motivos comerciais (ZUCKERBERG, 2018). Sob essa égide, o Facebook retira um pouco da responsabilidade nos efeitos da moderação, terceirizando as decisões polêmicas fora da empresa, como no caso do banimento do perfil do ex-presidente dos Estados Unidos, Donald Trump.

Para além das motivações benevolentes destacadas por Zuckerberg, Douek sinaliza quatro motivos aparentes pelos quais essa estrutura de governança do Oversight Board é extremamente positiva mesmo diante dos dilemas de moderação de conteúdo consequentes do grande volume de dados compartilhados nas plataformas: (i) confere às decisões de moderação de conteúdo uma aura de legitimidade, ajudando nas relações com os usuários; (ii) oportunidade de moldar uma regulamentação governamental mais ampla, (iii) terceirizar decisões polêmicas para fora da empresa; e (iv) facilitar uma melhor aplicação das normas existentes (DOUEK, 2019, p. 17). Nota-se que há nesses pilares a disposição das ideias postuladas pelas teorias estudadas pelo constitucionalismo digital, na construção estrutural, sendo assim:

“O Conselho de Supervisão proposto pelo Facebook é um dos projetos de constituição mais ambiciosos da era moderna. Com a deslegitimação da governança pré-existente das plataformas tecnológicas no “techlash” em curso, isso representa um momento crucial em que podem surgir novas formas constitucionais que moldarão o futuro do discurso on-line.” (DOUEK, 2019, p. 5)

Antes do começo da atuação do Oversight Board, o órgão já era comumente referido como uma “Suprema Corte do Facebook”, termo cunhado pelo professor de Harvard, Noah Feldman (SULLIVAN, 2019). O uso dessa terminologia espelha legitimidade associada a uma suprema corte convencional, a qual pode, por sua vez, se estender ao próprio Comitê (COLWS, 2022, p. 5). Colws expõe que o uso dessas metáforas constitucionais “carregam conotações socioculturais e políticas, e sua ascensão discursiva mapeia essas conotações no objeto discutido, na medida em que isso acaba aumenta a legitimidade e o poder do objeto” (2022, p. 19).

Em contrapartida, autoras como Kate Klonick, Evelyn Douek e Rachel Griffin criticam esse uso. Griffin (2021) acredita que essa descrição é extremamente problemática, pois retrata a instituição como autoridade legítima, o que legitima seu próprio poder sobre o discurso on-line. Klonick reconhece que “a analogia com um tribunal é valiosa, mas também imperfeita” (2020, p.2476). O problema é que a confiança esmagadora nessa analogia está excluindo outras analogias (ou, de fato, descrições literais), que podem ser mais valiosas para o debate.

Destarte, o presente estudo caminha com esse último posicionamento, enquadrando o Oversight Board como uma instituição de autorregulação corporativa voluntária, em vez de um tribunal, facilita debates acadêmicos, como tem sido comumente classificada (GRIFFIN, 2021), de maneira que, isso também captura com mais precisão o contexto no qual o Comitê opera.

Dessa forma, após consulta pública que incluiu contribuições de especialistas do mundo inteiro, como acadêmicos, especialistas técnicos, advogados, designers e tecnólogos, em setembro de 2019, publicou-se a versão definitiva do estatuto do órgão, na qual estabeleceu-se um fundo fiduciário independente e desenvolveu regulamentos internos do comitê. Expõe-se no estatuto uma preocupação, em especial, com a liberdade de expressão com parâmetro para regulação, como um direito fundamental que pode colocar em risco a capacidade de outras pessoas de se expressarem livremente (OVERSIGHT BOARD, 2019).

O Comitê começou a examinar casos em 2020, visando formular decisões binárias na autorização ou remoção do conteúdo no Facebook e Instagram, com base no respeito à liberdade de expressão e aos direitos humanos, bem como realizar recomendações para esclarecer melhor nossas políticas e aprimorar os processos de revisão de conteúdo. Os pilares principiológicos são a independência, autoridade, acessibilidade e transparência (OVERSIGHT BOARD, 2023).

Atualmente o Oversight Board (OB) é composto por 22 membros representantes de diversas partes do mundo, incluindo indivíduos de distintas origens culturais e profissionais, visando refletir a diversidade da comunidade (OVERSIGHT BOARD, 2023). Essa perspectiva diversificada e o equilíbrio geográfico são de suma importância, pois proporcionam uma ampla variedade de conhecimentos culturais e linguísticos para a moderação global de conteúdo, considerando as diversas jurisdições locais (BARBOSA, 2023), mas como será possível observar na análise das decisões, essa diversidade entre as regiões dos casos não é notada.

O Conselho reforça no Estatuto e em boa parte das publicações os seguintes pilares funcionais: independência, julgamento, transparência e legitimidade. É evidente aqui que o Comitê tem forte influência da corrente constitucionalista.

Ademais, no que se refere à transparência e comunicação de suas ações, o Conselho de Supervisão prevê, em seu Estatuto, a publicação anual de relatórios contendo resumos dos casos julgados, análises dos processos recebidos por região e uma prestação de contas das medidas adotadas pela Meta na implementação das decisões (OVERSIGHT BOARD, 2023).

Até o momento, o Comitê já endereçou 193 recomendações, sendo que a maioria discute questões de transparência na disposição dos termos de funcionamento da plataforma, essas não têm valor vinculativo para a Meta, sendo da escolha da empresa adotar as orientações. Em contrapartida, as decisões sob casos concretos são irrecorríveis. Até o fim de junho de 2023, o Comitê entrou em contato com 42 casos. Entre eles, 1 (um) foi cancelado/revogado e 41 (quarenta e um) têm decisões aplicadas e publicadas.

É de fácil interpretação que a atual estrutura não permite um maior acolhimento de demandas diante do volume de usuários da Meta. Atualmente, o Facebook tem 2,9 bilhões de usuários ativos, segundo relatório de abril de 2023, o que torna impossível que o formato do comitê, ou o algoritmo nas decisões automatizadas ou o flagging, mo-

dere o conteúdo compartilhado. Somente em 2021 o Oversight Board somou mais de 1 milhão de casos submetidos para julgamento, enquanto apenas 21 decisões foram publicadas neste ano.

Mesmo que fosse tecnologicamente possível treinar a IA para apreciar o infinito espectro de nuances humanas, Douek (2019) observa mais dois motivos principais pelos quais a IA não pode ser uma resposta completa ao problema da moderação de conteúdo. Primeiro, a IA não dá a uma pessoa que tem uma decisão tomada contra ela a substância ou a sensação do devido processo legal ou de ser ouvida. A IA também não dá uma justificativa pública para suas decisões, o que, conforme argumentado abaixo, é um componente essencial de um sistema legítimo de moderação de conteúdo.

Em segundo lugar, e ainda mais fundamental, a IA não pode ser uma resposta completa para as dificuldades da moderação de conteúdo porque a IA não pode determinar a questão anterior dos valores que devem ser codificados nos algoritmos. Diante desses pontos, nota-se que o Oversight Board possibilita que o usuário tenha a oportunidade de ser ouvido e, em seguida, considerará todo o contexto relevante e os valores concorrentes em jogo no caso em questão, como as questões culturais e linguísticas, as quais são grandes obstáculos em outras modalidades de moderação. Contudo, o mais importante é que o Comitê oferece uma explicação detalhada para o fato de o conteúdo ser ou não mantido/permitido na plataforma da Meta.

ESTUDO SOBRE AS DECISÕES DO COMITÊ

Tendo em vista esse panorama, pretende-se compreender o comportamento do Comitê de Supervisão da Meta através de uma busca nas decisões do órgão. A pesquisa quantitativa tem como principal característica a unicidade da forma de coleta e tratamento dos dados. Dessa forma, foram coletadas 42 decisões publicadas.

No entanto, uma decisão específica relacionada ao ex-primeiro ministro da Malásia, com data de 28 de janeiro de 2021, foi excluída do conjunto de dados, pois a postagem em questão havia sido removida posteriormente pelo próprio usuário. Esse fato foi considerado pelo Conselho como uma “decisão equivocada”, um caso de indisponibilidade, portanto, desconsiderado.

Posteriormente, todas as 41 decisões selecionadas foram organizadas em uma base de dados permitindo o levantamento das porcentagens e uma análise detalhada das matérias acolhidas pelo Comitê. O Conselho adota na identificação das decisões al-

guns padrões, como o título da decisão; o número do processo; a região do caso; os países envolvidos; a mídia social do ocorrido; a data da publicação; o conteúdo moderado/a; norma da comunidade que foi violada; os direitos envolvidos; se a decisão foi mantida ou revertida; e o resumo do processo. Todas essas etiquetas foram adotadas na base, assim como as classificações do conteúdo moderado e direitos envolvidos no caso.

Para a melhor exposição da revisão sistemática realizada, estão dispostas na tabela abaixo uma visão geral das decisões coletadas, listando o nome do caso, a classificação da matéria moderada, um breve resumo dos fatos relacionados ao caso e a decisão final tomada pelo Conselho de Supervisão. Em seguida, discutirá os dados levantados.

Tabela 1: Análise das decisões coletadas do Oversight Board de janeiro de 2021 a junho de 2023.

Caso	Conteúdo violado moderado	Resumo do Caso	Decisão Final
Publicação de Myanmar sobre muçulmanos	Discurso de ódio	Usuário fez publicação questionando a falta de reação dos muçulmanos, assumindo que “os muçulmanos devem ter algum problema psicológico”. Essa declaração foi considerada um discurso de ódio pela Meta, mas o Comitê considerou que, embora a publicação possa ser considerada pejorativa ou ofensiva em relação aos muçulmanos, ela não defendeu o ódio nem incitou intencionalmente qualquer forma de dano iminente.	Reverte a decisão da Meta de remover, mantendo o conteúdo na rede.
Armênios no Azerbaijão	Discurso de ódio	O usuário usou o termo “тазики” (“taziks”) para descrever os azerbaijanos, um termo depreciativo para os azerbaijanos que aparece na lista interna de termos de difamação da Meta. A maioria do Comitê também considerou a remoção da publicação do Facebook adequada.	Mantém a decisão da Meta de remover, retirando o conteúdo da rede.
Sintomas de câncer de mama e nudez	Nudez adulta e atividades sexuais	Publicação com imagens de mamilos femininos com um título em português cujo intuito era aumentar a conscientização sobre os sinais do câncer de mama. publicação foi removida por um sistema automatizado que seguiu o Padrão da Comunidade sobre nudez e atividade sexual de adultos. Após o Comitê selecionar este caso, o Facebook determinou ser um erro e restaurou a publicação.	Reverte a decisão da Meta de remover, mantendo o conteúdo na rede.
Citação nazista	Indivíduos e organizações perigosas	Usuário fez publicação afirmando que o ex-presidente dos Estados Unidos segue modelo fascista ao citar o líder nazista, Joseph Goebbels. Comitê concluiu que a citação não apoiava a ideologia do partido nazista ou os atos de ódio e violência do regime.	Reverte a decisão da Meta de remover, mantendo o conteúdo na rede.

Alegação sobre cura da COVID	Violência e incitação	O usuário criticou a falta de uma estratégia de saúde na França e incitou o uso da hidroxicloroquina, um “medicamento inofensivo”. Meta removeu alegando a desinformação, mas o Comitê enfatizou que a publicação não havia demonstrado que aumentaria o nível de danos iminentes e a decisão da Meta é limitadora à liberdade de expressão.	Reverte a decisão da Meta de remover, mantendo o conteúdo na rede.
Protesto na Índia contra a França	Violência e incitação	Usuário publicou meme com o texto incitando a violência ao franceses, “se o “kafir” (infiel) falar algo contra o Profeta, a espada deverá ser retirada da bainha”. Meta removeu dada a ameaça mesmo que velada, a maioria dos membros do Comitê concluiu que essa publicação não causaria grandes danos.	Reverte a decisão da Meta de remover, mantendo o conteúdo na rede.
Representação de Zwarte Piet	Discurso de ódio	O usuário compartilhou um vídeo que mostrava uma criança encontrando três adultos, dois retratando “Zwarte Piet”, com os rostos pintados de preto e usando perucas afro. O Comitê manteve a decisão da Meta de remover o conteúdo que violava a proibição expressa de publicar caricaturas na forma de blackface.	Mantém a decisão da Meta de remover, retirando o conteúdo da rede.
Preocupação de Punjab com a RSS na Índia	Indivíduos e organizações perigosas	O usuário alegou que os regimentos sikhs no exército informaram ao Primeiro-ministro Modi que estão dispostos a morrer para proteger os fazendeiros sikhs e suas terras em Punjab. Análise humana da Meta determinou que a publicação violava o Padrão da Comunidade e a removeu, entretanto para o Comitê essa decisão não consistente com a liberdade de expressão defendida pela empresa.	Reverte a decisão da Meta de remover, mantendo o conteúdo na rede.
Suspensão do ex-presidente Trump	Indivíduos e organizações perigosas	Ex-presidente fez publicações apoiando a violência e a invasão do capitólio. O Comitê considerou que as duas publicações do Sr. Trump, em 6 de janeiro, violaram gravemente os Padrões da Comunidade do Facebook e as Diretrizes da Comunidade do Instagram que proíbem exaltar ou apoiar pessoas envolvidas em violência.	Mantém a decisão da Meta de remover, retirando o conteúdo da rede.
Meme “dois botões”	Cruel e insensível	Usuário fez 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”. A minoria do Comitê concluiu que não ficou suficientemente claro se o usuário compartilhou o conteúdo para criticar o governo turco.	Reverte a decisão da Meta de remover, mantendo o conteúdo na rede.
Protestos a favor de Navalny na Rússia	Bullying e assédio	O Comitê de Supervisão revogou a decisão do Facebook de remover um comentário no qual um apoiador de Alexei Navalny, líder preso da oposição russa, chamou outro usuário de “robô covarde”. O Facebook removeu o comentário porque este continha a palavra “covarde”, que foi interpretada como uma alegação de caráter negativo. O Comitê concluiu que a decisão criava uma restrição desnecessária e desproporcional à liberdade de expressão.	Reverte a decisão da Meta de remover, mantendo o conteúdo na rede.
Isolamento de Öcalan	Indivíduos e organizações perigosas	O usuário escreveu que aquele era o momento de falar sobre o fim do isolamento de Öcalan na prisão na Ilha Imrali, na Turquia, incentivando os leitores a falar sobre a prisão e a natureza desumana do confinamento em solitária de Öcalan. O Comitê concluiu que a decisão original do Facebook de remover o conteúdo não foi consistente com os Padrões da Comunidade da empresa.	Reverte a decisão da Meta de remover, mantendo o conteúdo na rede.

Bot de Myanmar	Discurso de ódio	O post discutiu maneiras de limitar o financiamento aos militares de Myanmar. “O povo de Hong Kong, porque as porras dos chineses torturaram eles, mudaram seus bancos para o Reino Unido e agora eles [os chineses] não podem mais encostar neles”. A Meta removeu a publicação de acordo com os Padrões da Comunidade sobre discurso de ódio, devido ao linguajar adotado. O Comitê avaliou se esse conteúdo deve ser restaurado com base em três óticas: os Padrões da Comunidade, os valores da empresa e suas responsabilidades sobre os direitos humanos.	Reverte a decisão da Meta de remover, mantendo o conteúdo na rede.
Confinamentos devido à COVID no Brasil	Violência e incitação	O Comitê de Supervisão manteve a decisão do Facebook de deixar ativa uma publicação de um conselho médico de nível estadual no Brasil que afirmava que o confinamento é ineficaz e que foi condenado pela Organização Mundial da Saúde (OMS).	Mantém a decisão da Meta de deixar o conteúdo na rede.
Publicação compartilhada da Al Jazeera	Indivíduos e organizações perigosas	O usuário compartilhou publicação com a legenda “A liderança da resistência na área comum concede uma trégua para a ocupação retirar seus soldados da mesquita de Al-Aqsa e do bairro Sheikh Jarrah até às 18h. Quem avisa, amigo é. Abu Ubaida – Porta-voz militar das Brigadas Al-Qassam” Como resultado da escolha do caso pelo Comitê, o Facebook concluiu que havia removido o conteúdo por engano e o restaurou.	Reverte a decisão da Meta de remover, mantendo o conteúdo na rede.
Protestos na Colômbia	Discurso de ódio	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”. A Meta removeu o conteúdo devido ao linguajar. O Comitê manteve o conteúdo devido ao seu valor jornalístico, embora o uso do termo seja calunioso.	Reverte a decisão da Meta de remover, mantendo o conteúdo na rede.
Calúnias na África do Sul	Discurso de ódio	Um usuário postou em um grupo público discutindo o “multirracismo” 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, essa pessoa “não bate bem da cabeça”. Ele concluiu que a pessoa seria apenas um “escravo sofisticado”, “um preto esperto”, “n goeie kaffir” ou “a house ‘n***er’”. A Meta e o Comitê removeram o conteúdo por violar sua política que proíbe o uso de calúnias direcionadas a pessoas com base em raça, etnia e/ou nacionalidade.	Mantém a decisão da Meta de remover, retirando o conteúdo da rede.
Cinto Wampum	Discurso de ódio	Um artista indígena postou uma foto de um cinto de wampum (uma forma de arte em que conchas são tecidas para formar imagens, registrando histórias e acordos). O objeto feito por ele mesmo e tinha como objetivo aumentar a conscientização sobre o massacre de Kamloops. O título de seu trabalho era 'Kill the Indian/ Save the Man' ('Mate o índio/Salve o homem'). Como o Comitê selecionou o caso em apreço, a Meta identificou a remoção como um “erro de monitoramento” e restaurou o conteúdo.	Reverte a decisão da Meta de remover, mantendo o conteúdo na rede.

Cerveja de Ayahuasca	Produtos regulamentados	Usuário publicou vídeo da cerveja de ayahuasca com a legenda “AYAHUASCA É PARA QUEM TEM CORAGEM DE SE ENFRENTAR”. O Comitê revogou a decisão da meta de remoção da publicação e recomendou que a empresa altere suas regras para permitir que usuários discutam, de forma positiva, os usos tradicionais ou religiosos de drogas não medicinais.	Reverte a decisão da Meta de remover, mantendo o conteúdo na rede.
Alegação de crimes em Raya Kobo	Discurso de ódio	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 e outras cidades na região de Amhara, na Etiópia. Ainda assim, ele alegou que civis da etnia Tigray ajudaram o TPLF com essas atrocidades. Embora a Meta inicialmente tenha removido a publicação por violar os Padrões da Comunidade sobre discurso de ódio, a empresa restaurou o conteúdo depois que o Comitê selecionou o caso. O Comitê tomou sua decisão pela remoção ciente das tensões entre proteger a liberdade de expressão e reduzir a ameaça de conflito sectário.	Mantém a decisão da Meta de remover, retirando o conteúdo da rede.
Questionamento sobre Adderall®	Produtos regulamentados	Usuário fez publicação em grupo privado na qual se identifica como uma pessoa que tem TDAH e pergunta ao grupo como falar com um médico sobre medicamentos específicos. Ele declara que recebeu uma prescrição de Xanax, mas que o remédio Adderall funcionou para ele no passado. Como o Comitê selecionou o caso em apreço, a Meta identificou a remoção como um “erro de monitoramento” e restaurou o conteúdo.	Reverte a decisão da Meta de remover, mantendo o conteúdo na rede.
Jornalista sueca denuncia violência sexual contra menores	Nudez adulta e atividades sexuais	O usuário na Suécia publicou foto de um banco de imagens em que aparecia uma jovem sentada com a cabeça apoiada nas mãos, o que impossibilitava que o rosto dela fosse visto. A foto tem uma legenda em sueco que descreve incidentes de violência sexual contra duas menores de idade, argumentava que o sistema de justiça penal sueco é muito leniente e incentiva o crime. O Comitê concluiu que a publicação não violou a política dos Padrões da Comunidade contra exploração sexual, abuso e nudez infantil.	Reverte a decisão da Meta de remover, mantendo o conteúdo na rede.
Vídeo explícito do Sudão	Conteúdo violento e explícito	Vídeo mostra uma pessoa caída próxima a um carro com um grande ferimento na cabeça e um olho visivelmente para fora com legenda em árabe que pedia que as pessoas permanecessem juntas e não confiassem nos militares. Depois da apelação do usuário ao Comitê, no entanto, a Meta emitiu uma permissão para conteúdo de valor jornalístico, o que isenta a publicação da remoção.	Mantém a decisão da Meta de remover, retirando o conteúdo da rede.
Ressignificação de palavras em árabe	Discurso de ódio	Uma conta pública do Instagram descrita como um espaço para discutir narrativas queer 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, incluindo os termos “zamel”, “foufou” e “tante/tanta”. A Meta explicou que suas decisões iniciais de remover o conteúdo foram baseadas em análises das fotos que continham os termos acima. O Comitê está preocupado que os analistas podem não ter recursos suficientes em termos de capacidade ou treinamento para evitar o tipo de erro observado neste caso.	Reverte a decisão da Meta de remover, mantendo o conteúdo na rede.

Animação Knin	Discurso de ódio	Uma página do Facebook publicou uma versão editada da animação da Disney “O Flautista Encantado” com legenda: “O músico de Čavoglave e os ratos de Knin”. O vídeo retrata uma (“Knin”) invadida por ratos. O narrador descreve como os ratos decidiram viver em um “país de ratos puros” e começaram a assediar e perseguir as pessoas que moravam na cidade. Um gaiteiro de Čavoglave apareceu com sua “flauta mágica” e os ratos começaram a segui-lo para fora da cidade e desapareceram e “todos viveram felizes para sempre”. A letra da música homenageia Momčilo Đujić, um padre ortodoxo sérvio que foi líder das forças de resistência sérvias durante a Segunda Guerra Mundial. O Comitê de Supervisão anulou a decisão original da Meta de permitir uma publicação no Facebook que mostrava os sérvios como se fossem ratos. Apesar de a Meta ter removido a publicação por violar a política de discurso de ódio, cerca de 40 moderadores haviam decidido anteriormente que o conteúdo não violava essa política.	Reverte a decisão da Meta de remover, mantendo o conteúdo na rede.
Ilustração retratando policiais colombianos	Indivíduos e organizações perigosas	Usuário publicou ilustração parecida com o brasão oficial da Polícia Nacional da Colômbia, retratando três pessoas vestindo uniforme policial e segurando um cassete-te acima da cabeça. Elas parecem chutar e agredir outra pessoa, que está deitada no chão com a cabeça ensanguentada. O texto do brasão dizia: “Polícia Nacional – República da Colômbia – Cassetete e Chute”. O Comitê de Supervisão revogou a decisão original da Meta de remover a publicação, preocupado com o fato de que os bancos de dados que têm a capacidade de remover automaticamente imagens que violem as regras da Meta, amplificando o impacto de decisões incorretas.	Reverte a decisão da Meta de remover, mantendo o conteúdo na rede.
Menção do Talibã na divulgação de notícias	Indivíduos e organizações perigosas	Popular jornal baseado na Índia fez uma publicação anunciando que as escolas e faculdades femininas reabririam em março de 2022. A Meta determinou que a publicação violava a política de organizações e indivíduos perigosos que proíbe “elogios” a entidades consideradas “envolvidas em sérios danos no meio físico”, incluindo organizações terroristas. O Comitê de Supervisão anulou a decisão da Meta visto que não condiz com as responsabilidades da Meta em matéria de direitos humanos.	Reverte a decisão da Meta de remover, mantendo o conteúdo na rede.
Gabinete de Comunicações da Região do Tigré	Violência e incitação	Usuário publica post com as perdas sofridas pelas Forças de Defesa Nacional da Etiópia sob a liderança do primeiro-ministro Abiy Ahmed no conflito armado com a FLPT. A publicação encoraja o exército nacional a virar suas armas contra o grupo fascista de Abiy Ahmed para reparar o mal causado a pessoas pelo grupo. O Comitê de Supervisão manteve a decisão da Meta de remover uma publicação que continha ameaça de violência no conflito na Etiópia.	Mantém a decisão da Meta de remover, retirando o conteúdo da rede.
Poema russo	Discurso de ódio	O Comitê anulou a decisão original da Meta de remover uma publicação que comparava o exército russo na Ucrânia com os nazistas e citava um poema que pede o extermínio de fascistas, posto que considera improvável essa publicação aumentar a violência, ainda que existam complexidades adicionais na avaliação do discurso violento em situações de conflito internacional.	Reverte a decisão da Meta de remover, mantendo o conteúdo na rede.

Gênero musical drill do Reino Unido	Violência e incitação	A Polícia Metropolitana enviou um email para a Meta solicitando que a empresa analisasse todos os conteúdos que contivessem “Secrets Not Safe” sob o argumento que as letras de suas músicas incitasse a violência. A Meta decidiu pela revogação mesmo com provas suficientes para concluir que o conteúdo contivesse uma ameaça plausível, e a análise do Comitê não revelou nenhuma prova para corroborar tal conclusão, sendo assim, na ausência de provas, a Meta deveria ter dado mais peso à natureza artística do conteúdo.	Reverte a decisão da Meta de remover, mantendo o conteúdo na rede.
Vídeo após ataque a igreja na Nigéria	Conteúdo violento e explícito	Usuário do Instagram na Nigéria publicou um vídeo que mostrava corpos ensanguentados, resultado do ataque terrorista. A Meta removeu o conteúdo, entretanto o Comitê concluiu que a restauração da publicação com uma tela de aviso protege a privacidade das vítimas e permite a discussão de eventos que alguns estados podem querer suprimir, dando importância ao valor jornalístico.	Reverte a decisão da Meta de remover, mantendo o conteúdo na rede.
Vídeo de assédio sexual na Índia	Exploração sexual de adultos	Conta do Instagram que se descreve como uma plataforma que mostra a perspectiva dos dalits publicou um vídeo da Índia em que um grupo de homens agredia uma mulher. Não é possível ver o rosto da mulher no vídeo, e não há nudez. A Meta a removeu por violar a política de exploração sexual de adultos, que proíbe conteúdo que “representa, ameaça ou promove violência, agressão sexual ou exploração sexual”, entretanto restaurou o conteúdo e colocou uma tela de aviso que impede que menores de 18 anos vejam o vídeo, decisão compactuada com o Comitê.	Mantém a decisão da Meta de remover, retirando o conteúdo da rede.
Slogan de protesto no Irã	Violência e incitação	Usuário publicou conteúdo protestando contra o governo iraniano, que contém o slogan “marg bar... Khamenei”. Isso se traduz literalmente como “morte a Khamenei”, mas é frequentemente usado como retórica política para significar “abaixo Khamenei”. A análise do Comitê sobre as políticas de conteúdo foi fundamentada pelo compromisso com a voz da Meta, que a empresa descreve como “primordial”, e seus valores de segurança e dignidade.	Reverte a decisão da Meta de remover, mantendo o conteúdo na rede.
Identidade de gênero e nudez	Proposta de cunho sexual	Conta de um casal americano que se identifica como transgênero e não binário teve duas publicações removidas na qual estão com os seios expostos e os mamilos cobertos. As legendas da imagem discutem a saúde transgênero e dizem que um membro do casal em breve será submetido a uma cirurgia importante. A Meta decidiu sob o padrão que proíbe imagens que contenham mamilos femininos, exceto em circunstâncias específicas, como amamentação e cirurgia de afirmação de gênero. O Comitê considera que as políticas da Meta sobre nudez adulta resultam em mais barreiras à expressão de mulheres, trans e pessoas não binárias nas plataformas.	Reverte a decisão da Meta de remover, mantendo o conteúdo na rede.

Produtos farmacêuticos do Sri Lanka	Produtos regulamentados	A publicação em questão tinha um pedido do sindicato médico no Sri Lanka, solicitando doações de medicamentos e produtos médicos para o país e fornecendo um link para isso. A empresa notou que a publicação violava seu Padrão da Comunidade de Produtos e Serviços Rescritos, que proíbe conteúdo pedindo medicamentos, mas aplicou a tolerância à política. O Comitê de Supervisão confirmou a decisão.	Mantém a decisão da Meta de remover, retirando o conteúdo da rede.
Vídeo de prisioneiros de guerra armênios	Planejamento de dano e divulgação de crimes	O usuário publicou vídeo em uma página que se identifica como documentação de supostos crimes de guerra cometidos pelo Azerbaijão contra armênios no contexto do conflito de Nagorno-Karabakh. Na legenda, afirma que o vídeo retrata soldados azeris torturando prisioneiros de guerra armênios. O Comitê constatou que, embora o conteúdo viole os Padrões da Comunidade contra coordenação de danos e incitação ao crime, a Meta aplicou corretamente a permissão de conteúdo de valor jornalístico e deixou o conteúdo ativo.	Mantém a decisão da Meta de remover, retirando o conteúdo da rede.
Discurso de general brasileiro	Planejamento de dano e divulgação de crimes	Usuário do Facebook publicou um vídeo relacionado às eleições brasileiras de 2022. Na legenda em português, havia um apelo para "situar" o Congresso brasileiro como "a última alternativa". No dia 8 de janeiro, apoiadores do ex-presidente Jair Bolsonaro invadiram o Congresso Nacional, o Supremo Tribunal Federal e a Presidência da República, localizados na Praça dos Três Poderes, em Brasília, intimidando policiais e destruindo o patrimônio público. Como resultado da seleção do caso pelo Comitê, a Meta determinou que as várias decisões de manter o conteúdo no Facebook foram um equívoco. Em 20 de janeiro de 2023, depois que o Comitê selecionou esse caso, a Meta removeu o conteúdo.	Reverte a decisão da Meta de remover, mantendo o conteúdo na rede.
Discurso desumanizador contra uma mulher	Discurso de ódio	Um usuário recorreu da decisão da Meta de manter uma publicação no Facebook que atacava uma mulher identificável e a comparava a um veículo motorizado ("truck", ou caminhão). Depois que o Comitê trouxe a apelação à atenção da Meta, a empresa reverteu a decisão original e removeu a publicação.	Reverte a decisão da Meta de remover, mantendo o conteúdo na rede.
Declaração metafórica contra o presidente do Peru	Violência e incitação	Um usuário fez uma apelação em relação à decisão da Meta de remover uma publicação do Facebook que incluía uma declaração metafórica contra o então presidente do Peru, Pedro Castillo. A Meta inicialmente removeu a publicação do Facebook nos termos da política sobre violência e incitação. Dado o caráter metafórico da declaração e o contexto do processo de impeachment contra Pedro Castillo, a Meta concluiu que o usuário parece defender a "suspensão" do então presidente, e não agir com violência contra ele. Portanto, a remoção inicial estava incorreta, e a Meta restaurou o conteúdo no Facebook.	Reverte a decisão da Meta de remover, mantendo o conteúdo na rede.
Líder anticolonialista Amílcar Cabral	Indivíduos e organizações perigosas	Um usuário fez uma apelação à decisão da Meta de remover uma publicação do Facebook que tinha um poema sobre Amílcar Cabral, líder anticolonialista da Guiné-Bissau. Após o Comitê levar a apelação ao conhecimento da Meta, a empresa reverteu a decisão anterior e restaurou a publicação.	Reverte a decisão da Meta de remover, mantendo o conteúdo na rede.

Primeiro-ministro do Camboja	Planejamento de dano e divulgação de crimes	A página oficial do primeiro-ministro do Camboja realizou uma transmissão ao-vivo na qual o chefe de Estado, Hun Sen, pede que os adversários políticos responsáveis pelas alegações escolham entre o “legal system” (sistema jurídico) ou “a bat” (um cassete) e afirma ainda que, caso não escolham o sistema jurídico, ele vai convocar apoiadores do CPP para protestar e agredi-los. O Comitê conclui que o vídeo em questão nesse caso incluía afirmações indiscutíveis de intenção de cometer violência contra os adversários políticos de Hun Sen, o que claramente viola a política sobre violência e incitação.	Reverte a decisão da Meta de manter, removendo o conteúdo na rede.
------------------------------	---	---	--

Tendo em vista o levantamento, buscou-se analisar como esse inicial modelo de apelação para a moderação de conteúdo privado toma decisões e quais são as características em unicidade dos casos selecionados.

RESULTADOS E DISCUSSÃO

Diante da coleta e revisão sistemática das decisões mencionadas anteriormente, o presente tópico tem como objetivo (i) apresentar o padrão de estrutura observado em todas as decisões; (ii) identificar os temas mais recorrentes entre os casos e; (iii) analisar a direção dos posicionamentos do Comitê nas decisões coletadas.

Em primeiro momento, é relevante ressaltar que todas as decisões seguem uma estrutura semelhante, com exceção das decisões sumárias. Esta categoria distinta é relativamente recente, contando apenas com três deliberações publicadas até junho de 2023. A principal distinção reside no fato de que essas decisões sumárias dizem respeito a casos nos quais a Meta reverteu sua decisão original sobre um conteúdo após a intervenção do Comitê. Esse tipo de decisão visa fornecer transparência quanto às correções efetuadas pela Meta, embora não estabeleça precedentes. As decisões sumárias são mais concisas e consistem em um resumo do caso, descrição do caso e histórico, autoridade e alcance do Comitê, importância do caso e a decisão.

Por outro lado, as decisões vinculantes apresentam um corpo mais detalhado, começando com informações concisas, como resumo do caso, contexto do caso, principais conclusões e a decisão do Comitê de Supervisão. Posteriormente, na parte da versão completa das decisões, destacam-se a descrição do caso, o escopo e a autoridade do Comitê, as fontes de autoridade e diretrizes, contribuições dos usuários e da Meta, comentários públicos, análise do Comitê, decisão do Comitê, recomendações e informações processuais.

Neste contexto, ao considerar as fontes de autoridade e diretrizes, juntamente com a análise do Comitê, é possível notar que as deliberações são influenciadas por duas bases distintas: os padrões da comunidade e os valores internos da Meta (Community Guidelines and Standards), bem como os Tratados do Direito Internacional de Direitos Humanos. Em verdade, torna-se evidente que as políticas internas da Meta têm um peso menor quando comparadas à legislação internacional, de acordo com Barbosa (2023) que confirma que, no decorrer da análise dos casos, os padrões estabelecidos pela Meta foram empregados de forma preliminar. Os valores de voz, dignidade e segurança foram mencionados de forma sucinta, enquanto os padrões da comunidade foram predominantemente utilizados para classificar o grau de discurso de ódio. Como resultado, o Oversight Board rebaixou os padrões da Meta para uma posição secundária e os interpretou à luz dos Tratados Internacionais de Direitos Humanos, considerando-os como a máxima autoridade

Nessa perspectiva, após uma análise das decisões apresentadas frisa-se também que, em todos os casos vinculantes, o Oversight Board fundamenta suas decisões com base no Pacto Internacional dos 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). Nota-se também a ausência de legislação nacional ou de organismos regionais de direitos humanos.

Ademais, outro padrão observado em todas as decisões foi a referência ao direito à liberdade de expressão, conforme estabelecido no artigo 19 do PIDCP. Este fato por si só, indica a direção predominante das decisões do Comitê. Na verdade, em todas as deliberações, ficou evidente uma preocupação com a garantia da livre expressão de opiniões nas publicações revisadas.

Esse posicionamento também é caracterizado pela constante reafirmação do compromisso com o princípio da "Voz", um dos valores fundamentais da Meta, que se faz presente em todas as decisões. Veja-se:

Os valores da Meta são descritos na introdução aos Padrões da Comunidade do Facebook. O valor "Voz" é definido como "fundamental": o objetivo dos nossos Padrões da Comunidade sempre foi criar um local de expressão e dar voz às pessoas. (...) Queremos que as pessoas possam falar abertamente sobre assuntos importantes para elas, ainda que sejam temas que geram controvérsias e objeções. (...) A análise do Comitê foi fundamentada pelo compromisso com o valor

“Voz” da Meta, que a empresa descreve como “primordial”, e seus valores de “Segurança” e “Privacidade” e “Dignidade” (OVERSIGHT BOARD, 2023).

Conforme observado, a liberdade de expressão e a relevância da voz são argumentos frequentemente reiterados pelo Comitê. Como resultado dessa abordagem, dos 41 (quarenta e um) casos examinados, apenas 10 (dez) tiveram seus conteúdos removidos das plataformas da Meta, enquanto 31 (trinta e um) permaneceram ativos. Esses dados evidenciam a propensão do Oversight Board a manter os conteúdos ativos, mesmo em casos de alta controvérsia, com o intuito de preservar a liberdade de expressão nos espaços de interação online.

Alguns exemplos emblemáticos desse posicionamento incluem casos como o relacionado ao confinamento durante a pandemia de Covid no Brasil, e à recomendação de hidroxicloroquina, na França. Nestas situações, os usuários questionaram a eficácia do lockdown, mesmo diante das recomendações da OMS, e incentivaram o uso de medicamentos contraindicados, o que resultou em disseminação de desinformação. O Comitê optou por manter ambas publicações ativas, respeitando o princípio da liberdade de expressão.

Outro aspecto a ser observado é que as decisões do Comitê estabelecem precedentes, no entanto, ainda existem casos que apresentam problemas semelhantes aos já abordados pelo Comitê. Isso ocorre, por exemplo, nos casos envolvendo "Sintomas de câncer de mama e nudez" e "Identidade de gênero e nudez", nos quais os conteúdos foram removidos pela Meta devido à política de proibição de imagens contendo mamilos femininos, com exceção de circunstâncias específicas, como amamentação e cirurgia de afirmação de gênero. Em ambos os casos, a intenção era promover campanhas sobre temas da saúde, mas o Comitê considerou que as políticas da Meta em relação à nudez adulta resultam em obstáculos adicionais para a expressão de mulheres, pessoas trans e não binárias nas plataformas, revertendo as decisões da empresa de remover as publicações.

Em contínua análise, utilizando os critérios de classificação dos processos do Board, observou-se que a violação mais comum analisada foi a de discurso de ódio, totalizando 12 casos. É interessante notar que, em quase 80% desses casos, a questão está relacionada a assuntos políticos. Essa é a temática que mais resultou na remoção de conteúdo, com destaque para a suspensão do acesso do ex-presidente Donald

Trump ao Facebook e Instagram após ele exaltar e apoiar a invasão ao Capitólio e a remoção de conteúdos relacionados ao 8 de janeiro no Brasil.

A partir deste ponto, avança-se para a discussão final, considerando a conclusão dos objetivos delineados neste tópico. Foi observado um padrão recorrente nas decisões que tende a favorecer a liberdade de expressão por parte do Comitê de Supervisão.

O Oversight Board tem, de fato, desempenhado um papel eficaz na promoção da transparência das decisões, na abertura à participação pública nos comentários sobre os casos e na utilização de ferramentas do Constitucionalismo Digital para conferir autoridade e legitimidade às suas determinações. Isso tem dado origem a um modelo que se assemelha a uma forma de governança privada, embora tenha características que remetem ao domínio do direito público. Entre as características notáveis, incluem-se a natureza vinculativa das decisões para a Meta, o estabelecimento de precedentes, a preocupação em divulgar publicamente os motivos subjacentes às decisões para manter a transparência do Comitê, a referência a várias fontes jurídicas, a adoção de um estilo de argumentação com nuances judiciais e o envolvimento de uma variedade de partes interessadas impactadas pelas decisões do Oversight Board.

É digno de nota que a incorporação da norma internacional de direitos humanos na interpretação dos termos contratuais de um serviço prestado por uma multinacional de tecnologia é um desenvolvimento positivo. Dada a dimensão e relevância da empresa, como é o caso da Meta, esse aspecto assume uma relevância significativa para o domínio público.

Por outro lado, subsistem questões não plenamente compreendidas ou que demandam maior atenção. Essas incluem os limites de uma decisão após a regulamentação, especialmente nos casos envolvendo discurso de ódio, quando o dano já foi causado. Há também a necessidade de esclarecer a verdadeira independência entre o Oversight Board e a Meta, para determinar se a estrutura atual é genuinamente independente ou meramente uma fachada. Além disso, é pertinente indagar se o Comitê utiliza uma linguagem excessivamente jurídica como artifício para criar uma imagem de comprometimento com os direitos humanos e assuntos correlatos. Outro ponto relevante é o impacto do valor dos precedentes estabelecidos pelas decisões e recomendações nas estruturas e políticas da comunidade da Meta, que atualmente parece limitar-se aos casos específicos. Este estudo, predominantemente descritivo

e narrativo, serve como base para investigações futuras que possam abordar essas questões de forma mais aprofundada e abrangente. A intenção é colaborar e incentivar esse debate.

CONSIDERAÇÕES FINAIS

Em síntese, após uma análise detalhada da influência do Constitucionalismo Digital na origem do Oversight Board e analisar a estrutura do Conselho de Supervisão da Meta, podemos identificar tanto aspectos positivos quanto desafios inerentes ao atual modelo de moderação de conteúdo. O Comitê tem demonstrado um esforço significativo para promover a autoridade, acessibilidade e transparência na plataforma, oferecendo aos usuários a oportunidade de serem ouvidos e considerando, cuidadosamente, todos os contextos relevantes e os valores concorrentes em cada caso.

O modelo proposto pelo Oversight Board apresenta princípios procedimentais valiosos que podem servir como base para o desenvolvimento de um ambiente de moderação mais transparente e equitativo. Em contrapartida, há questões importantes permanecem em aberto, incluindo o impacto real das decisões do Comitê além dos casos específicos.

É fundamental reconhecer que o modelo atual enfrenta obstáculos significativos que precisam ser superados. O volume massivo de usuários e conteúdos presentes na plataforma dificulta o trabalho do Comitê em atender a todas as demandas e garantir uma moderação eficaz. Além disso, desafios como o overblocking, vieses algorítmicos, barreira da língua e falta de transparência continuam representando grandes desafios a serem enfrentados.

Para avançar, é imperativo buscar soluções capazes de equilibrar adequadamente a liberdade de expressão e a proteção dos direitos humanos. O desenvolvimento de mecanismos de governança mais robustos, com a ampla participação de especialistas e partes interessadas, pode contribuir para aprimorar a eficácia e a legitimidade do processo de moderação de conteúdo.

Ademais, é relevante considerar o Constitucionalismo Digital como uma abordagem promissora para lidar com os desafios da moderação de conteúdo. Esse enfoque pode oferecer uma base sólida para o desenvolvimento de políticas e diretrizes mais abrangentes, que levem em consideração os princípios fundamentais dos direitos humanos e as demandas da sociedade digital. No entanto, é essencial que

o setor privado utilize essas ferramentas com prudência, evitando viés de marketing e autoridade.

Diante disso, é de suma importância que haja um compromisso contínuo com a melhoria da governança e da responsabilidade na esfera digital. Somente por meio de esforços conjuntos, envolvendo tanto as plataformas quanto a sociedade em geral, poderemos enfrentar de maneira eficaz os desafios complexos e em constante evolução da moderação de conteúdo.

Portanto, instamos a Meta e outras plataformas a continuarem investindo na melhoria de seus processos de moderação, considerando os feedbacks dos usuários e colaborando com especialistas e organizações relevantes. Somente dessa forma poderemos construir uma internet mais inclusiva, segura e responsável, onde a liberdade de expressão seja garantida sem comprometer a dignidade e os direitos de todos os usuários.

Em última análise, esta discussão ressalta a necessidade contínua de aprimorar a governança e a responsabilidade na esfera digital. A colaboração entre as plataformas, especialistas e organizações relevantes é fundamental para enfrentar os desafios complexos da moderação de conteúdo e construir um ambiente online que promova a liberdade de expressão, sem comprometer a dignidade e os direitos de todos os usuários. A evolução do Constitucionalismo Digital e o papel das entidades de autorregulação, como o Oversight Board, representam passos importantes nessa jornada em constante evolução. Este estudo serve como um ponto de partida para uma discussão contínua e aprofundada sobre a moderação de conteúdo nas plataformas digitais e o Constitucionalismo Digital.

REFERÊNCIAS

BARBOSA, Laíse Milena. A regulação do discurso de ódio nas redes sociais: o caso do oversight board da Meta Curitiba, 2023. (Trabalho de dissertação do Programa de Pós-graduação em Direito, Setor de Ciências Jurídicas - UFPR).

BIETTI, Elettra. From ethics washing to ethics bashing: a view on tech ethics from within moral philosophy. In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT* '20). Association for Computing Machinery, New York, NY, USA, 210–219, 2020.. <https://doi.org/10.1145/3351095.3372860>

CELESTE, Edoardo. Digital constitutionalism: a new systematic theorisation. *International Review of Law, Computers and Technology*, v. 33, n. 1. 2019.

COWLS, Josh et al. Constitutional metaphors: Facebook's "supreme court" and the legitimization of platform governance. *New Media & Society*, v. 00, n. 0, p. 1-25, 2022.

DOUEK, Evelyn. Facebook's "Oversight Board:" Move Fast with Stable Infrastructure and Humility. *North Carolina Journal of Law & Technology*, V. 21, Issue 1. October, 2019. Disponível em: <<https://ncjolt.org/articles/6436-2/>> . Acesso em 07 de março de 2023.

GILL, Lex; REDEKER, Dennis; GASSER, Urs. Towards Digital Constitutionalism? Mapping Attempts to Craft an Internet Bill of Rights. *Research Publication*, No. 2015-15 November 9, 2015, v. 7641, 2015.

GILLESPIE, Tarleton. Content moderation, AI, and the question of scale. *Big Data & Society*. Washington, DC, 2020. Disponível em: <<https://journals.sagepub.com/doi/pdf/10.1177/2053951720943234>>. Acesso em: 05/03/2023.

GORWA, Robert; BINNS, Reuben; KATZENBACH, Christian. Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society*. Washington, DC, 2020. Disponível em: <<https://journals.sagepub.com/doi/pdf/10.1177/2053951719897945>>. Acesso em: 05/03/2023.

GRIFFIN, Rachel,. Metaphors matter: why we shouldn't call the Facebook Oversight Board a court. 20 Years Hertie School, Berlim, 11 de fevereiro de 2021. Disponível em: <<https://www.hertie-school.org/en/news/detail/content/metaphors-matter-why-we-shouldnt-call-the-facebook-oversight-board-a-court>>. Acesso em: 05/03/2023.

HATZIPANAGOS, Rachel. How online hate turns into real-life violence. *Washington Post*, 2018. Disponível em: <https://www.google.com/url?q=https://www.washingtonpost.com/nation/2018/11/30/how-online-hate-speech-is-fueling-real-life-violence/&sa=D&source=docs&ust=1678570494337568&usg=AOvVaw3Fna8sfNrVIGXII78Ou_LO>. Acesso em: 08/03/2023.

KLONICK, K. The Facebook Oversight Board: Creating an Independent Institution to Adjudicate Online Free Expression. *The Yale Law Journal*, v. 129, p. 2418-2499, 2020.

LLANSÓ, Emma J. No amount of "AI" in content moderation will solve filtering's prior restraint problem. *Big Data & Society*. Washington, DC, 2020. Disponível em: <<https://journals.sagepub.com/doi/pdf/10.1177/2053951720920686>>. Acesso em: 05/03/2023.

LAVILLE, Christian A construção do saber: manual de metodologia da pesquisa em ciências humanas / Christian Laville e Jean Dionne; tradução: Heloísa Monteiro e Francisco Settineri. — Porto Alegre : Artmed; Belo Horizonte: Editora UFMG, 1999.

MENDES, Gilmar Ferreira; OLIVEIRA FERNANDES, Victor. Constitucionalismo digital e jurisdição constitucional: uma agenda de pesquisa para o caso brasileiro. *Revista Brasileira de Direito*, Passo Fundo, v. 16, n. 1, p. 1-33, out. 2020. ISSN 2238-0604. Disponível em: <<https://seer.atitus.edu.br/index.php/revistadedireito/article/view/4103>>. Acesso em: 09 abr. 2023. doi:<https://doi.org/10.18256/2238-0604.2020.v16i1.4103> .

META. Atualizações trimestrais da Meta sobre o Comitê de Supervisão. Disponível em: <<https://transparency.fb.com/pt-br/oversight/meta-quarterly-updates-on-theoversight-board>>. Acesso em: 08 mar. 2023.

META. Nossa abordagem para conteúdo interessante. Disponível em: <<https://transparency.fb.com/pt-br/features/approach-to-newsworthycontent/>>. Acesso em: 08 mar. 2023.

OVERSIGHT BOARD. Conheça o Comitê. Disponível em: <<https://oversightboard.com/meet-the-board/>>. Acesso em: 08 mar. 2023.

OVERSIGHT BOARD. Decisões. Disponível em: <<https://www.oversightboard.com/decision/>>. Acesso em 20 jun. 2023.

OVERSIGHT BOARD. Rulebook for Case Review and Policy Guidance. Disponível em: <<https://oversightboard.com/sr/rulebook-for-case-review-andpolicy-guidance>>. Acesso em: 08 mar. 2023.

OVERSIGHT BOARD. Governança. Disponível em: <<https://www.oversightboard.com/governance/>>. Acesso em: 08 mar. 2023.

OVERSIGHT BOARD. Oversight Board Bylaws. Disponível em: <<https://www.oversightboard.com/sr/governance/bylaws>>. Acesso em: 08 mar. 2023.

PETTRACHIN, Andrea. Towards a universal declaration on internet rights and freedoms. The International Communication Gazette, v. 80, n. 4, 2018. Disponível em: <<https://journals.sagepub.com/doi/abs/10.1177/1748048518757139>>. Acesso: 08/03/2023.

SULLIVAN, Mark. Exclusive: The Harvard professor behind Facebook's oversight board defends its role. Fast Company. 2019. Disponível em: <<https://www.fastcompany.com/90373102/exclusive-the-harvard-professor-behind-facebooks-oversight-board-defends-its-role>>. Acesso em: 05 de maio de 2023.

ZUCKERBERG, Mark. A Blueprint for Content Governance and Enforcement. Facebook. November of 2018. Disponível em: < https://www.facebook.com/notes/751449002072082/?hc_location=ufi >. Acesso em: 07 de março de 2023.