

MODERAÇÃO DE CONTEÚDOS POLÍTICOS: LIBERDADE DE EXPRESSÃO HIGH-PROFILE¹

Camilla Tavares de Albuquerque 

Universidade Presbiteriana Mackenzie, São Paulo, Brasil 

Fernando Rister de Sousa Lima 

Universidade Presbiteriana Mackenzie, São Paulo, Brasil 

Contextualização: Se a questão da moderação de conteúdos nas plataformas já levanta discussões tormentosas, a celeuma é ainda maior quando se trata de assuntos políticos. O tema especialmente sensível em anos eleitorais, mas detém um impacto social a qualquer tempo, especialmente quanto às suas possíveis repercussões no jogo democrático. Entre erros e acertos, as empresas envolvidas nesse desafio têm criado estratégias específicas para lidar com a matéria, como é o caso da Meta, ao criar o programa de verificação cruzada (cross-check) para os perfis denominados high-profiles. A partir desse cenário, pergunta-se: Qual ou quais os critérios aplicados pela Meta à revisão de conteúdos políticos?

Objetivo: O objetivo deste trabalho é demonstrar como o arranjo normativo da Meta de moderação de conteúdos publicados por políticos tem sido construído, especialmente a partir do caso Trump, que deu ensejo à mais complexa das recomendações exaradas pelo Comitê de Supervisão criado pela empresa.

Método: Metodologicamente, adota-se a linha sociojurídica, analisando a moderação de conteúdos sob uma perspectiva mais ampla de um observador externo. Utilizando-se predominantemente do raciocínio indutivo, a técnica consistiu no levantamento e estudo documental e bibliográfico, com especial enfoque para aqueles oriundos do Comitê de Supervisão da Meta.

Resultados: A título conclusivo, foi possível verificar com clareza as divergências e convergências nos critérios adotados pela Meta e nas próprias decisões do Comitê, relevando eventuais riscos de uma aplicação ad hoc do direito à liberdade de expressão, consubstanciando o que seria possível chamar “liberdade de expressão high-profile”.

Palavras-chave: Moderação de conteúdos; Eleições; Meta; Programa de verificação cruzada; Liberdade de expressão.

¹ O artigo é produto do projeto de pesquisa interinstitucional: “Tribunais, Cidadania e Acesso às Atividades Econômicas na Sociedade Digital: um exame sociojurídico das demandas econômicas das plataformas em juízo” com financiamento do CNPQ, mediante chamada universal, e da CAPES, por meio do PAEP.

POLITICAL CONTENTS MODERATION: HIGH-PROFILE FREEDOM OF EXPRESSION

Contextualization: If the issue of content moderation on platforms already raises stormy discussions, the problem is even greater when it comes to political issues. The topic is especially sensitive in election years, but it has a social impact at any time, especially regarding its possible repercussions on the democratic game. Between mistakes and successes, companies involved in this challenge have created specific strategies to deal with the matter, as is the case with Meta, by creating the cross-check program for the called high-profiles. Considering this scenario, the question arises: What criteria or criteria does Meta apply to the review of political content?

Objective: The objective of this study is to demonstrate how the normative arrangement of the Meta for moderation of content published by politicians has been constructed, especially following the Trump case, which gave rise to the most complex of the recommendations made by the Supervisory Committee created by the company.

Method: Methodologically, it adopts the socio-legal line, analyzing content moderation from a broader perspective of an external observer. Predominantly using inductive reasoning, the technique consisted of surveying and studying documents and bibliography, with a special focus on those coming from the Meta's Oversight Board.

Results: Ultimately, it was possible to clearly verify the divergences and convergences in the criteria adopted by the Meta and in the Committee's own decisions, highlighting possible risks of an ad hoc application of the right to freedom of expression, substantiating what could be called "high-profile freedom of expression".

Keywords: Content Moderation; Elections; Meta; Cross-check Program; Freedom of Expression.

MODERACIÓN DE CONTENIDOS POLÍTICOS: LIBERTAD DE EXPRESIÓN HIGH-PROFILE

Contextualización: Si la cuestión de la moderación de contenidos en las plataformas ya suscita acalorados debates, el problema es aún mayor cuando se trata de cuestiones políticas. El tema es especialmente sensible en años electorales, pero tiene un impacto social en cualquier momento, especialmente en lo que respecta a sus posibles repercusiones en el juego democrático. Entre errores y aciertos, las empresas involucradas en este desafío han creado estrategias específicas para afrontar el asunto, como es el caso de la Meta, mediante la creación del programa de verificación cruzada (cross-check) denominados high-profiles. De este escenario surge la pregunta: ¿Qué criterios o criterios aplica Meta para la revisión de contenidos políticos?

Objetivo: El objetivo de este trabajo es demostrar cómo se ha construido el ordenamiento normativo de la Meta para la moderación de contenidos publicados por políticos, especialmente a partir del caso Trump, que dio lugar a la más compleja de las recomendaciones formuladas por el Comité de Supervisión creado por la compañía.

Método: Metodológicamente se adopta la línea sociojurídica, analizando la moderación de contenidos desde una perspectiva más amplia de un observador externo. Utilizando predominantemente el razonamiento inductivo, la técnica consistió en el levantamiento y estudio de documentos y bibliografía, con especial atención a los provenientes del Comité de Supervisión de Objetivos.

Resultados: Concluyentemente, fue posible verificar claramente las divergencias y convergencias en los criterios adoptados por la Meta y en las propias decisiones del Comité, destacando posibles riesgos de una aplicación ad hoc del derecho a la libertad de expresión, fundamentando lo que podría llamarse "libertad de expresión high-profile".

Palabras clave: Moderación de contenidos; Elecciones; Meta; Programa de Verificación Cruzada; Libertad de Expresión.

INTRODUÇÃO

Às vésperas de novas eleições, a questão da moderação de conteúdos pelas plataformas digitais permanece na ordem do dia. Se nas mais usuais interações sociais a moderação de conteúdos pode se mostrar controversa, no que tange ao conteúdo político o desafio é ainda maior. Desde o caso Cambridge-Analytica e, especialmente a partir das eleições de 2018, percebe-se uma crescente preocupação das plataformas quando do trato de conteúdos políticos.

O tema, no entanto, não se cinge ao período eleitoral. Embora tais momentos sejam permeados por uma potencialidade danosa maior, não há que se menosprezar o peso de conteúdos publicados por agentes políticos mesmo em períodos alheios às eleições. Não por acaso, o Facebook (ou a empresa Meta) esforçou-se em implantar critérios diversos de moderação de conteúdos que pudessem ser mais adequados não só a agentes políticos, mas a pessoas ilustres de um modo geral: os chamados high-profiles. Aqui nos interessam, especialmente, os casos de agentes políticos, cujas publicações podem gerar não só uma repercussão midiática considerável, como também consequências diversas para o jogo democrático, para além do mero escrutínio geral.

É esse cenário até então explanado que justifica a pesquisa. A importância da moderação de conteúdos é inegável e ganha espaço privilegiado no debate na esfera pública, especialmente quando se refere a conteúdos políticos. O tema demanda atenção e constante alerta ante as atitudes das plataformas, sendo de se ressaltar a importância a nível global dos posicionamentos adotados pelo Comitê de Supervisão da Meta (no original, Oversight Board) quando da regulação de postagens de políticos. Não por outro motivo, a atuação do Comitê tem sido objeto de discussões sob os mais diversos enfoques na comunidade científica, suscitando novos questionamentos acerca dos limites da liberdade de expressão, especialmente quando analisada sob o enfoque de autoridades em posição privilegiada no meio social.

A questão central é que o papel da plataforma não pode ser o de calar discursos divergentes, indispensáveis para a dinâmica democrática, nem o de permitir que os políticos utilizem de sua influência para finalidades escusas. Assim, a compreensão dos limites do direito à liberdade de expressão frente a outros valores sociais é matéria de interesse global, mostrando-se cada vez mais relevante no discurso jurídico-político. Considerando estas circunstâncias, pretende-se responder à seguinte pergunta: Qual ou quais os critérios aplicados pela Meta à revisão de conteúdos políticos?

O objetivo do estudo é demonstrar como o arranjo normativo da Meta de moderação de conteúdos publicados por políticos tem sido construído, especialmente a partir do caso Trump (2021-001-FB-FBR), o que justifica a sua escolha paradigmática.

Especificamente, o objetivo consiste em analisar quais alterações têm sido promovidas pela Meta em sua política de revisão, tendo por base as diretrizes fixadas na recomendação exarada pelo Conselho de Supervisão em relação ao tema (PAO-2021-02).

Metodologicamente, adota-se a linha sociojurídica, analisando a moderação de conteúdos sob uma perspectiva mais ampla de um observador externo. Utiliza-se o raciocínio indutivo, sem crer, no entanto, que este possa levar a uma constante de generalizações. A técnica de pesquisa consiste, quase que exclusivamente, em levantamento documental e bibliográfico, dando-se especial atenção ao conteúdo oriundo do Comitê de Supervisão.

A pesquisa se dá em três frentes de trabalho. Primeiramente, é brevemente analisado o programa de dupla verificação adotado pela plataforma, especificando as particularidades quando do trato de conteúdos divulgados por agentes políticos. Após, tem-se a análise da decisão do caso Trump, bem como das recomendações oriundas do Comitê de Supervisão para a moderação de tais conteúdos.

O esforço final consiste em analisar eventuais critérios empregados na moderação de conteúdos políticos – sejam eles oriundos de high-profiles ou não –, utilizando-se, para tanto, comparativamente, o caso brasileiro do Discurso de General (2003-001-FB-UA). A título conclusivo, é possível, ao menos, apontar divergências e convergências nos critérios adotados pela Meta e nas próprias decisões do Comitê, relevando eventuais riscos de uma aplicação ad hoc do direito à liberdade de expressão, consubstanciando o que seria possível chamar “liberdade de expressão high-profile”.

Em que pese se reconheça que o arranjo normativo da plataforma relativo ao tema esteja em constante transformação, um estudo inicial deste poderá esclarecer o caminho a ser percorrido a partir de então. Se a velocidade do direito é diversa daquela do mundo tecnológico, espera-se que tais considerações, mesmo incipientes, possam auxiliar no desenvolvimento de um prognóstico jurídico – ainda que precário – sobre a questão.

1. PROGRAMA CROSS-CHECK E REVISÃO HIGH-PROFILE

Já em 2019, o *Facebook* idealizou um órgão de controle interno especificamente para tratar do tema da moderação de conteúdos. O *Oversight Board*, hoje traduzido como Comitê de Supervisão (doravante CS) detém uma estrutura que em muito se assemelha à de um tribunal, reunindo uma comissão de juízes “leigos” para decidir sobre a remoção ou manutenção de conteúdos em casos paradigmáticos. Trata-se de uma organização altamente sofisticada e que consubstancia um dos mais complexos processos de moderação de conteúdos realizado pelas plataformas na atualidade e cuja atuação hoje abrange todas as plataformas integrantes da *Meta*.

O Comitê passou a funcionar em 2020 e, já em janeiro do ano seguinte, deparou-se com a contenda referente à suspensão da conta do presidente Trump (caso 2021-001-FB-FBR). Tamanha a repercussão política do caso que, no mesmo ano, o Comitê de Supervisão emitiu a mais completa de suas recomendações (PAO-2021-02), na qual analisa o programa de dupla verificação de conteúdos da plataforma (*cross-check program*) e fez 32 apontamentos especialmente direcionados ao tratamento de seus “usuários mais poderosos” (do original, “most powerful users”²).

O núcleo central das recomendações pertinentes ao caso se concentrou na necessidade de aprimoramento dos mecanismos de *compliance* da plataforma, dando especial enfoque à transparência das políticas aplicadas a tais personalidades. No mais, o CS faz expressa menção ao caso Trump, a partir do qual releva ter ficado mais atento acerca das políticas de *cross-check*. Foi a partir deste que o paradigma revisional referente aos agentes políticos passou a ser repensado, como declara o próprio CS, à luz do compromisso da *Meta* com os direitos humanos e com seus valores declarados³.

A questão central pautou-se no tipo de moderação de conteúdos aplicada a pessoas pública, evidenciando a existência de dois métodos revisionais distintos, um deles baseado na pessoa que posta o conteúdo e outro baseado no conteúdo em si, conforme se verá adiante. Antes, no entanto, cumpre esclarecer brevemente o funcionamento do processo de moderação de conteúdos da plataforma.

De acordo com os Padrões da Comunidade, a moderação de conteúdos passa por algumas etapas, sendo a primeira a moderação ordinária, que pode ser realizada por pessoas ou automaticamente, por meio de algoritmos desenhados a fim de fazer cumprir as políticas de conteúdo da Meta. A verificação cruzada é aplicável tanto à moderação de conteúdos como à remoção de páginas e perfis e, de acordo com a empresa:

[...] fornece níveis adicionais de revisão para determinados conteúdos que os nossos sistemas internos assinalam como estando em infração (através da automatização ou revisão humana), com o objetivo de prevenir ou minimizar os erros mais propensos de moderação de falsos positivos que possam ocorrer devido a vários fatores, como a necessidade de compreender o contexto ou os detalhes.

A grande maioria dos casos de moderação finaliza neste estágio e eventuais relevâncias são detectadas proativamente⁴, sem necessidade de escalamento revisional.

² META'S cross-check program. **Oversight Board**. Disponível em: <https://www.oversightboard.com/decision/PAO-NR730OF/> Acesso em 29 maio 2023.

³ POLICY advisory opinion on Meta's cross-check program. **Oversight Board**. 2021, s.p. Disponível em: <https://transparency.meta.com/pt-br/pao-cross-check-policy/> Acesso em 29 maio 2023.

⁴ ANÁLISE precisa de conteúdo de alto impacto pelo nosso sistema de verificação cruzada. **Meta Transparency Center**. 2023, s.p. Disponível em: <https://transparency.fb.com/pt->

Apenas após tal etapa é que os conteúdos podem, a depender do caso, ser direcionados ao chamado *cross-check*.

Havendo dúvida acerca da necessidade de remoção do conteúdo, a própria plataforma dá início ao *cross-check*, mantendo o conteúdo na rede, disponível para acesso – mesmo após a moderação inicial – até que a etapa revisional se encerre. Os objetos de revisão, em tais casos, são os chamados “falso positivos”, ou seja, conteúdos que, embora tenham sido apontados inicialmente como violadores dos padrões da comunidade, ostentam um certo grau de complexidade que exigiria uma observação mais apurada, antes mesmo do seu eventual bloqueio.

Se, durante a verificação cruzada um revisor confirmar que os conteúdos desrespeitam os nossos Padrões da Comunidade, aplicamos essas políticas e lidamos com os conteúdos em infração ou em conformidade. Consoante a complexidade dos conteúdos, podemos aplicar vários níveis de revisão, incluindo, em casos raros, a revisão por liderança. Se o revisor final determinar que os conteúdos em questão não desrespeitam os nossos Padrões da Comunidade, o revisor pode “anular” a ação inicial e deixar os conteúdos na plataforma.⁵

Apurada a complexidade do conteúdo, será ele objeto de uma moderação mais acurada pela plataforma, por meio de instâncias revisionais distintas. O escalonamento revisional se dá sob duas formas típicas, quais sejam, a Revisão Secundária Geral (*General Secondary Review – GSR*) e a Revisão Secundária da Resposta Prévia (*Early Response Secondary Review – ERSR*). A primeira representa a moderação aplicada ao público em geral; já a segunda é aplicada aos perfis de alto impacto, notadamente o de pessoas políticas, tal qual Donald Trump, então ex-presidente dos EUA.

No caso da revisão geral (GSR) o critério para a revisão é a natureza do conteúdo e, diferentemente do que se dá na revisão secundária (ERSR), não há qualquer garantia de que tais conteúdos sejam alvo de revisão humana. Inclusive, após analisar o processo, o Comitê de Supervisão deixou claro que a Meta não teria investido “os recursos necessários para que todos os conteúdos identificados por meio da GSR recebam a revisão humana adicional”⁶. Esta, portanto, é a revisão aplicada à grande maioria dos conteúdos postados na plataforma.

A ERSR, por seu turno, é aplicada taxativamente, conforme a autoria do conteúdo,

[pt/enforcement/detecting-violations/reviewing-high-visibility-content-accurately/](https://www.metatransparencycenter.org/enforcement/detecting-violations/reviewing-high-visibility-content-accurately/) Acesso em 13 maio 2023.

⁵ ANÁLISE precisa de conteúdo de alto impacto pelo nosso sistema de verificação cruzada. **Meta Transparency Center**. 2023, s.p.

⁶ No original: “Meta has not invested the resources needed for all content identified through GSR to receive additional human review” (POLICY advisory opinion on Meta’s cross-check program. **Oversight Board**. p. 10.)

obedecendo a uma lista⁷ que inclui celebridades, jornalistas em zonas de conflito, partidos políticos, agentes políticos e parceiros de negócios da Meta em geral. Em tais casos, embora o conteúdo seja mantido nas redes até a etapa revisional – como ocorre com o GSR –, o segundo nível de moderação é obrigatoriamente realizado por pessoas, no que contrasta com a revisão aplicada ao público em geral. Trata-se da chamada verificação cruzada ou *cross-check*, uma técnica que tenta dar conta do desafio de moderação de conteúdos da Meta, especialmente no intuito de evitar remoções incorretas.

Em qualquer caso, a potencial vantagem ou desvantagem do *cross-check* se consubstanciam no fato de que o conteúdo é mantido na rede a título de falso positivo. Se, de um lado, a manutenção evita eventual acusação de censura, por outro, a depender da natureza do conteúdo, pode levar a consequências irreversíveis ou, de fato, danos irreparáveis. Ou seja, de acordo com as políticas de moderação atuais, os conteúdos postados por agentes políticos, ainda que suspeitos, são mantidos na rede, passando por uma revisão humana específica, após a qual poderão, então sim, ser removidos.

Embora esse tipo de tratamento privilegiado já fosse alvo de controvérsia antes mesmo do caso Trump, o qual será abordado adiante. Já em 2019, Thomas E. Kadri e Kate Klonick, por exemplo, além de criticarem a falta de transparência do processo, fizeram uma digressão histórica para compreender o tratamento diferenciado reservado a pessoas públicas, sem deixar de lado a possibilidade de imposições mais gravosas⁸. No entanto, foi a partir do caso Trump que a atuação da Meta passou a ser questionada pelo Comitê, especialmente a partir das recomendações públicas por este formuladas.

O caso Trump serviu, portanto, como uma espécie de divisor de águas, exigindo que a Meta empregasse maior transparência nos seus processos revisionais, a partir do que ficou evidente a necessidade de alterações em suas políticas. De fato, o próprio CS admitiu falhas nas respostas da plataforma em face de seus questionamentos acerca da política de *cross-check*, já que, em certas ocasiões, não forneceu informações ou forneceu-as de forma incompleta⁹.

Se foi partir do caso da invasão do Capitólio, as políticas de ERSR passaram a ser alvo de importante discussão no bojo do Comitê de Supervisão, o estudo não pode prescindir

⁷ De fato, não se trata de mera força de expressão; a plataforma mantém uma lista de *high-profiles*, os quais serão objeto de Revisão Secundária. Em 2021, *The Wall Street Journal* noticiou que a “safe list” já contava com mais de 5.8 milhões de perfis (HORWITZ, Jeff. Facebook Says Its Rules Apply to All. Company Documents Reveal a Secret Elite That’s Exempt. **The Wall Street Journal**. New York City. 13 set. 2021. Disponível em: <https://www.wsj.com/articles/facebook-files-xcheck-zuckerberg-elite-rules-11631541353> Acesso em 29 maio 2023).

⁸ KADRI, Thomas E.; KLONICK, Kate. Facebook v. Sullivan: Public Figures and Newsworthiness in Online Speech. **Southern California Law Review**. v. 93, n. 19-0020, p. 72, 2019.

⁹ POLICY advisory opinion on Meta's cross-check program. **Oversight Board**. 2021, p. 3.

da uma prévia análise do evento, ou seja, do caso Trump, do qual se tratará adiante.

2. O CASO TRUMP

O *case* refere-se aos acontecimentos que a mídia convencionou chamar de “invasão do Capitólio”. O evento teve início em 6 de janeiro de 2021, durante a contagem dos votos das eleições presidenciais de 2020, quando uma multidão adentrou forçosamente o Capitólio em Washington, causando um tumulto de consequências graves, incluindo a morte de cinco pessoas e vários indivíduos feridos. A participação no evento do então presidente dos Estados Unidos, Donald Trump, deu-se por meio de publicações no *Facebook* e no *Instagram*, durante a invasão, em que o homem fez acusações reiteradas acerca de uma suposta fraude ocorrida no curso das votações eleitorais.

A primeira publicação do então presidente se deu às 16h21min (horário dos EUA) e foi removida pela plataforma às 17h41 (EUA) sob o argumento de violação dos Padrões da Comunidade sobre organizações e indivíduos perigosos. Nela, o presidente pedia “paz” e “ordem”, mas fazia expressa menção ao que chamou de “fraude ou roubo nas eleições”¹⁰.

Após a remoção do conteúdo, às 18h07min (EUA), enquanto o tumulto perdurava, o presidente fez uma nova postagem aludindo a “uma vitória esmagadora e sagrada” que teria sido “tomada de maneira vil e sem cerimônia de grandes patriotas que vêm sendo injustamente maltratados há tanto tempo”¹¹ e dizia aos norte-americanos: “vão para casa com amor e paz. Lembrem-se deste dia para sempre”¹². Mais uma vez, a plataforma removeu o conteúdo, porém de forma muito mais célere, apenas oito minutos após a postagem, sob a mesma justificativa anterior. Além disso, no mesmo ato o usuário foi bloqueado, sendo impedido de publicar no *Facebook* ou no *Instagram* pelo lapso de 24h, prazo que foi estendido indefinidamente, “pelo menos pelas próximas duas semanas até a conclusão da transferência pacífica do poder”¹³.

O caso Trump foi encaminhado ao CS após a posse do novo presidente dos EUA, Joe Biden, em 20 de janeiro. A própria plataforma pediu ao Comitê que julgasse a sua decisão de bloquear o ex-presidente por prazo indefinido, solicitando recomendações acerca das providências a serem tomadas em casos tais, envolvendo líderes políticos. Com efeito, apenas o fato de ter sido o caso encaminhado pelo próprio *Facebook* denota o grau de sensibilidade

¹⁰ FORMER President Trump's Suspension. **Oversight Board**, 2021, s.p. Disponível em: <https://www.oversightboard.com/decision/FB-691QAMHJ>. Acesso em 29 maio 2023.

¹¹ FORMER President Trump's Suspension. **Oversight Board**, 2021, s.p.

¹² FORMER President Trump's Suspension. **Oversight Board**, 2021, s.p.

¹³ FORMER President Trump's Suspension. **Oversight Board**, 2021, s.p.

e controvérsia acerca do conteúdo.

Registrado sob a nomenclatura 2021-001-FB-FBR, o caso foi alvo de intenso debate no Comitê o qual, por fim, exarou decisão favorável ao bloqueio da conta, porém não de forma indefinida:

[...] por manter uma narrativa sem fundamento de fraude eleitoral e chamadas para ação persistentes, o Sr. Trump criou um ambiente que fomentou um risco grave de possíveis atos de violência. Quando o Sr. Trump fez a publicação, havia um risco claro e imediato de dano e as palavras dele de apoio aos envolvidos no tumulto legitimou as ações violentas dessas pessoas. Como presidente, o Sr. Trump tinha muita influência. O alcance das publicações dele era grande, com 35 milhões de seguidores no Facebook e 24 milhões no Instagram.

Devido à gravidade das violações e ao risco contínuo de violência, foi justificada a decisão do Facebook de suspender as contas do Sr. Trump, em 6 de janeiro, e de estender essa suspensão, em 7 de janeiro.

Entretanto, não foi apropriada a imposição do Facebook de uma suspensão “indefinida”.

Não é admissível que o Facebook mantenha um usuário fora da plataforma por um período indefinido, sem critérios em relação a quando ou se a conta será restaurada.

Ao aplicar essa penalidade, o Facebook não seguiu um procedimento claro e publicado. Suspensões “indefinidas” não estão descritas nas políticas de conteúdo da empresa. As penalidades normais do Facebook incluem remoção do conteúdo em violação, imposição de uma suspensão com prazo determinado ou desativação permanente da página e da conta.¹⁴

Resumidamente, embora o Comitê tenha apontado os graves riscos de violência decorrentes da publicação dos conteúdos pelo ex-presidente à época, confirmando tratar-se de violação das políticas da comunidade, refutou a fixação de uma suspensão por prazo indefinido, sanção que não atenderia aos critérios de necessidade e adequação defendidos em decisões anteriores e previstos no Plano de Rabat.

Assim, é justamente a partir da controversa decisão do *Facebook* de bloquear indefinidamente o perfil do presidente, o processo revisional ERSR passou a ser questionado e repensado, de modo que, quando o caso do capitólio foi enviado ao Comitê de Supervisão, suscitou não apenas uma decisão específica, como também um parecer consultivo, ou seja, a formalização de uma recomendação acerca do funcionamento da verificação cruzada, tratando especialmente da ERSR.

O núcleo central das recomendações pertinentes ao caso se concentrou na

¹⁴ FORMER President Trump's Suspension. **Oversight Board**, 2021, s.p.

necessidade de aprimoramento dos mecanismos de *compliance* da plataforma, dando especial enfoque à transparência das políticas aplicadas a tais personalidades¹⁵. O Comitê fez expressa menção ao caso Trump, a partir do qual releva ter passado a demandar mais atenção às políticas de *cross-check*. Foi a partir deste que o paradigma revisional referente aos políticos passou a ser repensado, como declara o próprio CS, à luz do compromisso da *Meta* com os direitos humanos e com seus valores declarados¹⁶.

Desde então, a *Meta* tem demonstrado relativo empenho em aprimorar suas diretrizes de moderação, precisamente no intuito de permitir mais transparência no controverso processo de *cross-check*. O esforço principal consiste em parametrizar e publicizar novas práticas de verificação cruzada, revendo, em especial, a clareza dos seus padrões, bem como os critérios que determinam quais indivíduos e entidades poderão ser objeto da ERSR.

3. AS RECOMENDAÇÕES DO COMITÊ

Ao tratar do caso Trump, a primeira linha da ementada decisão do Comitê deixa clara a determinação de reversão do bloqueio do ex-presidente nas plataformas *Facebook* e *Instagram*. O fundamento se deu, em especial, em razão de uma alegada desproporcionalidade entre a violação e uma sanção de suspensão por prazo indeterminado. De fato, já em casos anteriores, o Comitê vinha insistindo na proporcionalidade da atuação da plataforma, para cuja consecução seria necessário o incremento de sanções menos danosas do que a remoção de conteúdo. Em se tratando de suspensão indeterminada, sanção ainda mais gravosa, o CS manteve a coerência.

Na contramão, embora o próprio *Facebook* admita não ter lançado mão de qualquer exceção em relação a notícias de valor jornalístico, o CS deixou clara que a categoria não deveria ser priorizada, especialmente quando da presença de um “risco de dano significativo”¹⁷. Nesse sentido, apesar de assentir com a maior potencialidade danosa de conteúdos propagados por chefes de Estado e altos funcionários, o Comitê foi incisivo em negar a possibilidade de uma sanção perpétua – ou, ao menos, indeterminada –, exigindo que a fixação de penalidade seja condizente à potencialidade danosa futura.

Embora não tenha sido taxativo nesse sentido, tudo indica que o Comitê concordou com a reprovabilidade da conduta do usuário, mas discordou da sanção de

¹⁵ VUKČEVIĆ. Ivana. Facebook Oversight Board's Decision on the Indefinite Suspension of Donald Trump's Account. *PravniZapisi*. v. 12, n. 1, p. 302, 2021.

¹⁶ FORMER President Trump's Suspension. *Oversight Board*, 2021, s.p.

¹⁷ POLICY advisory opinion on Meta's cross-check program. *Oversight Board*. 2021, p. 3.

suspensão por tempo indeterminado então aplicada pela Meta. Apesar disso, as recomendações do Comitê, como se verá adiante, caminharam muito mais no sentido de legitimar a conduta de um tratamento diferenciado da plataforma, do que de propor soluções para eventual conduta reprovável.

Em suas recomendações, o Comitê tratou sobre a dicotomia entre as limitações das capacidades humanas no campo de moderação de conteúdos e da necessidade de observância de direitos humanos, inclusive pela falta de transparência em suas decisões:

[...] o Conselho considera que, embora o Meta caracterize a verificação cruzada como um programa para proteger vozes vulneráveis e importantes, parece ser mais diretamente estruturado e calibrado para satisfazer a interesses empresariais. Apesar de o Conselho entender que a Meta é uma empresa e deveria ser capaz de desenvolver políticas que atendam às demandas empresariais, essas mesmas políticas não devem ser caracterizadas como mitigação de riscos aos direitos humanos se não atenderem a esse objetivo. Além disso, se as escolhas de modelo de negócios da Meta impactam negativamente os direitos humanos, esta deverá identificar e então prevenir, mitigar ou cessar esses impactos negativos através de melhorias no programa.¹⁸

Com base nisso, conclui que, ao tempo em que a Meta estabelece uma moderação específica para determinadas pessoas *high-profile*, cria uma distinção a princípio não justificada que, segundo afirma, não contribuiria com a prevenção de erros na moderação.¹⁹

Especificamente quanto à listagem, o Comitê determinou que a Meta tomasse medidas de transparência tanto para a pessoa objeto de revisão específica, como em relação à comunidade. Resumidamente, é preciso que a personalidade saiba, de antemão, acerca da excepcionalidade de sua moderação, bem como dos seus critérios e padrões, assentindo em observá-los. Trata-se de uma garantia que se alinha com os princípios da ampla defesa e do contraditório e que, segundo as recomendações, deve abarcar, inclusive, a possibilidade de que o perfil alvo do *cross-check* abra mão da sua condição específica de moderação²⁰.

Além disso, a Meta deveria, segundo o Comitê, reunir esforços para que a escolha das personalidades se dê com participação da comunidade, atendendo a diversidades culturais que não podem ser invisibilizadas. Assim, apontou-se a necessidade de adaptar as

¹⁸ Tradução livre de: "...the Board finds that while Meta characterizes cross-check as a program to protect vulnerable and important voices, it appears to be more directly structured and calibrated to satisfy business concerns. While the Board understands that Meta is a business and should be able to design policies that meet business concerns, these same policies should not be characterized as serving as human rights risk mitigation measures if they do not meet that objective. Additionally, if Meta's business design choices negatively impact human rights, it should identify and then prevent, mitigate, or cease those negative impacts through program improvements." (POLICY advisory opinion on Meta's cross-check program. **Oversight Board**. 2021, p. 29).

¹⁹ POLICY advisory opinion on Meta's cross-check program. **Oversight Board**. 2021, s.p.

²⁰ POLICY advisory opinion on Meta's cross-check program. **Oversight Board**. 2021, s.p.

políticas de moderação às especificidades locais, dado que nem todos os Estados detêm a mesma visibilidade a nível mundial. Eventualmente, é preciso se atentar da possibilidade de que a comunidade aponte critérios específicos para a criação de uma lista de perfis que serão excepcionalmente moderados, apontando para a imprescindibilidade de *local inputs* na dinâmica de *cross-check*.

O requisito de transparência foi repisado ao longo de todas as recomendações, exigindo que a Meta estabeleça regras e métricas precisas para a moderação de conteúdos de personalidades. É interessante notar que a preocupação do Comitê abrangeu, inclusive, a independência e imparcialidade dos moderadores, além da sua inclusão no contexto cultural em apreço. Nesse ponto, a opacidade foi dita particularmente prejudicial, não só para a moderação como atividade interna ou para o risco suportado pelo moderador, como também por inviabilizar a participação da sociedade civil e dos atores públicos na tomada de decisão.

Outra questão relevante para reduzir erros se consubstanciou na obrigatoriedade de que a Meta não viesse a priorizar a moderação em razão da pessoa, do que em razão do conteúdo. É interessante notar que, a essa altura, o Comitê admite um *distinguishing* relativizado, desde que os *standards* para tanto estejam bem definidos e sejam constantemente revistos. Há uma demonstração evidente de preocupação com os impactos das decisões de moderação de conteúdos de personalidades, impondo-se até mesmo a obrigatoriedade de que eventuais moderadores destes estejam num nível tanto performativo como hierárquico de aplicar exceções, particularmente no que tange ao contexto em questão.

Por fim, destacam-se as recomendações que demandam a criação de equipes de auditoria, bem como a divulgação de relatórios públicos acerca do *quantum* de conteúdos removidos e de sua caracterização final, seja como positivo, falso-positivo ou falso-negativo. Além disso, a Meta ficou advertida de que deverá ofertar ao Comitê relatórios e dados específicos, cujo acesso, embora a princípio restrito, possa ser viabilizado à comunidade científica para fins de aprimoramento, especialmente em razão de eventual interesse público.

Trata-se da recomendação mais longa e específica formulada pelo Comitê até o momento e cujas determinações têm sido corroboradas constantemente em sua atuação. De fato, em especial o enfrentamento à opacidade das decisões das plataformas se coaduna com as críticas mais recorrentes feitas à moderação de conteúdos por elas realizada.

O que se observa é que, ao tratar do caso em apreço, o Comitê passou ao largo de qualquer recomendação especificamente de mérito em relação ao tratamento que deve ser dado a personalidades. Determinou a transparência de políticas, sem delas efetivamente tratar e impôs obrigações de caráter exclusivamente instrumental, numa tentativa de propor alterações estruturais que conduziram a uma melhor apreciação da matéria de fundo. O mesmo se deu em 2023, quando, sob argumentos formais, o Comitê manteve na plataforma

um vídeo alterado do então Presidente dos EUA Joe Biden datado de julho de 2023, que tinha o intuito de denegrir sua imagem pública (2023-033-FB-UA).

Uma das aduções mais contundentes, inclusive, é apontada por Ivana Vukčević, que questiona o nível de arbitrariedade das decisões da *Meta*, denotando possíveis influências comerciais e políticas na escolha feita pelos moderadores²¹. No mesmo sentido, Majó-Vásquez *et al*, ao debaterem sobre a suspensão de contas relevam que “essas decisões são tomadas quase unilateralmente por empresas privadas com fins lucrativos com pouca responsabilidade, supervisão ou transparência”.²²

O fato é que, como negócio, a Meta não pode se permitir criar um ambiente hostil aos usuários, que podem se sentir intimidados por atos recorrentes de bloqueio, especialmente quando comparados a atos de censura. Entretanto, se, por um lado, a condição de garantir a segurança dos usuários em participar da rede criada é fundamental para que a plataforma seja bem sucedida como empresa, por outro, também não é autorizado que esta se exima da moderação de conteúdos potencialmente danosos, especialmente quando seu impacto é de tal grandeza a interferir em regras democráticas indissociáveis a um Estado. Trata-se de um impasse que encontra respaldo já nos valores estabelecidos pela empresa, dando lugar a um conflito aparente entre as medidas relativas à “Voz” e à “Segurança”.

O dilema apenas corrobora a necessidade de adaptação das políticas de moderação de conteúdos da plataforma, que, para alcançar um ponto de equilíbrio entre a liberdade de expressão e a segurança dos usuários, deve estar sempre atenta a questões culturais e contextuais, o que demanda, ademais, constante atualização e revisão dos parâmetros fixados. A propósito, a menção Greyson K. Young, fazendo expressa referência ao caso Trump:

É improvável que a mídia social receba uma solução perfeita para como eles lidam com a moderação de conteúdo. Conteúdo que pode ter sido aceitável há um ano, ou mesmo um mês atrás, pode ser estritamente proibido agora. Os termos de uso de uma plataforma precisam ser específicos o suficiente para capturar e remover postagens que precisam ir, mas amplos o suficiente para incluir assuntos em evolução, como a insurreição do Capitólio.²³

²¹ VUKČEVIĆ, Ivana. Facebook Oversight Board's Decision on the Indefinite Suspension of Donald Trump's Account. **PravniZapisi**. p. 302.

²² No original: “...these decisions are taken almost unilaterally by private, for-profit companies with little accountability, oversight, or transparency.” (MAJÓ-VÁZQUEZ, Silvia; CONGOSTO, Mariluz; NICHOLLS, Tom; NIELSEN, Rasmus Kleis. The Role of Suspended Account in Political Discussion on Social Media: Analysis of the 2017 French, UK and German Elections. **Social Media + Society**. p. 13, jul./set. 2021).

²³ No original, “It is unlikely that social media is ever going to be given a perfect solution for how they handle content moderation. Content that may have been acceptable a year ago, or even a month ago, may be strictly prohibited now. A platform's terms of use need to be specific enough to capture and remove posts that need to go but broad enough where evolving matters like the Capitol insurrection can also be included.” (YOUNG, Greyson K. **How much is too much**: the difficulties of social media content

Com efeito, mesmo se afastada eventual apreciação de mérito, a tarefa não admite desídia, exigindo um esforço que, embora custoso, encerra-se no âmbito do risco da atividade comercial.

4. O CASO BRASILEIRO

Se feita uma análise comparada da decisão relativa ao caso Trump, as similaridades são também evidentes em relação ao caso do “8 de janeiro” brasileiro (2023-001-FB-UA - Discurso de General²⁴). Os eventos guardam semelhanças, especialmente em razão do pano de fundo eleitoral e das condutas potencialmente lesivas, aptas a gerar ou agravar episódios de violência. Em ambos os casos, a população se dividiu em duas posições basilares, uma delas mais liberal, pela prevalência da liberdade de expressão e outra mais conservadora, dizendo que seria o caso de bloqueio do conteúdo, a fim de preservar a integridade das eleições e, ao fim e ao caso, do processo democrático²⁵.

No caso dos EUA, ambas as posições se basearam na aplicação da Primeira Emenda, mas sob óticas interpretativas diversas, atendendo a uma casuística pouco ou nada esclarecedora.

Muitos críticos dentre a direita política argumentam que as plataformas de redes sociais estão abusando do seu poder e removendo uma enorme quantidade de material de base conservadora. Alegações de fraude eleitoral após as eleições presidenciais de 2020 levaram o Twitter e o Facebook, dentre outros, a sinalizar postagens que faziam esse tipo de acusação e a incluir isenções de responsabilidade alertando outros usuários sobre informações potencialmente falsas. Conteúdo associado a esses ou que espalham teorias perigosas do popular grupo de conspiração QAnon foi direcionado e removido da maioria das plataformas de mídia social após a eleição e subsequente insurreição do Capitólio dos EUA. Entre muitos conservadores, estas ações das plataformas de redes sociais parecem uma invasão das suas crenças e uma violação dos seus direitos da Primeira Emenda. (...) No extremo oposto deste espectro, muitos acreditam que as empresas de redes sociais não estão removendo conteúdo suficiente. Embora algumas pessoas possam aplaudir as plataformas de redes sociais por tomarem posição e sinalizarem ou removerem publicações relacionadas a temas como fraude eleitoral ou conspirações QAnon, essas mesmas pessoas culpam as redes sociais por permitirem que esses utilizadores e as suas mensagens prosperem em primeiro lugar. No entanto, de alguma forma surpreendente, ambos os lados concordam que a mudança precisa ocorrer – quer isso inclua a

moderation. 30 jan. 2021, p. 28, Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3792647 Acesso em 13 maio 2023).

²⁴ BRAZILIAN General Speech (2023-001-FB-UA). **Oversight Board**. 22 jun. 2023. Disponível em: <https://www.oversightboard.com/news/6509720125757695-oversight-board-overturns-meta-s-original-decision-in-brazilian-general-s-speech-case/> Acesso em 03 mar. 2023.

²⁵ YOUNG, Greyson K. **How much is too much**: the difficulties of social media content moderation. 30 jan. 2021, p. 4-5.

modificação das políticas de conteúdo da plataforma ou um mandato de intervenção por lei federal.²⁶

No Brasil, entretanto, as controvérsias foram mais evidentes, de modo que, mesmo findo o processo eleitoral, direita e esquerda deram continuidade a hostilidades públicas, sem sinais de reconciliação à vista. Talvez esse tenha sido motivo por que, quando do julgamento do caso brasileiro, o Comitê pareceu ter firmado uma posição mais contundente.

Ainda sob o fundamento da proporcionalidade, o CS reverteu a decisão de manutenção, no *Facebook*, de um vídeo brasileiro em que se dizia haver um general incitando a violência. O argumento do Comitê era de que a remoção do conteúdo seria consistente com as responsabilidades em relação aos direitos humanos, de modo que, embora louváveis, os esforços da Meta para garantir a integridade das eleições brasileiras ainda se mostravam desproporcionais em face dos riscos presentes²⁷. E acrescentou:

O Comitê reconhece que, em muitos ambientes políticos, desafiar a integridade das eleições ou do sistema eleitoral é um exercício legítimo dos direitos das pessoas à liberdade de expressão e protesto, mesmo que haja incidentes isolados de violência. Devido à mensagem política, há um nível maior de proteção (Comentário Geral n.º 37, parágrafos 19 e 32). No entanto, o Comitê observa que esse não é o caso aqui. Há uma linha crucial que distingue o discurso político protegido da incitação à violência com vistas a anular os resultados de uma eleição popular legal. Com base nos fatores delineados no Plano de Ação de Rabat, o limite da restrição de expressão foi claramente ultrapassado nesse caso.²⁸

Assim como se deu no caso Trump, o Comitê fez referência ao Plano de Rabat, que remonta aos artigos 19 e 20 do Pacto Internacional dos Direitos Civis e Políticos, ratificando os requisitos de eventual *distinguishing* de casos envolvendo a liberdade de expressão, então dividida em três categorias:

²⁶ Tradução livre: “many critics among the political right argue that social media platforms are abusing their power and removing an egregious amount of conservative-based material. Claims of election fraud following the 2020 Presidential Election led Twitter and Facebook, among others, to flag posts making such accusations and include disclaimers warning other users of potentially false information. Content associated with or spreading dangerous theories from the popular conspiracy group, QAnon have been targeted and removed from most social media platforms following the election and subsequent insurrection of the U.S. Capitol. Among many conservatives, these actions by social media platforms feel like an invasion of their beliefs and an infringement on their First Amendment rights. (...) On the opposite end of the spectrum, many believe that social media companies are not removing nearly enough content. While some people may applaud social media platforms for taking a stand and flagging or removing posts regarding things such as election fraud or QAnon conspiracies, those same people blame social media for allowing those users and their messages to flourish in the first place. However, somewhat surprisingly, both sides agree that change needs to occur—whether that includes modifying platform content policies or an interventional mandate by federal law. (YOUNG, Greyson K. **How much is too much: the difficulties of social media content moderation.**, p. 4-5).

²⁷ BRAZILIAN General Speech (2023-001-FB-UA). **Oversight Board**. s.p.

²⁸ BRAZILIAN General Speech (2023-001-FB-UA). **Oversight Board**. s.p.

[...] expressão que constitui uma ofensa criminal; expressão que não é criminalmente punível, mas pode justificar um processo civil ou sanções administrativas; expressão que não resultam em sanções criminais, civis ou administrativas, mas ainda levantam preocupações em termos de tolerância, civilidade e respeito ao direito de outros.²⁹

No caso do discurso atribuído ao General, há referência ao tipo de expressão que caracteriza, em si, uma violação criminal, um ato terrorista. No entanto, não teria sido esse o caso também do Capitólio? E mais, considerando a distinção crucial de que, neste, a pessoa que publicou o conteúdo era uma autoridade de visibilidade global, o presidente preterido das eleições, não haveria então uma ameaça ainda mais evidente, exigindo sanções mais graves – tal qual suspensão por tempo indeterminado – ao menos até o fim do possível motim?

Não foi essa a posição da Meta, mas, se aplicada a rigorosa argumentação levada a cabo no curso da apreciação do “8 de janeiro”, poderia ter sido. Ao menos teoricamente, a questão central parece ter se circunscrito, precisamente, à proporcionalidade da medida tomada, matéria também tratada no Plano de Rabat, que estabeleceu o teste tripartite de proporcionalidade rotineiramente empregado pelo Comitê em sua argumentação. De acordo com o plano, as restrições aplicadas à liberdade de expressão só serão legítimas se guardarem observância a três requisitos, quais sejam:

restrictions are clearly and narrowly defined and respond to a pressing social need; (...) the least intrusive measure available; (...) not overly broad (...); and proportionate so that benefit to protect interest outweighs the harm to freedom of expression.

Assim, no caso do Capitólio, embora a remoção fosse justificada, a suspensão por tempo indeterminado não seria, por razões de proporcionalidade.

A questão que se impõe a seguir é: a simples remoção continuada do conteúdo seria igualmente eficiente? E o tempo empregado para avaliar cada conteúdo antes de sua remoção seria suficiente para superar os riscos de sua viralização? Por fim, teria o Comitê julgado devida a suspensão por tempo determinado? Mas qual tempo? Numa escala de dias, horas, meses?

A única resposta dada pelo Comitê é de que a suspensão indeterminada não era necessária, embora suficiente, para impedir risco iminente de violência futura. Ou seja, uma resposta fundamentada num cálculo preditivo que, embora leve em conta questões

²⁹ No original: “...expression that constitutes a criminal offence; expression that is not criminally punishable, but may justify a civil suit or administrative sanctions; expression that does not give rise to criminal, civil or administrative sanctions, but still raises concern in terms of tolerance, civility and respect for the rights of others.” (UNITED NATIONS, **Annual report of the United Nations High Commissioner for Human Rights**, 2012, p. 9. Disponível em: https://www.ohchr.org/sites/default/files/Rabat_draft_outcome.pdf. Acesso em: 08/07/2023).

particulares de cada Estado envolvido no caso, foi formulada quando as consequências já tinham se dado.

Ora, esse cálculo *a posteriori*, considerando o risco de um evento já ocorrido quando da tomada da decisão parece sobremaneira primário. Beira a obviedade dizer que a imposição de uma tal medida foi insuficiente para evitar uma consequência quando esta, de fato, já ocorreu. A questão controversa situa-se, precisamente, em apontar o nexo causal entre a medida tomada pela plataforma – dentre tantas outras – e o resultado indesejado. Como fazer esse tipo de previsão, quais fatores devem ou não ser considerados *a posteriori*, objetivamente?

Contra esse raciocínio, é de se relevar a impossibilidade, aduzida por Douek, de pacificar uma única referencial global para a liberdade de expressão³⁰. A autora afirma que o Comitê tende a desconsiderar aspectos culturais e contextuais fundamentais para sua decisão, ensejando um conflito entre sensibilidade casuística e consistência legal³¹. A esse impasse soma-se o desafio já apontado por Young³², que ressalta que as políticas das plataformas precisam estar em constante mutação.

Ambos os problemas apontam para uma questão de fixação de critérios para o *distinguishing*. No caso brasileiro, a diferença crucial é não ter sido o conteúdo publicado por um *high-profile*, de modo que a revisão do conteúdo foi feita através do mecanismo do GSR, em que a reanálise se dá apenas após a recepção de eventual denúncia de um usuário. No caso do Capitólio, o conteúdo foi de autoria de um agente político constante da listagem de nomes especial relevância da Meta, o que motivou a aplicação do mecanismo revisional específico de *cross-check*, precisamente no intuito de prevenir a ocorrência de falso-positivos. Essa é a distinção crucial.

A questão é que os mecanismos revisionais, embora sejam levados a efeito de modo diverso, de acordo com critérios subjetivos, deveriam, ao menos em tese, decidir de acordo com os mesmos parâmetros objetivos em relação ao conteúdo e subjetivos em face do contexto. Se o aspecto contextual é de difícil apreciação, ao menos os critérios objetivos do conteúdo mereceriam a fixação de parâmetros determinados, para além do que consta das políticas da Meta, que fazem uma menção de todo insuficiente à potencialidade danosa futura. Isso atende à recomendação instrumental da transparência, mas também exige uma análise substancial do tema, em especial dos limites da liberdade de expressão.

Mais uma questão merece ser pontuada. No caso brasileiro, logo se verificou que

³⁰ DOUEK, Evelyn. Content Moderation as Systems Thinking. **Harvard Law Review**. v. 136 n. 2 p. 569, dez. 2022.

³¹ DOUEK, Evelyn. Content Moderation as Systems Thinking. **Harvard Law Review**. p. 569.

³² YOUNG, Greyson K. **How much is too much**: the difficulties of social media content moderation. p. 4-5.

o vídeo teria sido falsamente atribuído a um general, razão por que pretendia propagar uma *fake news*. O vídeo, embora divulgado poucos dias antes do “8 de Janeiro”, é de autoria de um vereador, datado de novembro de 2018. Isso, se não justifica integralmente, ao menos explica parcialmente a decisão inicial da plataforma pela manutenção do conteúdo, mas demonstra sua inabilidade na apreciação de contextos fáticos. O vídeo era verdadeiro, não se tratava de qualquer montagem, mas o contexto em que tinha sido colocado o tornava, além de tudo, uma *fake news*. Além disso, o conteúdo tinha sido publicado num momento de sensibilidade democrática e nele se verificava a fala de um agente político, embora possivelmente não constante da lista *high-profile*.

Com efeito, todos esses aspectos indicam, por si sós, a inadequação do conteúdo. Apesar disso, talvez essas mesmas circunstâncias tornassem dispensável, ou melhor, desproporcional, a sua remoção. A simples sinalização, pela plataforma, de tratar-se de uma distorção dos fatos, de uma *fake news*, talvez se mostrasse suficiente para fins de prevenir o risco futuro de violência. Aliás, embora o Comitê tenha decidido pela remoção, ainda que extemporânea, tem reiteradamente exigido que as plataformas se utilizem de medidas menos restritivas, sendo o bloqueio do conteúdo realizado somente em último caso. Avisos de conteúdo sensível, tarjas e restrições de impulsionamento do conteúdo, de acordo com o Comitê, podem ser medidas suficientes para evitar o dano, ao mesmo tempo em que logram preservar a liberdade de expressão.

Esse tipo de alternativa incrementa a tão discutida transparência na conduta da plataforma na moderação de conteúdos, sendo especialmente útil para conteúdos de caráter jornalístico, que visam a expor determinada distorção de fatos. No entanto, qualquer que seja a gradação da medida, os requisitos objetivos para sua aplicação devem ser claros, prévios e públicos, a fim de evitar tratamentos desiguais em circunstâncias similares. Em qualquer caso, as razões históricas e contextuais que justificariam um tratamento mais rigoroso devem ser expostas e fundamentadas. Esse é o parâmetro mínimo, mas com o qual, talvez, não nos devemos contentar

CONSIDERAÇÕES FINAIS

O ponto fulcral das recomendações que se deram a partir do caso Trump se fundamenta no papel social desempenhado pela plataforma, que não pode ser o de calar discursos divergentes, indispensáveis para a dinâmica democrática, nem o de permitir que os políticos utilizem de sua influência para finalidades escusas. O objetivo parece ter sido, precisamente, viabilizar uma moderação de conteúdos políticos em conformidade com os direitos humanos e demais compromissos da plataforma, prevenindo, por outro lado, uma moderação efetivamente política, no sentido mais lobista e pejorativo do termo.

A garantia da segurança dos usuários e, portanto, a legitimidade do jogo democrático demanda das plataformas muito mais do que a observância de aspectos instrumentais como o tão frequentemente repisado requisito de transparência. A ação do CS, embora tenha excedido as melhores expectativas, pode ser mais incisiva, mormente quando observada a sua legitimidade em função das repercussões intraestatais dos eventos do Capitólio e do “8 de Janeiro”. Em ambos os casos verificam-se processos criminais em andamento e condenações substanciais dos envolvidos, o que corrobora o alinhamento do Comitê com a justiça pública, sem prejuízo à soberania.

Por outro lado, merece atenção a preocupação fundada do Comitê de evitar julgamentos *ad hoc*. O mecanismo revisional *high-profile* não pode ser confundido com uma moderação de exceção e, para tanto, é fundamental a transparência em todos os seus processos, especialmente pelos indivíduos aos quais se aplica. Se se trata de uma prerrogativa legítima, reconheça-se que qualquer privilégio somente faz sentido se acompanhado de excepcional grau de responsabilidade.

Por fim, vale a pena notar que as recomendações formuladas pelo Comitê, embora despojadas de caráter vinculante, têm sido pouco a pouco implementadas pela *Meta*. Ainda que o processo esteja mais lento do que o desejável, o CS tem acompanhado as providências tomadas e exigido constante atualização dos esforços empreendidos pela empresa. Os dados pertinentes têm sido reunidos e divulgados em campo próprio no site do Comitê, de forma notavelmente acessível e clara. Enquanto a regulamentação das plataformas não se consolida no país, espera-se que tais esforços sejam suficientes para impedir que atores privilegiados façam mau uso da liberdade de expressão *high-profile*.

REFERÊNCIAS DAS FONTES CITADAS

ANÁLISE precisa de conteúdo de alto impacto pelo nosso sistema de verificação cruzada. **Meta Transparency Center**. 2023, s.p. Disponível em: <https://transparency.fb.com/pt-pt/enforcement/detecting-violations/reviewing-high-visibility-content-accurately/> Acesso em 13 maio 2023.

BRAZILIAN General Speech (2023-001-FB-UA). **Oversight Board**. 22 jun. 2023. Disponível em: <https://www.oversightboard.com/news/6509720125757695-oversight-board-overturns-meta-s-original-decision-in-brazilian-general-s-speech-case/> Acesso em 03 mar. 2023.

DOUEK, Evelyn. Content Moderation as Systems Thinking. **Harvard Law Review**. v. 136 n. 2 p. 528-607, dez. 2022.

FORMER President Trump’s Suspension. **Oversight Board**, 2021. Disponível em: <https://www.oversightboard.com/decision/FB-691QAMHJ>. Acesso em 29 maio 2023.

HORWITZ, Jeff. Facebook Says Its Rules Apply to All. Company Documents Reveal a Secret Elite That's Exempt. **The Wall Street Journal**. New York City. 13 set. 2021. Disponível em: <https://www.wsj.com/articles/facebook-files-xcheck-zuckerberg-elite-rules-11631541353> Acesso em 29 maio 2023

KADRI, Thomas E.; KLONIK, Kate. Facebook v. Sullivan: Public Figures and Newsworthiness in Online Speech. **Southern California Law Review**. v. 93, n. 19-20, p. 37-99, 2019.

MAJÓ-VÁZQUEZ, Silvia; CONGOSTO, Mariluz; NICHOLLS, Tom; NIELSEN, Rasmus Kleis. The Role of Suspended Account in Political Discussion on Social Media: Analysis of the 2017 French, UK and German Elections. **Social Media + Society**. p. 1-20, jul./set. 2021.

META'S cross-check program. **Oversight Board**. Disponível em: <https://www.oversightboard.com/decision/PAO-NR730OFI/> Acesso em 29 maio 2023.

POLICY advisory opinion on Meta's cross-check program. **Oversight Board**. 2021. Disponível em: <https://transparency.meta.com/pt-br/pao-cross-check-policy/> Acesso em 29 maio 2023.

YOUNG, Greyson K. **How much is too much**: the difficulties of social media content moderation. 30 jan. 2021. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3792647 Acesso em 13 maio 2023.

UNITED NATIONS. **Annual report of the United Nations High Commissioner for Human Rights**. 2012. Disponível em: https://www.ohchr.org/sites/default/files/Rabat_draft_outcome.pdf. Acesso em: 08/07/2023.

UNITED NATIONS. **International Covenant on Civil and Political Rights**. Disponível em: <https://www.ohchr.org/en/instruments-mechanisms/instruments/international-covenant-civil-and-political-rights>. Acesso em: 08/07/2023.

VUKČEVIĆ, Ivana. Facebook Oversight Board's Decision on the Indefinite Suspension of Donald Trump's Account. **PravniZapisi**. v. 12, n. 1, p. 295-311, 2021.

INFORMAÇÕES DO AUTOR

Camilla Tavares de Albuquerque

Doutoranda em Direito Político e Econômico pela Universidade Presbiteriana Mackenzie. Mestra em Direito da Sociedade da Informação (FMU). Graduada em Direito (UFPE). Docente em Direito nos cursos de Pós-graduação da FMU/SP. Delegada de Polícia do Estado de São Paulo. Lattes: <https://lattes.cnpq.br/9570276997068133>. E-mail: albuquerque.camilla33@gmail.com.

Fernando Rister de Sousa Lima

Professor da Universidade Presbiteriana Mackenzie (PPGDPE) e da Universidade Católica de São Paulo (Departamento de Teoria Geral do Direito).

COMO CITAR

ALBUQUERQUE, Camilla Tavares de; LIMA, Fernando Rister de Sousa. Moderação de Conteúdos Políticos: Liberdade de Expressão High-profile. **Novos Estudos Jurídicos**, Itajaí (SC), v. 29, n. 3, p. 540-560, 2024. DOI: 10.14210/nej.v29n2.p540-560.

Recebido em: 02 de jun. de 2024

Aprovado em: 13 de nov. de 2024