

**Liberdade de expressão e a regulamentação da moderação em plataformas online:
considerações sobre o Projeto de Lei 2630/2020**

**Freedom of expression and regulation of moderation in online platforms: considerations
about the bill 2630/2020**

Lucas Aparecido de Jesus Pinheiro*

RESUMO

Preocupações com a liberdade de expressão são constantemente externadas quando o Estado atua no âmbito das plataformas digitais. No presente artigo, busca-se averiguar se a regulamentação da moderação praticada por plataformas *online* seria lesiva à liberdade de expressão no Brasil. Parte-se da hipótese de que a regulamentação não seria danosa ao livre discurso no país, porquanto obrigaria as plataformas a tratarem seu procedimento de moderação com transparência, de modo que a sociedade e o Estado pudessem melhor apurar os detalhes do caso concreto e julgar o acerto ou erro da decisão tomada pelos moderadores. Na primeira seção, a liberdade de expressão será analisada no contexto brasileiro, esclarecendo-se que as normas jurídicas existentes e a jurisprudência praticada já restringem qual tipo de expressão é legalmente admissível no país. Na segunda seção, são trazidos os conceitos de moderação e plataformas digitais e os riscos inerentes à moderação de conteúdo. Na terceira seção, são expostos os argumentos presentes na literatura sobre a regulamentação da moderação. Conclui-se com a verificação da hipótese à luz do marco teórico. Adotou-se a metodologia analítica crítica de revisão bibliográfica.

Palavras-chave: liberdade de expressão; regulamentação; plataformas online; moderação.

ABSTRACT

Concerns about freedom of expression are constantly expressed when the State acts within the scope of digital platforms. This article seeks to determine whether regulating moderation practiced by online platforms would be harmful to freedom of expression in Brazil. The hypothesis is that regulation would not be harmful to free speech in the country, since it would oblige platforms to treat their moderation procedures with transparency, so that society and the State could better investigate the details of the specific case and judge the correctness or incorrectness of the decision taken by the moderators. In the first section, freedom of expression will be analyzed in the Brazilian context, clarifying that existing legal norms and case law already restrict what type of expression is legally admissible in the country. The second section presents the concepts of moderation and digital platforms and the risks inherent in content moderation. The third section presents the arguments present in the literature on the regulation of moderation. It concludes with the verification of the hypothesis in light of the theoretical framework. The critical analytical methodology of bibliographic review was adopted.

Keywords: freedom of expression; regulation; online platforms; moderation.

Artigo submetido em 5 de julho de 2024 e aprovado em 4 de agosto de 2024. Revisado em 6 de novembro de 2024.

*Bacharel em Direito pela Pontifícia Universidade Católica de Minas Gerais, *campus* Coração Eucarístico. Mestrando em Direito pela Universidade Federal de Minas Gerais. Pós-graduando em Direito Processual pela Pontifícia Universidade Católica de Minas Gerais. E-mail: lpinheiro@dir.mest.ufmg.br

1 INTRODUÇÃO: LIBERDADE DE EXPRESSÃO E A DEMOCRACIA BRASILEIRA

A liberdade de expressão é uma característica histórica de Estados que aspiram ser democráticos. Os Estados Unidos da América, por exemplo, têm-na como direito fundamental desde a promulgação da Primeira Emenda (*Amendment I*) a sua Constituição, o que se deu em 1791 (Estados Unidos da América, 1791). Dois anos antes, na França revolucionária, o monarca Louis XVI, sob pressão da Assembleia Constituinte, ratificou a Declaração dos direitos do homem e do cidadão, de 1789 (*La Déclaration des droits de l'homme et du citoyen*), que também previu a liberdade de opinião e de comunicação dos pensamentos como um dos direitos mais valiosos do homem (França, 1789).

Após a Segunda Guerra Mundial, em 1948, a Assembleia Nacional das Nações Unidas proclamou a Declaração Universal dos Direitos Humanos, que, assim como a declaração francesa do século XVIII, contém dois artigos resguardando a liberdade de pensamento e opinião (Organização das Nações Unidas, 1948)¹.

Também nos anos seguintes ao fim da Segunda Guerra Mundial, a Lei Fundamental da República Federal da Alemanha (*Grundgesetz für die Bundesrepublik Deutschland*), promulgada em 1949, referência central ao que hoje se entende por Estado Democrático de Direito (Barroso, 2024, p. 192), listou a liberdade de expressão como um de seus direitos fundamentais (Alemanha, 1949), sendo que o Tribunal Constitucional Federal alemão pronunciou-se no sentido de que a liberdade de expressão “[...] é absolutamente essencial para uma ordem estatal livre e democrática [...]” (Alemanha, 1958, tradução nossa)².

No Brasil, a Constituição de 1988 previu como direitos fundamentais a livre manifestação do pensamento (art. 5º, IV), a liberdade de consciência e de crença (art. 5º, VI) e a livre expressão da atividade intelectual, artística, científica e de comunicação (art. 5º, IX) (Brasil, [2023a]).

Portanto, a proteção à liberdade de expressão está ligada ao surgimento constitucionalismo liberal e, após a Segunda Guerra Mundial, é um elemento indispensável de democracias constitucionais (Masferrer, 2023, p. 1444). Porém seu significado, isto é, o alcance material da proteção ao discurso, depende da cultura e até mesmo do contexto histórico vivido em determinado país.

Para o direito brasileiro, embora se trate de um direito fundamental, tem-se que uma das características da liberdade de expressão (assim como de todos os direitos) é a relatividade. Isso significa que o direito à livre manifestação de ideias não é absoluto nem pode ser utilizada para legitimar práticas ilícitas. Os direitos fundamentais, segundo Alexandre de Moraes (2018, p. 126) “[...] encontram seus limites nos demais direitos igualmente consagrados pela Carta Magna”.

Limitações à liberdade de expressão no Brasil têm por fundamento a proteção aos direitos fundamentais elencados na Constituição. Todavia isso não significa que as restrições estarão expressamente listadas no texto constitucional, mas, sim, que serão nele embasadas. A título de exemplo, podem-se citar os crimes contra a honra, previstos no Capítulo V do Código Penal (Brasil, [2024]).

O que é aceitável e o que é repreensível enquanto expressão também têm estreita relação com a jurisprudência do país em análise. É inegável o papel dos tribunais, sobretudo dos

1 “Artigo 18. Todo ser humano tem direito à liberdade de pensamento, consciência e religião; [...].

Artigo 19. Todo ser humano tem direito à liberdade de opinião e expressão; esse direito inclui a liberdade de, sem interferência, ter opiniões e de procurar, receber e transmitir informações e idéias (*sic*) por quaisquer meios e independentemente de fronteiras.”

2 “Für eine freiheitlich-demokratische Staatsordnung ist es schlechthin konstituierend [...]”.

tribunais constitucionais, na definição dos contornos dos direitos fundamentais, bem como em sua ponderação, em um Estado Democrático de Direito (Barroso, 2024, p. 206)³.

Especificamente em relação à jurisprudência brasileira sobre a liberdade de expressão, Vitor Amaral Medrado verificou que, tal qual o Tribunal Constitucional Federal alemão, o Supremo Tribunal Federal, até 2018, tendia a conferir preferência à liberdade de expressão em relação a outros direitos fundamentais, mesmo em casos envolvendo direitos da personalidade (Medrado, 2018, p. 141).

O Supremo Tribunal Federal publicou, em 2023, o primeiro volume da linha editorial ‘Supremo Contemporâneo’, cujo tema eleito foi liberdade de expressão. A publicação lista os principais casos julgados pelo Tribunal entre 2007 e 2022 em que a liberdade de expressão foi o objeto de debate, bem como os argumentos utilizados para fundamentar a decisão tomada (Brasil, 2023b).

A seguinte imagem constante da obra exemplifica limitações que o Supremo Tribunal Federal entende serem legítimas à liberdade de expressão em relação à imunidade parlamentar para o direito brasileiro:



Fonte: (Brasil, 2023b, p. 18).

3 “A fórmula envolvia a constitucionalização dos direitos fundamentais, que ficavam imunizados contra a ação eventualmente danosa do processo político majoritário: sua proteção passava a caber ao Judiciário.”

Ao longo da obra são listados outros casos em que o Supremo Tribunal Federal entendeu serem constitucionais as limitações impostas à liberdade de expressão, à liberdade de culto ou de religião, bem como casos em que a liberdade de expressão foi priorizada em relação a outros direitos fundamentais (Brasil, 2023b).

À vista disso, argumenta-se que realidade brasileira e a jurisprudência construída no que tange à liberdade de expressão não deve ser menosprezada ao tratar da regulamentação da moderação, ainda que as regras sejam eventualmente criadas pelo Poder Legislativo.

Partindo de uma premissa liberal, entende-se que quanto menor a ingerência estatal nas liberdades dos indivíduos, mais livres esses indivíduos serão (Berlin, 1969, p. 16). Isso significa que a regulamentação do processo de moderação implicaria um aumento do poder do Estado sobre o âmbito privado das plataformas digitais que, por consequência, poderia ser danoso à liberdade de expressão. Jacob Mchangama, nesse sentido, pondera que plataformas submetidas a determinado governo poderiam se tornar os “[...] executores privados da censura governamental [...]” (Mchangama, 2022, p. 317, tradução nossa)⁴.

No entanto, a hipótese levantada no presente artigo é a de que a regulamentação da moderação seria benéfica à democracia do Brasil, pois a concentração de poder pelas plataformas e a sua falta de transparência no que tange aos procedimentos de moderação adotados são mais lesivos ao Estado Democrático de Direito e à liberdade de expressão do que eventual regulamentação que as obrigue a prestar contas ao poder público e à sociedade em que atuam.

2 PLATAFORMAS DIGITAIS E MODERAÇÃO

No mundo atual, a internet e sobretudo as plataformas digitais criaram canais de comunicação que mudaram substancialmente o fluxo de informação, quando comparados à mídia tradicional (Calderaro, 2017, p. 781).

Jacob Mchangama destaca que, no início da implementação da internet, havia uma crença otimista, quase utópica, de que limites à livre expressão online não existiriam, cabendo ao próprio indivíduo avaliar sua conduta, longe das amarras da opressão estatal, especialmente em Estados não democráticos: “A natureza global da internet significava que as ferramentas clássicas de censura não poderiam mais ter a confiança de Estados autoritários para isolar seus cidadãos de compartilhar e acessar informações.” (Mchangama, 2022, p. 309, tradução nossa).⁵

No entanto, essa ilusória liberdade aparentemente irrestrita, combinada ao surgimento de plataformas centralizadas, trouxe consigo consequências graves ligadas ao modelo de funcionamento dessas plataformas, que sempre buscam o engajamento:

[...] A ascensão de plataformas centralizadas, no entanto, deu a elas a capacidade de coordenar incitação ao ódio e atingir membros de grupos minoritários que nunca teriam encontrado um panfleto neonazista ou um obscuro blog supremacista branco de outra forma. Em alguns casos, extremistas de direita violentos até usaram as mídias sociais para transmitir ao vivo atos de assassinato em massa. Bandos itinerantes de *trolls* misóginos encontraram uma audiência cativa de mulheres para assediar, abusar e envergonhar até o medo e o silêncio. Os jihadistas podem não apoiar a liberdade de expressão para cartunistas, infiéis e apóstatas, mas grupos terroristas como o ISIS abraçaram com entusiasmo as mídias sociais para gerar fundos e recrutas por meio do “choque e pavor” de propaganda astuta e inteligente, cabeças decepadas e tudo. Desinformação e propaganda enganosa prosperaram no ecossistema de plataformas

4 “[...] private enforcers of government censorship [...]”.

5 “The global nature of the internet meant that the classic tools of censorship could no longer be relied upon by authoritarian states to isolate their citizens from sharing and accessing information.”

de mídias sociais projetadas para recompensar o engajamento do usuário por meio de curtidas e compartilhamentos (Mchangama, 2022, p. 315-316, tradução nossa).⁶

Essa também parece ser a opinião expressa no Relatório do Relator Especial da Organização das Nações Unidas (ONU) sobre a promoção e proteção do direito à liberdade de opinião e expressão, David Kaye: “O público vê [nas plataformas] ódio, abuso e desinformação no conteúdo gerado pelos usuários. Governos veem recrutamento terrorista ou dissidência e oposição desconcertantes” (Organização das Nações Unidas, 2018, p. 3, tradução própria)⁷.

Os abusos acima listados pelos autores quase certamente não seriam tidos pela jurisprudência do Supremo Tribunal Federal (ousar dizer, de nenhum tribunal constitucional) como exercício legítimo do direito de liberdade de expressão. No entanto, isso não necessariamente implica que as plataformas devam ser objeto de regulamentação a fim de coibir práticas ilícitas praticadas em seu âmbito. A autorregulação também poderia ser uma alternativa menos intrusiva à regulamentação imposta pelo Estado.

Antes de adentrar nesse ponto específico, serão abordados brevemente os significados de plataformas digitais e de moderação.

A moderação, em termos gerais, precede a internet. De acordo com Lahis Kurtz, Paloma do Carmo e Victor Vieira (2021, p. 9), “[a] confecção de jornais impressos, por exemplo, envolve desde seus primórdios um trabalho de seleção das informações que serão veiculadas a cada edição”. No entanto, com o advento da internet e com as plataformas digitais, a intervenção sobre o conteúdo ocorre *posteriormente* a sua publicação, o que diferencia esse tipo de moderação daquele adotado pelas mídias tradicionais.

Tarleton Gillespie *et al.* fornecem um conceito técnico para o tipo de moderação a que se refere quando se fala de plataformas digitais:

Moderação de conteúdo – a detecção, avaliação e intervenções realizadas sobre conteúdo ou comportamento considerados inaceitáveis por plataformas ou outros intermediários de informação, incluindo as regras que impõem, o trabalho humano e as tecnologias necessárias, bem como os mecanismos institucionais de adjudicação, execução e recurso que a sustentam (Gillespie *et al.*, 2020, tradução nossa⁸).

Assim, trata-se de um procedimento em que comportamentos inaceitáveis, com base no que as próprias plataformas ou terceiros entendem por aceitáveis, são detectados e julgados, a partir de regras criadas e impostas pelas plataformas. A forma de moderação, seja ela feita por humanos ou por tecnologias (inteligência artificial e programação algorítmica, por exemplo) e os mecanismos de adjudicação, execução e os recursos também são intrínsecos ao conceito.

Definição semelhante é apresentada por Sarah T. Roberts: “Moderação de conteúdo é a prática organizada de triagem de conteúdo gerado por usuários (CGU) postados em sites de

6 “[...] The rise of centralized platforms, however, gave them the ability to coordinate hateful incitement and target members of minority groups who would have never encountered a neo-Nazi pamphlet or obscure white-supremacist blog otherwise. 42 In some cases, violent right-wing extremists even used social media to livestream acts of mass murder. Roving bands of misogynist trolls found a captive audience of women to harass, abuse, and shame into fear and silence. Jihadists may not support free speech for cartoonists, infidels, and apostates, but terrorist groups like ISIS enthusiastically embraced social media to generate funds and recruits through the “shock and awe” of slick and savvy propaganda, severed heads and all. Disinformation and misleading propaganda thrived in the ecosystem of social media platforms designed to reward user engagement through likes and shares.

7 “The public sees hate, abuse and disinformation in the content users generate. Governments see terrorist recruitment or discomfiting dissent and opposition.”

8 “Content moderation – the detection of, assessment of, and interventions taken on content or behaviour deemed unacceptable by platforms or other information intermediaries, including the rules they impose, the human labour and technologies required, and the institutional mechanisms of adjudication, enforcement, and appeal that support it”

internet, redes sociais e outros locais online, para determinar a sua adequação [...]” (Roberts, 2017, p. 1, tradução nossa)⁹.

Por sua vez, plataformas, para Tarleton Gillespie (2018, p. 17-21), são sites e serviços online que: (1) hospedam, organizam e distribuem o conteúdo produzido por usuários ou as interações sociais (2) sem terem produzido o “grosso” do conteúdo (3) construídos sobre uma infraestrutura que, além da circulação de informações, processa dados para diferentes propósitos, incluindo a geração de lucro, e (4) moderam (e devem fazê-lo) o conteúdo e a atividade de seus usuários.

A moderação, segundo o autor, é a característica definidora das plataformas. É o que as torna o que elas são:

Moderação não é um aspecto auxiliar do que as plataformas fazem. É essencial, constitucional, definidor. As plataformas não só não podem sobreviver sem moderação, como não são plataformas sem ela. A moderação está lá desde o início e sempre; no entanto, deve ser amplamente rejeitada, escondida, em parte para manter a ilusão de uma plataforma aberta e em parte para evitar a responsabilidade legal e cultural. As plataformas enfrentam o que pode ser uma contradição irreconciliável: elas são representadas como meros condutores [de conteúdo] e têm como premissa fazer escolhas sobre o que os usuários veem e dizem (Gillespie, 2018, p. 21, tradução nossa)¹⁰.

Assim como Gillespie menciona brevemente, no Relatório do Relator Especial da ONU sobre a promoção e proteção do direito à liberdade de opinião e expressão, David Kaye cita que “(...) as companhias permanecem sendo reguladoras enigmáticas, estabelecendo uma forma de “lei das plataformas” nas quais clareza, consistência, prestação de contas e remediação são enganosas” (Organização das Nações Unidas, 2018, p. 3, tradução nossa¹¹).

Esse aspecto obscuro sobre a moderação, tal qual argumentado por Gillespie na definição acima transcrita e por David Kaye no Relatório da ONU, suscita preocupações com relação à transparência das decisões tomadas.

Como defendem por Kurtz, do Carmo e Vieira, (2021, p. 19), “[...] devido à centralidade do discurso na concretização da democracia, qualquer forma de interferência, manipulação ou controle deve ser acompanhada de perto. Logo, a transparência é uma prática aliada à promoção da democracia.”

Por serem empresas privadas, poder-se-ia argumentar que as plataformas são soberanas nas regras que impõem aos usuários, vez que estes escolhem livremente participar de sua atividade. Diferentemente do poder público, as plataformas não têm a obrigação de prestar contas, porque não são financiadas pelo Estado.

Entretanto, é inegável que, no mundo moderno, as principais plataformas dominam o debate na esfera pública e são um dos mecanismos mais poderosos de expressão de ideias. Nos termos de Barrie Sander (2020), os indivíduos têm pouca escolha a não ser participar das plataformas para estarem incluídos nos debates da esfera pública do mundo atual, sendo que as grandes plataformas são consideradas por alguns como as *governadoras do discurso online* e *tutoras da esfera pública digital*.

9 “Content moderation is the organized practice of screening user-generated content (UGC) posted to Internet sites, social media, and other online outlets, in order to determine the appropriateness of the content [...]”.

10 “Moderation is not an ancillary aspect of what platforms do. It is essential, constitutional, definitional. Not only can platforms not survive without moderation, they are not platforms without it. Moderation is there from the beginning, and always; yet it must be largely disavowed, hidden, in part to maintain the illusion of an open platform and in part to avoid legal and cultural responsibility. Platforms face what may be an irreconcilable contradiction: they are represented as mere conduits and they are premised on making choices for what users see and say.”

11 “[...] the companies remain enigmatic regulators, establishing a kind of “platform law” in which clarity, consistency, accountability and remedy are elusive.”

Nesse sentido, as atividades realizadas pelas plataformas – como a moderação – são de interesse público, pois, além de serem essenciais no discurso público e para a promoção da democracia na atualidade, as plataformas lidam diretamente com direitos fundamentais de seus usuários ao remover (ou não) determinado conteúdo.

Essas matérias de ordem pública – democracia e direitos fundamentais – são regulamentadas por normas implícitas ou explícitas de governança privada das próprias plataformas, que nem sempre estarão de acordo com a legislação do país em que se encontram, tampouco com a compreensão jurisprudencial sobre o significado e abrangência da liberdade de expressão para o contexto desse mesmo país. As plataformas são empresas privadas que buscam lucro e crescimento, ainda que isso signifique prejudicar o interesse público (Sander, 2020).

Por essa razão, a autorregulação não é suficiente para que as plataformas se alinhem com a legislação e com a jurisprudência de um determinado país sobre a liberdade de expressão e sobre os outros direitos fundamentais, visto que, na existência de conflito entre a preservação da legislação/jurisprudência e a obtenção de lucro, as empresas buscarão o lucro, a menos que sejam diretamente obrigadas pelo Estado a não fazê-lo, com a aplicação de sanções em caso de descumprimento.

Ademais, em razão de uma compreensão deslocada de liberdade de expressão, já que a maioria das grandes plataformas estão sediadas nos Estados Unidos da América, conteúdos pacificamente tidos como danosos a minorias pela jurisprudência de um determinado país podem ser mantidos acessíveis pelas plataformas, enquanto expressões já aceitas como legítimas podem ser indevidamente removidas (Kurtz; do Carmo; Vieira, 2021, p. 21-22)¹², sendo essa mais uma consequência danosa da falta de transparência relacionada à moderação.

Barrie Sander também chama atenção ao fato de que há plataformas que realizam acordos informais com instituições vinculadas ao Estado para que conteúdos específicos sejam removidos com mais celeridade. Isso viola o devido processo legal e frequentemente é danoso ao direito de liberdade de expressão, pois se trata artifício que evita a apreciação da matéria pelo Poder Judiciário e, em muitos casos, não há formas de se recorrer, no âmbito das próprias plataformas, quando uma decisão de moderação é tomada (Sander, 2020). O Relatório do Relator Especial da ONU sobre liberdade de expressão e opinião também critica essa prática, afirmando ser uma espécie de “terceirização” de funções públicas (Organização das Nações Unidas, 2018).

Em vista do que foi descrito acima, a imposição de regras estatais ao procedimento de moderação pode ser vista como uma necessidade cada vez mais urgente. No entanto, uma regulamentação nesse sentido deve ser realizada com muita parcimônia. A próxima seção tratará do tema.

12 “Alguns exemplos ilustram os riscos envolvidos em reforçar ou excluir ideias pela moderação de conteúdo. Recentemente, em 2020, vieram a público, por meio de uma reportagem, regras internas de remoção de conteúdo da plataforma TikTok, consideradas ofensivas aos direitos humanos. Entre elas, eram proibidas diversas formas de expressão por pessoas ou em ambientes que não pertencessem a classes sociais privilegiadas ou idealizadas como padrão de beleza. Havia, além disso, uma instrução nos documentos para a exclusão de publicações sobre a independência do Tibet, o que representa uma barreira à livre informação e debate sobre assuntos políticos. Há também o algoritmo de corte de imagem da plataforma Twitter, denunciado pela comunidade de usuários por priorizar pessoas brancas sobre negras, mesmo quando elas não seriam o destaque. Também foi amplamente divulgado o caso da modelo Nyome Nicholas-Williams, mulher negra que teve suas fotos excluídas pelo Instagram, cuja política considerava como violação a exibição parcial de seios de mulheres que não fossem consideradas magras. A repercussão de todos esses eventos levou à adoção de medidas e ajustes nas práticas de moderação de conteúdo. Isso demonstra a relevância de permanente discussão e reavaliação das medidas de moderação de conteúdo.”

3 REGULAMENTAÇÃO DA MODERAÇÃO

Vislumbram-se dois extremos entre os quais o meio termo ainda é turvo. Talvez assim o seja em decorrência do fato de que as plataformas e o impacto que elas trouxeram sobre assuntos tão sensíveis à ordem pública sejam fenômenos relativamente recentes.

Diante do que foi até aqui argumentado, conclui-se a autorregulação não é um modelo viável para que sejam tutelados os direitos de liberdade de expressão e proteção de outros direitos fundamentais contra abusos no âmbito das plataformas digitais. Por outro lado, uma regulamentação demasiadamente austera por parte do Estado pode causar um estrangulamento da liberdade de expressão nas plataformas, mesmo em Estados democráticos (Kurtz; do Carmo; Vieira, 2021, p. 8)¹³.

David Kaye, no Relatório da ONU sobre liberdade de expressão e opinião, exemplifica como os Estados interagem com conteúdo nas plataformas:

Estados regularmente exigem que empresas restrinjam conteúdo manifestamente ilegal, como representações de abuso sexual infantil, ameaças diretas e críveis de dano e incitação à violência, presumindo que eles também preencham as condições de legalidade e de necessidade. Alguns Estados vão muito além disso e contam com a censura e a criminalização para moldar o ambiente regulatório online. Leis restritivas amplamente formuladas sobre “extremismo”, blasfêmia, difamação, discurso “ofensivo”, “notícias falsas” e “propaganda” são frequentemente usadas como pretexto para exigir que empresas suprimam discursos legítimos. Cada vez mais, os Estados miram conteúdo em plataformas online. Outras leis podem interferir com a privacidade online de maneira que impedem o exercício da liberdade de opinião e expressão. Muitos Estados também implantam ferramentas de desinformação e propaganda para limitar a acessibilidade e a confiabilidade da mídia independente (Organização das Nações Unidas, 2018, p. 6, tradução nossa)¹⁴.

Reconhece-se que é legítima a preocupação dos Estados com assuntos como segurança nacional, privacidade e direitos humanos. Mas não se pode esquecer dos ricos à liberdade de expressão, especialmente considerando a pressão imposta sobre plataformas, que em muitos casos removem conteúdo lícito/duvidoso para evitar eventual responsabilização própria (Organização das Nações Unidas, 2018).

No mesmo Relatório, sugere-se a utilização de princípios baseados em direitos humanos previstos na legislação internacional para a moderação de conteúdo feita pelas empresas. Essa abordagem à moderação dá às plataformas ferramentas conceituais para avaliar e abordar de modo holístico os impactos que o processo moderativo pode trazer em contextos diversos (Organização das Nações Unidas, 2018). Contudo, como conclui Barrie Sander (2020), essa forma de abordagem, embora útil, não deve ser vista como uma panaceia e sua implementação envolve diversas complexidades, como o ‘transplante’ de princípios genéricos de direitos humanos em regras específicas a serem aplicadas por empresas privadas.

13 “A regulação, assim, pode ser uma ferramenta institucional de garantia de adequação da moderação de conteúdo a parâmetros justos e democráticos, ou pode ampliar de forma desmedida o poder de agentes privados e públicos sobre a livre expressão.”

14 “States regularly require companies to restrict manifestly illegal content such as representations of child sexual abuse, direct and credible threats of harm and incitement to violence, presuming they also meet the conditions of legality and necessity. Some States go much further and rely on censorship and criminalization to shape the online regulatory environment. Broadly worded restrictive laws on “extremism”, blasphemy, defamation, “offensive” speech, “false news” and “propaganda” often serve as pretexts for demanding that companies suppress legitimate discourse. Increasingly, States target content specifically on online platforms. Other laws may interfere with online privacy in ways that deter the exercise of freedom of opinion and expression. Many States also deploy tools of disinformation and propaganda to limit the accessibility and trustworthiness of independent media.”

Esses princípios, todavia, podem ser observados pelos Estados na elaboração de uma legislação que aborde o tema da moderação. Seguindo as disposições do Relatório, os princípios em questão estão previstos no artigo 19 do Pacto Internacional sobre Direitos Civis e Políticos (PIDCP), ratificado pelo Brasil em 1992 (Brasil, 1992).

Segundo dispõe o PIDCP, os princípios orientadores à limitação da liberdade de expressão online devem ser: (a) *legalidade*, no sentido de que limitações à liberdade de expressão devem ser criadas por leis claras, acessíveis e precisas em seu conteúdo; (b) *legitimidade*, já que limitações à liberdade de expressão serão legítimas quando os fins previstos no artigo 19 (3) do PIDCP forem observados; e (c) *necessidade*, indicando que restrições à livre expressão online devem ser necessárias para que os fins legítimos sejam atingidos e devem representar o instrumento menos intrusivo possível (Organização das Nações Unidas, 2018). Além disso, a análise da necessidade requer que seja demonstrado de modo preciso e individualizado a natureza da ameaça e a proporcionalidade da ação a ser tomada, devendo-se estabelecer uma conexão imediata e direta (nexo causal) entre a expressão sob análise e a ameaça de dano (Sander 2020).

Ademais, é preciso que a realidade local (incluindo a construção jurisprudencial sobre os direitos humanos, como a liberdade expressão) seja levada em consideração, atentando-se a peculiaridades sociais, históricas, culturais, linguísticas e contextuais (Sander 2020).

A imposição de uma obrigação de prestar de contas sobre as plataformas também é um aspecto muito relevante que deveria ser abordado pela legislação.

Na data de elaboração deste artigo, a Lei 12.965/2014 está em vigência no Brasil. Comumente chamada de Marco Civil da Internet, a norma “[...] iniciou a delimitação de algumas regras e limites tanto para a moderação quanto para a responsabilização e atuação das plataformas.” (Kurtz; do Carmo; Vieira, 2021, p. 49).

Ainda assim, o Projeto de Lei (PL) 2.630/2020, atualmente em trâmite na Câmara dos Deputados, “[...] desponta como expressão mais concreta da atenção dos diferentes setores da população brasileira sobre o tema da moderação de conteúdo e liberdade de expressão na internet”. (Kurtz; do Carmo; Vieira, 2021, p. 50).

O Relatório do Relator Especial da ONU sobre liberdade de expressão e opinião estabelece cinco recomendações para que Estados lidem com a regulamentação das plataformas online:

Recomendações para Estados

65. Os Estados devem revogar qualquer lei que criminalize ou restrinja indevidamente a expressão, online ou offline.

66. Regulamentação inteligente, não regulamentação baseada em pontos de vista severos, deve ser a norma, focada em garantir a transparência e remediação por parte das empresas, para permitir que o público faça escolhas sobre como e se deve se envolver em fóruns online. Os Estados devem apenas buscar restringir o conteúdo após uma ordem de autoridade judicial independente e imparcial, e de acordo com o devido processo e com os padrões de legalidade, necessidade e legitimidade. Os Estados devem se abster de impor sanções desproporcionais, sejam elas multas pesadas ou prisão, a intermediários da Internet, dado seu efeito inibidor significativo sobre a liberdade de expressão.

67. Estados e organizações intergovernamentais devem abster-se de estabelecer leis ou arranjos que exijam o monitoramento ou filtragem “proativos” de conteúdo, o que é inconsistente com o direito à privacidade e provavelmente equivale a censura pré-publicação.

68. Estados devem abster-se de adotar modelos de regulamentação em que agências governamentais, em vez de autoridades judiciais, se tornem árbitros da expressão legal. Eles devem evitar delegar responsabilidade a empresas como julgadores de conteúdo, o que capacita o julgamento corporativo sobre valores de direitos humanos em detrimento dos usuários.

69. Estados devem publicar relatórios detalhados de transparência sobre todas as solicitações relacionadas a conteúdo emitidas a intermediários e envolver contribuição pública genuína em todas as considerações regulatórias (Organização das Nações Unidas, 2020, p. 19-20, tradução nossa)¹⁵.

Dentre as recomendações acima, destaca-se a do parágrafo 68 do Relatório, que preserva a competência do Poder Judiciário, e não de outros atores, como árbitro da liberdade de expressão, o que é compatível com o modelo de Estado instituído com a Constituição de 1988.

Ao analisar perfunctoriamente o PL 2.630/2020, nota-se que há indícios de que alguns de seus dispositivos destoam das recomendações listadas no Relatório, notadamente a de evitar a criação de exigências de monitoramento e de filtragem proativos, bem como a transferência para as plataformas da responsabilidade de aferir se algum conteúdo viola ou não direitos humanos ou a legislação.

Exemplo disso é o art. 12, do PL.

Art. 12. Os provedores de aplicação de internet submetidos a esta Lei devem garantir o direito de acesso à informação e à liberdade de expressão de seus usuários nos processos de elaboração e aplicação de seus termos de uso, disponibilizando mecanismos de recurso e devido processo.

§ 1º Em caso de denúncia ou de medida aplicada em função dos termos de uso das aplicações ou da presente Lei que recaia sobre conteúdos e contas em operação, o usuário deve ser notificado sobre a fundamentação, o processo de análise e a aplicação da medida, assim como sobre os prazos e procedimentos para sua contestação.

§ 2º Os provedores dispensarão a notificação aos usuários se verificarem risco:

I – de dano imediato de difícil reparação;

II – para a segurança da informação ou do usuário;

III – de violação a direitos de crianças e adolescentes;

IV – de crimes tipificados na Lei nº 7.716, de 5 de janeiro de 1989;

V – de grave comprometimento da usabilidade, integridade ou estabilidade da aplicação

[...] (Câmara dos Deputados, 2020).

Além disso, a Associação Brasileira de Jornalismo Investigativo (ABRAJI) criticou a aprovação do PL 2.630/2020 no Senado Federal em razão da celeridade em sua tramitação, tida como desproporcional, assim como pelo fato de o PL supostamente criar “diversos problemas que podem provocar um impacto desastroso e amplo para milhões de brasileiros e para a economia do país, afetando significativamente o acesso à rede e direitos fundamentais como a liberdade de expressão e a privacidade dos cidadãos e cidadãs na Internet.” (Associação Brasileira de Jornalismo Investigativo, 2020).

15 “Recommendations for States

65. States should repeal any law that criminalizes or unduly restricts expression, online or offline.

66. Smart regulation, not heavy-handed viewpoint-based regulation, should be the norm, focused on ensuring company transparency and remediation to enable the public to make choices about how and whether to engage in online forums. States should only seek to restrict content pursuant to an order by an independent and impartial judicial authority, and in accordance with due process and standards of legality, necessity and legitimacy. States should refrain from imposing disproportionate sanctions, whether heavy fines or imprisonment, on Internet intermediaries, given their significant chilling effect on freedom of expression.

67. States and intergovernmental organizations should refrain from establishing laws or arrangements that would require the “proactive” monitoring or filtering of content, which is both inconsistent with the right to privacy and likely to amount to pre-publication censorship.

68. States should refrain from adopting models of regulation where government agencies, rather than judicial authorities, become the arbiters of lawful expression. They should avoid delegating responsibility to companies as adjudicators of content, which empowers corporate judgment over human rights values to the detriment of users.

69. States should publish detailed transparency reports on all content-related requests issued to intermediaries and involve genuine public input in all regulatory considerations.”

No entanto, a despeito da aprovação no Senado Federal, fato é que o projeto, na data de confecção deste artigo, está praticamente parado na Câmara dos Deputados há cerca de quatro anos, sem perspectiva de aprovação. Embora o debate para amadurecer e aprimorar o projeto seja muito necessário, a existência de normas jurídicas sobre a moderação também o é, visto que, sem elas, assuntos públicos ficam à mercê de interesses corporativos, além de que o Poder Judiciário não possui um parâmetro concreto e objetivo para tomar decisões minimamente uniformes sobre o tema.

4 CONCLUSÃO

Conclui-se que a regulamentação do processo de moderação para o Brasil tem o potencial de ser benéfico para a democracia do país. Como demonstrado pela pesquisa bibliográfica, a autorregulação é ineficaz, pois, em um contexto de conflito de interesses entre o bem público e o lucro, as plataformas tendem a preterir o primeiro em favorecimento do último.

É muito problemático, e talvez até mesmo inconstitucional, que o Estado transfira para as plataformas a sua competência de decidir se direitos de seus cidadãos foram violados, fazendo com que as empresas se tornem árbitras da liberdade de expressão, especialmente ao se considerar que as plataformas em questão, atualmente, não possuem dever legal de transparência sobre como o conteúdo é moderado.

Ainda assim, a própria regulamentação do Estado pode ser restritiva demais e, em razão disso, também se mostrar lesiva ao direito de liberdade de expressão.

O Relatório do Relator Especial da Organização das Nações Unidas sobre a promoção e proteção do direito à liberdade de opinião e expressão traz parâmetros pertinentes a serem observados na elaboração da legislação, concebendo um mínimo de uniformidade – se seguido por diversos países – e uma orientação concreta sobre como as plataformas deveriam lidar com o processo de moderação de conteúdo levado a cabo em seu âmbito.

Além disso, argumenta-se que tanto a legislação quanto a prática da moderação devem estar em constante sintonia com a jurisprudência de um determinado país sobre o que configura uso legítimo ou abuso do direito de liberdade de expressão, vez que o posicionamento do Poder Judiciário confere um mínimo de objetividade ao que se considera ‘contexto e realidade locais’. No entanto, os parâmetros baseados em direitos humanos não são e nem devem ser vistos como uma panaceia que resolverá todos os complexos problemas apresentados pela moderação de conteúdo realizada pelas plataformas digitais.

O PL 2.630 pode não ser ideal e ainda precisa de amadurecimento para que fique em conformidade com os parâmetros da ONU. Ainda assim, a inércia legislativa deve ser alvo de críticas, visto que, enquanto a matéria não for regulamentada, a liberdade de expressão permanece frágil, porque sua definição e seus limites no âmbito digital estão completamente nas mãos de agentes privados que visam primordialmente o lucro e não estão obrigados a prestar contas à sociedade em que atuam.

REFERÊNCIAS

ALEMANHA. [Constituição (1949)]. **Grundgesetz für die Bundesrepublik Deutschland**. Berlim: Deutscher Bundestag [s. d.]. Disponível em: [https://www.bundestag.de/parlament/aufgaben/rechtsgrundlagen/grundgesetz#:~:text=Das%20Grundgesetz%20\(GG\)%20ist%20die,und%20von%20den%20Alliierten%20genehmigt.](https://www.bundestag.de/parlament/aufgaben/rechtsgrundlagen/grundgesetz#:~:text=Das%20Grundgesetz%20(GG)%20ist%20die,und%20von%20den%20Alliierten%20genehmigt.) Acesso em: 4 ago. 2024.

ALEMANHA. Tribunal Constitucional Federal (Ersten Senats). **BVerfGE 7,198 ff.** Julgado em 15 jan. 1958. Disponível em: https://www.bundesverfassungsgericht.de/SharedDocs/Entscheidungen/DE/1958/01/rs19580115_1bvr040051.html. Acesso em: 4 ago. 2024.

ASSOCIAÇÃO BRASILEIRA DE JORNALISMO CORPORATIVO. **PL das fake news ameaça privacidade e liberdade de expressão.** ABRAJI, 25 jun. 2020. Notícias, liberdade de expressão. Disponível em: <https://www.abraji.org.br/noticias/pl-das-fake-news-ameaca-privacidade-e-liberdade-de-expressao>. Acesso em 07 ago. 2024.

BARROSO, Luís Roberto. **Curso de direito constitucional contemporâneo.** 12. ed. São Paulo: SaraivaJur, 2024.

BERLIN, Isaiah. **Two concepts of liberty.** Oxford: Oxford University Press, 1969.

BRASIL. [Constituição (1988)]. **Constituição da República Federativa do Brasil.** Brasília, DF: Presidência da República [2023a]. Disponível em: https://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm. Acesso em: 2 ago. 2024.

BRASIL. **Decreto nº 592, de 6 de julho de 1992.** Atos Internacionais. Pacto Internacional sobre Direitos Civis e Políticos. Promulgação. Brasília, DF: Presidência da República, 1992. Disponível em: https://www.planalto.gov.br/ccivil_03/decreto/1990-1994/d0592.htm. Acesso em: 7 ago. 2024.

BRASIL. **Decreto-lei n.º 2.848, de 7 de dezembro de 1940.** Código Penal. Brasília, DF: Presidência da República [2024]. Disponível em: https://www.planalto.gov.br/ccivil_03/decreto-lei/del2848compilado.htm. Acesso em: 2 ago. 2024.

BRASIL. Supremo Tribunal Federal. **Liberdade de expressão.** Brasília: STF, Secretaria de Altos Estudos, Pesquisas e Gestão da Informação, 2023b. *E-book*. Disponível em: <https://www.stf.jus.br/arquivo/cms/publicacaoPublicacaoTematica/anexo/LiberdadeExpressao.pdf>. Acesso em: 2 ago. 2024.

CALDERARO, Andrea. Social Media and Politics. In: OUTHWAITE, William; TURNER, Stephen P. (orgs.). **The SAGE Handbook of Political Sociology:** Two Volume Set. Londres: Sage, 2017, p. 781-796.

CÂMARA DOS DEPUTADOS. **Projeto de Lei nº 2.630, 03 jul. 2020.** (Do Sr. Alessandro Vieira). Institui a Lei Brasileira de Liberdade, Responsabilidade e Transparência na Internet. Disponível em: <https://www.camara.leg.br/proposicoesWeb/fichadetramitacao?idProposicao=2256735>. Acesso em: 7 ago. 2024.

ESTADOS UNIDOS DA AMÉRICA. [Constituição (1787)]. **Constitution of the United States.** Washington, DC: U.S. Senate [2023]. Disponível em: <https://www.senate.gov/about/origins-foundations/senate-and-constitution/constitution.htm>. Acesso em: 2 ago. 2024.

FRANÇA. [Declaração (1789)]. **La Déclaration des droits de l'homme et du citoyen**. Paris: Elysée [2022]. Disponível em: <https://www.elysee.fr/la-presidence/la-declaration-des-droits-de-l-homme-et-du-citoyen>. Acesso em: 2 ago. 2024.

GILLESPIE, Tarleton L. *et al.* Expanding the Debate About Content Moderation. **Internet Policy Review**, v. 9 n. 4, 21 out. 2020. Disponível em: <https://policyreview.info/articles/analysis/expanding-debate-about-content-moderation-scholarly-research-agendas-coming-policy>. Acesso em: 6 ago. 2024.

GILLESPIE, Tarleton. **Custodians of the Internet**: platforms, content moderation, and the hidden decisions that shape social media. Londres e New Haven: Yale University Press, 2018.

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

MASFERRER, Aniceto. The Decline of Freedom of Expression and Social Vulnerability in Western Democracy. **International Journal for the Semiotics of Law**. Dordrecht, v. 36, p. 1443–1475, 12 mar. 2023. Disponível em: <https://link.springer.com/article/10.1007/s11196-023-09990-1#citeas>. Acesso em: 4 ago. 2024.

MCHANGAMA, Jacob. **Free speech**: a history from Socrates to social media. Nova Iorque: Basic Books, 2022.

MEDRADO, Vitor Amaral. **Liberdade de expressão e democracia no Brasil**: a consolidação da jurisprudência brasileira em diálogo com o constitucionalismo norte-americano. 2018. Tese (Doutorado em Direito) – Pontifícia Universidade Católica de Minas Gerais, Belo Horizonte, 2018. Disponível em: https://bib.pucminas.br/teses/Direito_MedradoVA_2.pdf. Acesso em: 4 ago. 2024.

MORAES, Alexandre de. Título II: Dos direitos e garantias fundamentais. In: MORAES, Alexandre de *et al.* **Constituição Federal comentada**. 1. ed. Rio de Janeiro: Forense, 2018, p. 123-135.

ORGANIZAÇÃO DAS NAÇÕES UNIDAS. [Declaração (1948)]. **Declaração Universal dos Direitos Humanos**. Unicef Brasil [s.d.]. Disponível em: <https://www.unicef.org/brazil/declaracao-universal-dos-direitos-humanos>. Acesso em: 2 ago. 2024.

ROBERTS, Sarah T. Content Moderation. In: SCHINTLER, L.A., MCNEELY, C. L. (orgs.). **Encyclopedia of Big Data**. Nova Iorque: Springer, Cham, 2017, p. 1-4.

SANDER, Barrie. Freedom of Expression in the Age of Online Platforms: The Promise and Pitfalls of a Human Rights-Based Approach to Content Moderation. **Fordham Int'l L.J.**, v. 43, n. 4, mai. 2020, p. 939-1006. Disponível em: <https://ir.lawnet.fordham.edu/ilj/vol43/iss4/3/>. Acesso em: 6 ago. 2024.