

# A moderação de conteúdo online e as batalhas entre o bem e o mal

## *Online content moderation and the battles between good and evil*

*Luciano Meneguetti Pereira<sup>1</sup>*

**RESUMO:** O presente artigo tem como objetivo analisar um dos temas mais problemáticos da sociedade digital contemporânea que é a moderação dos conteúdos produzidos na Internet, notadamente nas mídias sociais. A pesquisa justifica-se em razão das dificuldades que vêm sendo enfrentadas tanto em âmbito nacional como internacional no enfrentamento das complexas questões que a temática tem levantado. De um lado, práticas nefastas como a violência extremada, o discurso de ódio, o racismo, a intolerância (em suas mais diversas facetas), o engano viral online, teorias da conspiração e o cancelamento de pessoas promovido pelos “tribunais” da Internet têm promovido incontáveis violações de vários direitos no ambiente digital; de outro, a necessidade de uma efetiva proteção dos direitos violados por aquelas desprezíveis práticas e a tutela dos interesses das empresas proprietárias das plataformas digitais têm se mostrado cada vez mais importantes. Pela presente pesquisa conclui-se pela necessidade da moderação de conteúdo pela via estatal que observe os standards e parâmetros internacionais de proteção dos direitos humanos. Na pesquisa foi empregado o método dedutivo, utilizando-se como metodologia a pesquisa básica, qualitativa, exploratória e bibliográfica, lançando-se mão de doutrina autorizada de índole nacional e internacional.

**Palavras-chave:** Internet. Mídias Sociais. Conteúdo. Moderação. Direitos Humanos. Liberdade de Expressão.

**ABSTRACT:** This article aims to analyze one of the most problematic issues in contemporary digital society, which is the moderation of content

---

<sup>1</sup> Doutorando em Direito Internacional pelo Programa de Pós-Graduação da Faculdade de Direito da Universidade de São Paulo (USP). Mestre em Direito Constitucional pelo Instituto Toledo de Ensino (ITE) - Bauru/SP (2011). Especialista em Direito Público com ênfase em Direito Constitucional pela Universidade Potiguar (UNP) - Natal/RN (2007). Especialista em Docência no Ensino Técnico e Superior pelo Centro Universitário Toledo (UNITOLEDO) - Araçatuba/SP (em andamento). Graduação em Direito pelo Centro Universitário Toledo (UNITOLEDO) - Araçatuba/SP (2005). Professor Universitário no Programa de Pós-Graduação em Direito da Faculdade Aldete Maria Alves (FAMA) - Iturama/MG. Professor Universitário no Programa de Pós-Graduação em Direito das Faculdades Integradas de Santa Fé do Sul (FUNEC) - Santa Fé do Sul/SP. Professor Universitário no Curso de Graduação em Direito do Centro Universitário Toledo (UNITOLEDO) - Araçatuba/SP e no Curso de Graduação em Direito da Faculdade de Birigui (FABI) - Birigui/SP. Líder do Grupo de Pesquisa em Direito Internacional Contemporâneo (GPDIC) do Centro Universitário Toledo (CNPq). Pesquisador no Núcleo de Estudos em Tribunais Internacionais da Faculdade de Direito do Largo São Francisco (NETI/USP). Membro da Rede Latino-Americana e Caribenha de Educação em Direitos Humanos; e da Rede Brasileira de Educação em Direitos Humanos (RBEDH). Membro da Academia Brasileira de Direito Internacional (ABDI). Advogado.

produced on the Internet, especially in social media. The research is justified by the difficulties that have been faced both nationally and internationally in facing the complex issues that the theme has raised. On the one hand, harmful practices such as extreme violence, hate speech, racism, intolerance (in its most diverse facets), viral deception online, conspiracy theories and the cancellation of people promoted by Internet "courts" have promoted countless violations of various rights in the digital environment; on the other, the need for an effective protection of rights violated by those despicable practices and the protection of the interests of companies that own digital platforms have shown themselves to be increasingly important. This research concludes that there is a need for content moderation by the state, which observes international standards and parameters for the protection of human rights. In the research, the deductive method was used, using as methodology the basic, qualitative, exploratory, and bibliographical research, making use of authorized doctrine of national and international nature.

**Keywords:** Internet. Social media. Contents. Moderation. Human rights. Freedom of expression.

## 1. Introdução

As batalhas entre o bem e o mal têm estado presentes na vida humana desde as mais remotas eras. A história tem demonstrado que ao longo dos séculos os seres humanos têm lutado interna e externamente em um dos lados e que em algum momento de suas vidas eles podem trocar de lado, passando para o “lado negro da força”. Esse é, aliás, um dos pontos mais intrigantes dos quais tem se ocupado a filosofia ao longo dos tempos. O fato é que tais batalhas têm atravessado os séculos, mudado de feição, e estão presentes na sociedade contemporânea, tendo alcançado o mundo digital e a vida *na e através* da Internet e das plataformas digitais.

Atualmente parece ser bastante óbvio para toda e qualquer pessoa que o advento da Internet revolucionou o mundo, transformando-o profunda e positivamente de maneira irreversível. Os benefícios por ela trazidos, aliados à toda evolução tecnológica ocorrida nas últimas décadas são incontáveis e seria de pouca utilidade procurar descrevê-los aqui. Mas um ponto que importa aqui ser destacado é que o *mal* também a atingiu, se apropriando dela, potencializando lutas antigas e dando origem a novos e complexos conflitos por meio de sua utilização aliada às novas tecnologias.

O cenário é paradoxal. Se de um lado os direitos humanos talvez tenham alcançado níveis de positividade, desenvolvimento e efetivação como nunca antes na história humana (a própria Internet já foi reconhecida pela Organização das Nações Unidas (ONU) como um direito humano)<sup>2</sup>, de outro, nunca foi tão fácil violá-los, notadamente por meio da rede mundial de computadores e do uso de toda a parafernália tecnológica hoje à disposição, especialmente por meio das plataformas digitais de mídias sociais.

Nesse contexto e mais especificamente, a *liberdade de expressão* e o *acesso à informação* foram, sem dúvida, direitos que foram potencializados no mundo digital. Milhões de indivíduos que antes não tinham acesso e voz em diversas arenas, agora podem expressar suas ideias, filosofias, ideologias e convicções através da Internet, bem como acessar praticamente todo e qualquer tipo de conteúdo. Do mesmo modo podem influenciar os processos de tomada de decisões na esfera pública e participar mais ativamente dos debates públicos sobre assuntos pertinentes à condução de uma nação.

Mas por outro lado, essas liberdades também passaram a ser objetos de constantes ataques, restrições e violações indevidas, que muitas vezes têm ocorrido de maneira massiva e sistemática. Práticas nefastas como a violência extremada, o discurso de ódio, o racismo, a xenofobia, a misoginia, a intolerância (em suas mais diversas facetas), teorias da conspiração e o cancelamento de pessoas promovido pelos “tribunais” da Internet são exemplos atuais dessas violações, assim como as leis que têm sido editadas pelos Estados visando a moderação do conteúdo produzido na Internet para o combate destas práticas. O tema é realmente sensível, problemático e de difícil enfrentamento.

A expansão e a proliferação das mídias sociais nas últimas décadas constituem fatores que incrementaram a *batalha entre o bem e o mal no mundo digital*. O que se verifica na atualidade é que essas mídias têm

---

<sup>2</sup> Vide nesse sentido as Resoluções A/HRC/RES/32/13, de 18 de julho de 2016 e A/HRC/38/L.10/Rev.1, de 04 de julho de 2018.

constituído palco de verdadeiras guerras, que muitas vezes têm redundado em sérios prejuízos para a democracia e os direitos humanos e por vezes até mesmo em fatalidades envolvendo a vida humana. Além disso, nesse contexto estão em jogo os interesses econômicos das grandes corporações de tecnologia (*Big Techs*), proprietárias e controladoras das plataformas digitais e mídias sociais onde se veiculam diariamente uma quantidade incalculável de dados e informações por meio de textos, imagens e vídeos produzidos por seus usuários e clientes.

É neste cenário que os Estados nacionais começaram a editar leis visando a prevenção e o combate de violações de direitos na Internet, inclusive e notadamente no âmbito das mídias sociais, por meio da edição de regras de moderação do conteúdo que ali é produzido pelos usuários. No entanto, referidas leis não raras as vezes têm implicado em restrições indevidas e que igualmente violam direitos como a liberdade de expressão, de acesso à comunicação e a privacidade, quadros que podem facilmente conduzir à censura.

Portanto, o presente artigo tem como objetivo analisar algumas destas legislações, a fim de verificar como a moderação de conteúdo na Internet tem sido impactada por elas, aferindo a sua compatibilidade com os parâmetros internacionais de direitos humanos, e apontando ao final algumas diretrizes que podem ser seguidas pelos Estados para o alcance de uma regulação normativa que não seja atentatória contra os direitos humanos, e ao mesmo tempo consiga ser eficaz na prevenção e no combate às violações desses direitos.

Para o alcance do objetivo proposto, num primeiro momento são abordados os aspectos conceituais da liberdade de expressão, do acesso à informação e da moderação de conteúdo na Internet, em seguida são analisadas criticamente algumas leis nacionais já editadas e relativas à moderação de conteúdo online e, ao final, são apontadas algumas diretrizes para o enfrentamento da temática tratada no texto.

Na pesquisa foi empregado o método dedutivo, utilizando-se como metodologia a pesquisa básica, qualitativa, exploratória e bibliográfica, lançando-se mão de doutrina autorizada de índole nacional e internacional.

## 2. Aspectos conceituais e o mundo digital

Antes de se passar a análise da problemática central da presente pesquisa, importa que os aspectos conceituais concernentes à liberdade de expressão e à moderação de conteúdo online sejam abordados, ainda que sucintamente, com o fim de possibilitar uma melhor compreensão da pesquisa como um todo, especialmente em razão do fato de que as pessoas vivem hoje na *Era da Informação*, onde o acesso à ela e a participação na circulação de todo tipo de informação é uma característica distintiva de nosso mundo (LAIDLAW, 2015).

### 2.1 A liberdade de expressão e o acesso à informação na era da Internet

A consolidação normativa da liberdade de expressão e do acesso à informação é amplamente identificada em âmbito nacional na esmagadora maioria das Constituições do mundo, o mesmo ocorrendo no plano internacional, onde tais direitos encontram-se consagrados em uma extensa gama de instrumentos de direitos humanos tanto em nível universal como regional, tais como a Declaração Universal dos Direitos Humanos (art. 19), o Pacto Internacional de Direitos Civis e Políticos (art. 19), a Declaração Americana de Direitos e Deveres do Homem (art. 4º), a Convenção Americana Sobre Direitos Humanos (art. 13), a Convenção Europeia dos Direitos do Homem (art. 10) e a Carta Africana dos Direitos Humanos e dos Povos (art. 9º).

Mas como afirmou Norberto Bobbio (1992, p. 10) há mais de duas décadas, *uma coisa é proclamar um determinado direito, outra bem diferente é desfrutá-lo efetivamente*, especialmente quando se trata de direitos de vocação universal. Esse alerta continua mais vivo e importante do que nunca na era da Internet e da tecnologia de massa, especialmente no contexto das batalhas que têm sido travadas entre governos e Big Techs ao redor do mundo, em que o *mal* das supressões e restrições indevidas, que importam em escancaradas violações da liberdade de expressão e do acesso à informação, precisa ser combatido pelo *bem*, isto é, pelo amplo diálogo e por um esforço conjugado dos governos e das gigantes da tecnologia em prol da garantia desses direitos, conforme se verá adiante.

É evidente que a liberdade de expressão e o acesso à informação são direitos humanos que integram o patrimônio conquistado pela humanidade ao longo da história e não há dúvida de que o seu exercício foi amplamente potencializado pela Internet e pela tecnologia, um fenômeno que trouxe consigo várias implicações. As diversas mídias sociais hoje existentes, assim como bilhões de sites localizados na Internet “revolucionaram as comunicações modernas” (MCGOLDRICK, 2013, p. 125) e têm exigido importantes debates sobre o exercício da liberdade de expressão e do direito de acesso à informação e seus limites, pois neste contexto, “toda expressão ou informação gerada e compartilhada na internet ganha maior – e mais rápida – repercussão, seja ela positiva ou negativa” (RAMOS; BUCCI, 2021, p. 194-195).

As plataformas de mídias sociais, mais notadamente, revolucionaram a capacidade das pessoas em se conectar com esferas sociais, políticas e geográficas historicamente distintas e até mesmo incomunicáveis. Além disso, onde anteriormente imperava a atuação de guardiões que mitigavam e negociavam o acesso às plataformas de mídia de massa, hoje, potencialmente, qualquer pessoa e qualquer conteúdo podem alcançar milhões de pessoas em instantes. Esse desenvolvimento é capaz de oferecer

grandes oportunidades para a democratização da expressão e a diversificação do discurso público, mas também de ampliar ofensas e violações de direitos, com impactos prejudiciais e danosos à vida de milhões de pessoas.

Neste cenário, não há dúvidas de que o acesso à informação e a liberdade de expressão, pilares das democracias contemporâneas, precisam ser amplamente assegurados em toda a sua extensão no mundo cibernético<sup>3</sup>, especialmente num momento em que os governos ao redor do mundo começam a dar respostas às novidades e desafios que as mídias sociais passaram a representar.

Toda pessoa tem naturalmente a *liberdade de pensamento*, uma vez que no âmbito de sua consciência não vigoram normas, vontades ou forças externas capazes de controlar o pensar. Desse modo, é indiscutível que todo ser humano, em circunstâncias normais, detém a soberania sobre os seus pensamentos. No entanto, isso não é suficiente, pois é da própria natureza humana o não se contentar com a possibilidade de formar e manter suas opiniões e convicções dentro de si. Pelo contrário, é de sua natureza ir além e *expressar*, extrapolar os limites internos do pensamento para amplamente compartilhar suas ideias, ideologias, opiniões e convicções e, em muitos casos, até mesmo convencer os outros, fazer o proselitismo (BASTOS, 2002, p. 329).

E é por isso que aqui se afirma que o acesso à informação, bem como as liberdades de informação e de expressão são direitos que devem ser exercidos conforme todo o seu significado e extensão, bem como desfrutar de toda a proteção jurídica nacional e internacional já alcançada, o que compreende, por óbvio, tanto uma adequada e efetiva tutela da “liberdade de pensamento, que se restringe aos juízos intelectivos, como também o

---

<sup>3</sup> Não se afirma aqui que se trata de direitos absolutos, mas que tais direitos não podem ser restringidos sem a observância dos parâmetros nacionais e, sobretudo, internacionais de proteção dos direitos humanos.

externar sensações” (TAVARES, 2020, p. 497) enquanto um legítimo exercício da liberdade de expressão.

A liberdade de expressão possui muitos significados, especialmente devido a sua amplitude e longínqua construção histórica. O que se constata é que estudioso ao redor do mundo, inclusive no Brasil, por vezes se desencontram ao tentar conceituá-la (BALBINO; PEREIRA, 2018, p. 17), não obstante concordem sobre a existência de um núcleo comum. Nesse sentido, Laurence H. Tribe (1988, p. 789, *tradução nossa*) afirma que “Qualquer conceituação adequada da liberdade de expressão deve, ao invés, passar por diversas modalidades de teorias para que se possa proteger a rica variedade de formas de expressão”.

Sob uma perspectiva bastante abrangente, a liberdade de expressão pode ser entendida como a *ampla e irrestrita faculdade de qualquer ser humano de expressar o que sente e pensa a uma pessoa, grupos ou mesmo multidões, pelo canal que melhor lhe convir, prescindindo de qualquer formulação de convicções ou juízo de valor*. Trata-se de um *direito-mãe*, de um *gênero* que abarca várias outras liberdades ou de um *guarda-chuva* em cuja sombra se encontram várias outras manifestações específicas, tais como a livre manifestação do pensamento, as liberdades de consciência, de crença, religiosas, de comunicação (inclusive de imprensa), de informação, de opinião, de ensino e pesquisa, a livre expressão artística, intelectual e científica, enfim, todo um complexo de liberdades comunicativas.

Portanto, a liberdade de expressão deve ser tratada e tutelada da forma mais ampla possível pelos poderes constituídos em cada Estado nacional, *sem distinção alguma em relação ao meio pelo qual esse direito é exercido, se presencial ou fisicamente, ou através da Internet, virtualmente*.

Conforme a precisa síntese de Ingo Wolfgang Sarlet (2016, p. 493),

[...] para assegurar a sua máxima proteção e sua posição de destaque no âmbito das liberdades fundamentais, o âmbito de proteção da liberdade de expressão deve ser interpretado como o mais extenso possível, englobando tanto a manifestação de opiniões, quanto de ideias, pontos de vista, convicções, críticas,

juízos de valor sobre qualquer matéria ou assunto e mesmo proposições a respeito de fatos. Neste sentido, em princípio todas as formas de manifestação, desde que não violentas, estão protegidas pela liberdade de expressão, incluindo gestos, sinais, movimentos, mensagens orais e escritas, representações teatrais, sons, imagens, bem como as manifestações veiculadas pelos modernos meios de comunicação, como as mensagens de páginas de relacionamento, *blogs* etc.

Importante ainda para a compreensão do presente estudo é a distinção feita por Jónatas Machado entre as dimensões *substantiva* e *instrumental* da liberdade de expressão. Para o autor, “A *dimensão substantiva* compreende a actividade de pensar, formar a própria opinião e exteriorizá-la. A *dimensão instrumental* traduz a possibilidade de utilizar os mais diversos meios adequados à divulgação do pensamento” (*apud* TAVARES, 2020, p. 498).

Resta evidente que ambas as dimensões devem ser protegidas no âmbito da Internet e quando da utilização das plataformas digitais pelos usuários, inclusive as de mídias sociais. A *dimensão substantiva* aponta claramente para o fato de que todas as pessoas têm o direito de se expressar livremente na Internet, v.g., manifestando seu pensamento e suas ideias; emitindo opiniões de qualquer natureza e sobre qualquer assunto; assim como externando seus posicionamentos políticos, econômicos, ideológicos, religiosos etc. Nesse sentido, verifica-se que *essa dimensão constitui a pedra angular da liberdade de expressão*. Mas não menos importante é a *dimensão instrumental* dessa liberdade, que “compreende a possibilidade de escolher livremente o suporte físico ou técnico que se considere adequado à *comunicação* que se pretende realizar” (MACHADO, *apud* TAVARES, 2020, p. 499). Resta evidente que essa dimensão assume especial importância quando se trata do exercício e da tutela da liberdade de expressão na Internet.

Por sua vez, o direito à informação enquanto uma das manifestações da liberdade de expressão, pode ser compreendido como o

[...] *direito a ser informado*, e o *direito de acesso à informação*, que constituem a face positiva do *direito de se informar*, [e] abarcam,

nas suas diversas refrações subjetivas, um direito de não ser impedido de se informar, seja no que diz com a liberdade individual de recolher informações, seja no que diz respeito à liberdade de busca e escolha das fontes de informação (SARLET, 2021, n.p.).

Não há dúvida de que o direito à informação, tanto na perspectiva do direito de ser informado, como de acesso à informação, constituem na quadra histórica atual, elementos fundamentais para o exercício da cidadania e salvaguarda dos demais direitos e da dignidade humana, razão pela qual são componentes centrais no âmbito dos Estados Democráticos de Direito, de modo que, qualquer Estado que se afirme democrático, deve tutelá-lo efetivamente, inclusive e notadamente quando exercido na Internet, por meio das plataformas digitais de mídia social. Nesse sentido, André de Carvalho Ramos e Daniela Bucci (2021, p. 196) afirmam:

Estando a liberdade de expressão e de informação asseguradas também no âmbito da internet, o papel de destaque exercido pelas mídias e redes sociais na vida cotidiana, como dito acima, possibilitaram a democratização das informações e também a comunicação interativa, em que os indivíduos não apenas recebem a informação, como também interagem com seu interlocutor, servindo, ainda, como importante ferramenta para proliferação de movimentos sociais em prol do exercício de direitos humanos, políticos, ambientais, etc, a partir das quais inúmeras pessoas são mobilizadas ao redor do mundo.

Conforme se pode aferir pelas considerações feitas neste tópico, o *conteúdo* substantivo e instrumental, revelador da amplitude da liberdade de expressão e do direito à informação deve sempre ser interpretado e atuado na extensão mais ampla possível, notadamente na Internet, especialmente em razão do fato de que a vida das pessoas na atualidade é composta inarredavelmente por uma dimensão *real* e *virtual*.

## 2.2 A moderação de conteúdo online

No mundo digital integrado e possibilitado pela Internet existe um estado dinâmico de fluxo massivo e constante de dados e informações geradores de riqueza, que tem como um de seus poderosos propulsores o

conteúdo criado pelos usuários diariamente. De fato, quantidades inimagináveis de textos, imagens e vídeos são publicados todos os dias na Internet, notadamente nas famosas mídias sociais como o Facebook, Instagram, TikTok, Twitter e YouTube, dentre muitas outras.

A digitalização acelerada das sociedades e das economias criou uma situação em que algumas grandes plataformas controlam ecossistemas importantes na economia digital global. São empresas que surgiram como guardiãs dos mercados digitais com o poder de agir como legisladores privados, fato que tem ensejado diversos problemas para governos, usuários e clientes.

Hoje é evidente que as empresas, principalmente as proprietárias de grandes marcas, estão atentas ao conteúdo que é produzido e hospedado em suas plataformas<sup>4</sup>, bem como estabelecem maneiras de analisar e controlar esse conteúdo, a fim de evitar violações de direitos, responsabilizações legais e, sobretudo, para o alcance de seus interesses econômicos, consistentes principalmente na maximização dos lucros. Isso é fundamental para que elas possam manter um ambiente seguro e confiável para seus usuários e clientes, bem como para monitorar as influências sociais na percepção de suas marcas e cumprir as regulamentações oficiais dos Estados em que atuam. Desse modo, a *moderação de conteúdo* acaba sendo o método mais eficaz para o alcance dessas finalidades, embora seja também bastante polêmico.

Mas em que consiste essa moderação de conteúdo produzido online? Nos dicionários, em geral a palavra *moderar* implica em regular, controlar, refrear, reprimir, atenuar, diminuir, conter dentro de certos limites. Por sua

---

<sup>4</sup> Na literatura especializada existem várias definições de plataforma digital, desde as mais simples até aquelas mais técnicas e sofisticadas. Em geral, uma plataforma digital pode ser entendida como um lugar específico, um espaço destinado à troca de dados, informações, bens ou serviços entre pessoas (jurídicas e/ou privadas), além de consistir em um ambiente de participação e interação de uma comunidade. Em um conceito bastante amplo, “*Plataforma digital* é qualquer ferramenta eletrônica para comunicação que inclui software de desktop, celular, social e de e-mail que cobre sites e mídias sociais” (IGI GLOBAL, 2021, n.p., *tradução nossa*).

vez, a palavra *conteúdo*, embora abarque diversas concepções, normalmente tem a ver com um conjunto de conceitos e ideias, com algum assunto, tema, tópico, argumento, ideia, teor, matéria, objeto, questão, opinião, tese etc.

Transpondo-se a palavra para o contexto aqui estudado, moderar um conteúdo produzido na Internet significa literalmente regular e controlar o que foi produzido e postado/publicado pelo usuário em determinada plataforma digital. Por outras palavras, a moderação no sentido aqui proposto implica numa espécie de triagem de um conteúdo potencial ou realmente impróprio ou ilegal que os usuários publicam em uma determinada plataforma. Nesse sentido Sarah T. Roberts (2019, p. 33) explica que “a moderação de conteúdo comercial é a prática organizada de triagem de conteúdo gerado pelo usuário postado em sites da Internet, mídia social e outros meios de comunicação online”.

Trata-se de um processo que envolve a aplicação de regras predefinidas pelas empresas proprietárias das plataformas digitais, com base em suas próprias políticas e na legislação (nacional e internacional) aplicável, e que tem como finalidade precípua monitorar o conteúdo nelas veiculado. Como resultado desse processo, se determinado conteúdo não atende às políticas e diretrizes da plataforma, é então sinalizado como impróprio ou ilegal e providências são solicitadas aos usuários, podendo até mesmo ocorrer a sua remoção sumária. Os motivos pelos quais isso pode ocorrer são diversos, tais como violência, extremismo, discurso de ódio, intolerância, *fake news*, nudez, ofensas e violações de direitos autorais, dentre muitos outros.

Como se nota, o objetivo explícito da moderação de conteúdo é garantir que as plataformas digitais sejam um ambiente seguro e confiável para utilização pelos seus usuários e clientes e, conseqüentemente, possibilitar o alcance dos interesses econômicos das empresas proprietárias, uma vez que pelo processo de moderação é possível fazer com que o conteúdo

gerado pelos usuários siga as regras e diretrizes específicas da plataforma, em razão do estabelecimento de sua adequação para publicação.

Estes são os principais motivos pelos quais essa moderação vem sendo amplamente utilizada pelas diversas mídias sociais, sites e aplicativos hoje existentes. O fato é que em razão da imensa quantidade de conteúdo que é criado a cada segundo ao redor do mundo, as plataformas digitais baseadas em conteúdo gerado pelos usuários estão lutando para manter o controle sobre textos, imagens e vídeos inadequados, ofensivos e ilegais.

Diversos fatores entram em cena quando se trata de decidir qual é a melhor maneira de lidar com a moderação de conteúdo em uma determinada plataforma, tais como os objetivos do negócio empresarial, os tipos de conteúdo gerados pelos usuários e as especificidades da base de usuários, dentre outros. Em razão disso, diversos tipos de moderação têm surgido ao longo dos últimos tempos, conforme hoje demonstra a literatura especializada.

O primeiro tipo é a *pré-moderação*, em que o conteúdo produzido e/ou publicado pelos usuários é revisado por moderadores antes de ser liberado pela plataforma e se tornar visível para o público em geral. Nesse caso, quando um usuário publica algum texto, imagem ou vídeo, o conteúdo é enviado para a fila de revisão, “indo ao ar” somente depois que um moderador de conteúdo o aprovar. Embora seja uma das maneiras mais seguras (se não a mais segura) de controlar e, se for o caso, bloquear um conteúdo prejudicial, esse processo é bastante lento, o que não se coaduna com o acelerado mundo digital on-line.

A *pós-moderação* é um segundo tipo, em que os usuários publicam um determinado conteúdo sempre que desejarem e ele se torna visível para todos imediatamente, mas tudo o que é publicado ingressa em uma fila para moderação. Assim sendo, se algum conteúdo for sinalizado como inadequado ou ilegal e, portanto, prejudicial, será removido para a proteção de todos os demais usuários. Neste tipo de moderação, nota-se que algumas plataformas

têm se esforçado para reduzir o tempo de revisão, com o intuito de que o conteúdo prejudicial não permaneça online por muito tempo. Essa tem sido a maneira mais comum e utilizada pelas plataformas para fazer a triagem de conteúdo.

Um terceiro tipo é a *moderação reativa*, que conta com o auxílio dos próprios usuários para sinalizar o conteúdo que consideram inadequado, que ofende as regras, políticas ou diretrizes da plataforma ou é ilegal. Esse tipo de moderação pode ser utilizado como um método independente ou combinado com a *pós-moderação* para o alcance de melhores resultados. Neste último caso, os usuários podem sinalizar o conteúdo mesmo depois de ele ter passado por outros processos de moderação, o que faz com que se obtenha uma dupla rede de segurança.

Há também a *moderação distribuída*, um tipo em que a decisão de remover o conteúdo é distribuída entre os membros da comunidade, que votam pela sua manutenção ou não. Neste caso a moderação é totalmente dependente da comunidade online para revisar o conteúdo inadequado ou ilegal e removê-lo, caso seja necessário. Os usuários empregam um sistema de classificação para marcar se parte do conteúdo atende (ou não) às diretrizes das plataformas.

Por fim, um quinto tipo é a *moderação automatizada*, em que algoritmos com tecnologia de inteligência artificial (IA) analisam o conteúdo publicado pelos usuários em questão de segundos, sinalizando-o como impróprio ou ilegal e, conforme o caso, removendo-o sumariamente. É possível verificar que com a ajuda da automação, a detecção de publicações ofensivas se tornou mais rápida e eficaz, possibilitando também que os endereços dos usuários na Internet (IP – *Internet Protocol*), responsáveis pelo conteúdo prejudicial, sejam facilmente detectados e bloqueados em um curto espaço de tempo.

Na literatura aponta-se ainda a *não moderação*, caso em que nenhum controle de conteúdo é feito por uma determinada plataforma ou site, o que possibilita uma anarquia e abusos por parte dos usuários.

Seja qual for a espécie, cabe aqui ressaltar que a moderação de conteúdo é um tema complexo e polêmico, que tem dividido a opinião de especialistas ao redor do mundo<sup>5</sup>, notadamente em razão dos problemas que levanta quanto à proteção da liberdade de expressão e do direito à informação, conforme se verá a seguir.

### 3. Direitos humanos em risco: as leis nacionais de moderação de conteúdo

Nos últimos tempos as críticas dirigidas às empresas proprietárias de plataformas digitais, especialmente de mídias sociais, têm aumentado consideravelmente em relação à forma como elas têm lidado com a moderação de conteúdo produzido e publicado por seus usuários e clientes.<sup>6</sup> Tais críticas, que não serão aqui aprofundadas dada a finalidade da presente pesquisa, vão desde o abuso do poder e arbitrariedades diversas até os transtornos mentais causados aos funcionários ou terceiros que atuam como moderadores de conteúdo.<sup>7</sup>

Outro fator importante é que essas empresas passaram a ter que lidar com sérios dilemas éticos e jurídicos relacionados aos direitos humanos, v.g., o dever de prevenir e combater o conteúdo inadequado ou ilegal que atente

---

<sup>5</sup> Para um estudo mais aprofundado sobre o tema vide obra de Emily B. Laidlaw, intitulada *Regulating Speech in Cyberspace: Gatekeepers, Human Rights and Corporate Responsibility* (Cambridge University Press), e a obra de Sarah T. Roberts, cujo título é *Behind the Screen: Content Moderation in the Shadows of Social Media* (Yale University Press).

<sup>6</sup> O tema da violação de direitos pelas próprias *Big Techs* não será aqui analisado, pois escaparia aos objetivos da presente pesquisa, além de constituir uma temática que por si só ensejaria novas e distintas pesquisas.

<sup>7</sup> Por exemplo, o poder sobre o discurso; a suspensão e exclusão sumária de perfis de usuários; a utilização indevida dos dados dos usuários; problemas com a privacidade; a falta de transparência; a ingerência indevida em questões políticas e em processos eleitorais; a tomada de posições ideológicas com prejuízo a conteúdos contrários etc.

contra suas políticas e diretrizes ou mesmo contra a legislação em vigor e que lhe é aplicável, e que, portanto, é prejudicial aos seus usuários e clientes e aos seus próprios interesses, sem que isso implique em restrições e silenciamento indevido do *discurso protegido* pela liberdade de expressão que, de acordo com o direito doméstico dos Estados nacionais e o Direito Internacional, deve ser permitido e devidamente tutelado. Em síntese, a moderação de conteúdo feita por essas empresas, que implica na intervenção e remoção de conteúdo, pode afetar os direitos à liberdade de expressão, de acesso à informação e a privacidade, o que por sua vez pode facilmente acarretar uma inadmissível censura.

Neste cenário e diante da necessidade de tornar efetiva a responsabilização destas empresas, nos últimos anos muitos governos passaram a editar leis visando regulamentar a atividade das *Big Techs* que atuam em seus territórios, o conteúdo que é produzido em suas respectivas plataformas e as atividades de moderação por elas realizadas. O que se constata é que governos em todo o mundo estão cada vez mais afirmando sua autoridade sobre as plataformas digitais. Como resultado, verifica-se que atualmente há uma batalha global entre Estados e empresas de tecnologia pela regulamentação da Internet, com sérias implicações para os direitos humanos, ou seja, os principais prejudicados têm sido os usuários.

Conforme o Escritório do Alto Comissariado das Nações Unidas para os Direitos Humanos, nos últimos dois anos (2019-2020) cerca de 40 novas leis de mídia social foram adotadas em todo o mundo e outras 30 estão sendo consideradas (OHCHR, 2021, n. p.). De acordo com o relatório anual *Freedom on the Net 2021*, produzido pela *Freedom House*, autoridades de pelo menos 24 países aprovaram ou anunciaram no último ano novas leis ou regras que versam sobre como as empresas devem tratar o conteúdo produzido em suas plataformas digitais (FREEDOM HOUSE, 2021, p. 13). Tais normativas

[...] incluem, de forma variada, requisitos para a remoção de conteúdo ilegal, penalidades para certas formas de remoções, a

nomeação de representantes legais para gerenciar solicitações estatais e provisões mais fortes de transparência e devido processo legal. As medidas mais problemáticas podem resultar no aumento da censura de dissidência política, reportagens investigativas, e expressões de identidade étnica, religiosa, sexual ou de gênero, particularmente entre comunidades marginalizadas (FREEDOM HOUSE, 2021, p. 13, *tradução nossa*).

Como resultado deste quadro, o que se constata é a diminuição e/ou restrição da liberdade de expressão em nível global e uma pressão sem precedentes sobre o exercício desse direito humano, não havendo dúvidas de que esse impulso global para o controle da *Grande Tecnologia* precisa ser discutido de forma séria e comprometida pelos diversos sujeitos e atores que compõem as sociedades nacionais e a sociedade internacional.

A partir deste ponto serão analisadas brevemente algumas das mais recentes legislações nacionais e suas principais implicações para a proteção e o livre exercício dos direitos humanos, notadamente, a liberdade de expressão e o direito à informação.

### 3.1 Turquia: İnternet Ortamında Yapılan Yayınların Düzenlenmesi ve bu Yayınlar Yoluyla İşlenen Suçlarla Mücadele Edilmesi Hakkında Kanun (2020)<sup>8</sup>

A Turquia é um país que historicamente tem se destacado no cenário internacional em razão dos constantes ataques e restrições à liberdade de expressão. A sua mais recente regulamentação sobre as mídias sociais ocorreu por meio da Lei n. 7.253<sup>9</sup>, que entrou em vigor em outubro de 2020, alterando a lei de Internet turca (Lei n. 5.651/2007). Dentre as principais implicações para os direitos humanos, a nova legislação consolidou a

---

<sup>8</sup> *Lei de Regulamentação de Transmissões na Internet e Combate aos Crimes Cometidos por meio Destas Transmissões*. O conteúdo da referida lei pode ser consultado na íntegra em: <https://bit.ly/3GSNsDp>.

<sup>9</sup> O conteúdo da referida lei pode ser consultado na íntegra em: <https://bit.ly/3mbKAtC>.

censura de notícias sob o manto do “direito de ser esquecido”<sup>10</sup>, bem como simplificou a forma como os tribunais podem ordenar que as notícias sejam bloqueadas ou removidas, o que implica em restrições mais severas à liberdade de expressão (ÖĞRET, 2021).

Em meio às principais disposições desse conjunto normativo destacam-se as seguintes. As plataformas digitais de mídia social com mais de um milhão de usuários diários são obrigadas, dentre outras coisas: (i) a nomear um cidadão turco como seu representante na Turquia ou a estabelecer uma pessoa jurídica local, que ficam então responsáveis pelo cumprimento da legislação turca e sujeitos a multas aplicadas administrativa ou judicialmente pelo não cumprimento da lei e das decisões turcas, v.g., as ordens de remoção de conteúdo; (ii) a remover o conteúdo que for considerado “ofensivo” dentro do prazo de 48 horas após a notificação, sob o risco de agravamento das penalidades impostas, que incluem multas, proibições de novas publicações e limitações gradativas de largura da banda do tráfego de Internet; (iii) a apresentar relatórios preparados em turco semestralmente, contendo informações estatísticas e categóricas sobre a implementação das decisões para remover e/ou bloquear o acesso a conteúdo que lhe for notificado; e a (iv) tomar as medidas necessárias para hospedar os dados de seus usuários na Turquia (art. 6º)<sup>11</sup> (TURQUIA, 2021b).

Como resultado, verifica-se que as essas novas imposições legislativas acabaram por reduzir a capacidade das empresas de mídia social de resistir às solicitações das autoridades turcas destinadas a censurar mais severamente as vozes contrárias e de oposição, bem como o jornalismo independente e a liberdade de expressão em geral. Desde a entrada em vigor da nova lei, a maioria das empresas estabeleceu uma entidade legal na

---

<sup>10</sup> Em termos genéricos, o direito de ser esquecido, já reconhecido por tribunais nacionais e internacionais, permite que as pessoas façam petições a empresas proprietárias de plataformas digitais, para que removam resultados de pesquisa online sobre si mesmas, desde que as informações não sejam consideradas de interesse público.

<sup>11</sup> Esse artigo 6º da Lei n. 7.253/2020 alterou a redação original do art. 4º da Lei n. 5.651/2007.

Turquia, embora algumas tenham prometido que não haverá mudanças em suas políticas de moderação de conteúdo (FREEDOM HOUSE, 2021, p. 14).

### 3.2 Indonésia: Peraturan Menteri Komunikasi dan Informatika Nomor 5 (2020)<sup>12</sup>

A mais recente normativa de mídia social da Indonésia é o *Regulamento do Ministro das Comunicações e Informática n° 5*, de novembro de 2020, que rege o funcionamento dos operadores de sistemas eletrônicos privados (art. 1º) acessíveis no país, conceito que inclui as plataformas de mídia social, os mecanismos de pesquisa e as plataformas de comércio eletrônico, de jogos e serviços de comunicação. Como se nota, o escopo da lei é bastante amplo.

Dentre as principais disposições do referido regulamento em relação às plataformas de mídia social destacam-se: (i) a imposição de registro perante órgãos nacionais (art. 2º), assim como o dever de manter tal registro atualizado, sob pena de sanções administrativas como a advertência, a suspensão temporária e cessação de acesso aos serviços eletrônicos (art. 7º); (ii) a determinação de novos e severos requisitos de remoção de conteúdo (arts. 9º e 13 a 19) a uma ampla gama de empresas de tecnologia, independentemente de seu tamanho, incluindo aplicativos de mídia social, serviços de compartilhamento de conteúdo e mecanismos de pesquisa, sob pena de imposição das sanções descritas anteriormente e também da revogação de licenças para operar no país (art. 45); (iii) a necessidade de disponibilização, pelas plataformas, de um mecanismo de denúncia, que deve contar com instalações e serviços no país para a resolução das reclamações feitas pelos usuários (art. 10); (iv) a obrigação de fornecimento de acesso aos sistemas eletrônicos e/ou dados eletrônicos às autoridades,

---

<sup>12</sup> *Regulamento do Ministro das Comunicações e Informática n° 5*. O conteúdo da referida lei pode ser consultado na íntegra em: <https://bit.ly/3F2NUOZ>.

inclusive para fins penais (arts. 21 a 24 e 26 a 42); e, (v) a obrigação de nomear pelo menos uma pessoa de contato domiciliada no território da Indonésia, cuja tarefa é facilitar as solicitações de acesso a sistemas eletrônicos e/ou dados eletrônicos enviados pelas autoridades do país (art. 25) (INDONÉSIA, 2021).

É importante frisar que de acordo com a legislação em comento, uma vez notificada, a plataforma tem apenas 4 horas em situações “urgentes” (art. 15, 8) ou 24 horas nos demais casos (art. 15, 6) para remover o *conteúdo “proibido”*, amplamente conceituado como o discurso que viola qualquer lei nacional, que cria ansiedade na comunidade ou perturba a ordem pública (art. 9º).

O que se verifica é que a nova legislação constitui a mais recente adição ao arsenal legal usado pelo governo indonésio para reprimir a liberdade de expressão. Além das preocupações com os direitos humanos em relação ao seu amplo escopo, conforme se demonstrou, os prazos exíguos para a remoção de conteúdo levantam a questão de saber se as empresas proprietárias das plataformas digitais (exceto as maiores e mais poderosas economicamente falando), terão recursos para cumprir e, assim, sobreviver no mercado indonésio. Além disso, os prazos apertados também incentivam as empresas a implantar sistemas de monitoramento automatizados que muitas vezes sinalizam e censuram o conteúdo produzido pelos usuários de forma excessiva ou inconsistente.

Não há dúvidas de que nesse contexto muitos abusos podem ocorrer por parte das autoridades estatais. Sob o amparo desta lei, v.g., as autoridades da Indonésia já aplicaram as leis existentes para censurar conteúdo LGBTQIA+, críticas ao Islã e comentários sobre o movimento de independência nas províncias de Papua e Papua Ocidental (FREEDOM HOUSE, 2021, p. 14).

Ao analisar a lei em questão, a ONG de direitos humanos, Artigo 19, destacou as seguintes preocupações: a lei conta com definições muito amplas

e imprecisas, requer a localização de dados, concede às autoridades acesso aos dados sem as salvaguardas processuais adequadas, consagra avisos e ordens de remoção de conteúdo muito abrangentes e inclui penalidades excessivas pelo descumprimento de suas disposições (ARTICLE 19, 2021, p. 3).

### 3.3 Austrália: Online Safety Act (2021)<sup>13</sup>

A *Lei de Segurança Online* foi adotada na Austrália em junho de 2021. A nova legislação australiana está baseada em uma estrutura regulatória online já existente no país, estabelecida pelo *Enhancing Online Safety Act 2015*<sup>14</sup> e, como as suas congêneres, também introduziu obrigações de conformidade para as empresas que operam online por meio das plataformas digitais.

Dentre as suas principais disposições vale destacar que a lei (i) estabelece um Comissário de Segurança Eletrônica (*eSafety Commissioner*), um órgão estatutário independente e apoiado pela *Australian Communications and Media Authority*, que tem amplos poderes e como principais funções promover a segurança online no país, administrar sistemas de reclamações, gerir um esquema de remoção de conteúdo online e coordenar as ações dos órgãos e autoridades ligados à segurança online no país (art. 4º); (ii) autoriza que o Comissário de Segurança Eletrônica ordene que as empresas forneçam a identidade e dados dos usuários (art. 194); (iii) permite que o Comissário solicite ou exija que um provedor de serviços de Internet bloqueie o acesso a nomes de domínio, URLs ou endereços IP contendo material que retrata, promove, incita ou instrui em “conduta violenta abominável” (arts. 94 e 95), e (iv) autoriza que o Comissário de Segurança Eletrônica ordene que as empresas removam o conteúdo

<sup>13</sup> O conteúdo da referida lei pode ser consultado na íntegra em: <https://bit.ly/3p6xK1o>.

<sup>14</sup> O conteúdo da referida lei pode ser consultado na íntegra em: <https://bit.ly/32cJl67>.

publicado em mídias digitais, que é vagamente descrito como o abuso baseado em imagens, o abuso cibernético, a intimidação virtual ou qualquer material que seja de outra forma prejudicial, dentro do exíguo lapso temporal de 24 horas após a notificação (art. 109).

Como ocorre com outras legislações do gênero, o que se afere é que ao exigir a rápida remoção do conteúdo e incluir definições pouco claras sobre o que pode ser considerado um conteúdo proibido, a lei autoriza condutas que podem afetar desproporcionalmente o discurso legítimo de grupos marginalizados, em franca ofensa à liberdade de expressão. Dentre outros aspectos, vale destacar também que a legislação não é precisa sobre como o Comissário deve tomar as decisões, oferece poucas oportunidades para que os usuários respondam a reclamações sobre seu conteúdo, bem como engloba e dispensa o mesmo tratamento a uma variedade de empresas de Internet diferentes, em vez de diferenciá-las para fins obrigacionais com base em seu tamanho e função.

### 3.4 Rússia: Um labirinto normativo autoritário (2018-2021)

A Rússia conta com um labirinto de regulamentações pelo qual as empresas de tecnologia devem navegar para que possam atuar no país. Em geral, trata-se de leis que reforçam o autoritarismo e o controle estatal sobre a Internet, restringindo conseqüentemente a liberdade de expressão e outros direitos humanos.<sup>15</sup> Algumas delas serão analisadas a seguir.

O que se verifica na atualidade é um quadro de significativa expansão de leis e regulamentos que aumentam o controle sobre a infraestrutura da Internet, o conteúdo online e a privacidade das comunicações, com real potencial de prejudicar seriamente a capacidade das pessoas no país quanto

---

<sup>15</sup> O relatório *Online and On All Fronts. Russia's Assault on Freedom of Expression*, publicado pela ONG Human Rights Watch, demonstra como a liberdade de expressão tem sido atacada na Rússia por meio das constantes investidas legislativas opressoras. O seu conteúdo pode ser consultado na íntegra em: <https://bit.ly/3p7qDG0>.

ao exercício de seus direitos humanos online, notadamente a liberdade de expressão e o acesso à informação (HUMAN RIGHTS WATCH, 2020).

A *Lei Federal de 27 de junho de 2018 n. 155-FZ*<sup>16</sup> estabeleceu muitas administrativas por violação da lei de anonimato, que proíbe o oferecimento de serviços para contornar o bloqueio de sites proibidos. Desse modo, mecanismos de pesquisa que fornecem acesso a serviços de proxy, como redes privadas virtuais (VPNs), que permitem a um determinado usuário o acesso a conteúdo que foi banido pelo governo ou fornecem instruções para obter acesso a tal conteúdo, podem ser multados em valores significativos (RÚSSIA, 2021a).

Também em 2018, alterações<sup>17</sup> na legislação antiterrorismo da Rússia (Lei Yarovaya)<sup>18</sup> entraram em vigor, passando a exigir que as empresas de telecomunicações e Internet registradas perante autoridade russa (Roskomnadzor)<sup>19</sup> como “organizadoras de disseminação de informações” (v.g., aplicativos de mensagens instantâneas e mídias sociais) armazenem e compartilhem dados e informações<sup>20</sup> sobre seus usuários sem a necessidade de ordem judicial (art. 10.1) (RÚSSIA, 2021b). É evidente que esse tipo de determinação onera financeiramente sobremaneira as empresas que atuam no país, principalmente as menores, bem como atenta contra o direito à privacidade dos usuários.

---

<sup>16</sup> O conteúdo da referida lei pode ser consultado na íntegra em: <https://bit.ly/3yGqdd5>.

<sup>17</sup> Nesse particular, foi alterada a Lei Federal de 27 julho de 2006 n° 149-FZ, que dispõe sobre a sobre Informação, Tecnologias da Informação e Proteção da Informação. O seu conteúdo pode ser consultado na íntegra em: <https://bit.ly/32e3Pf0>.

<sup>18</sup> Também conhecido como Pacote Yarovaya, trata-se de um conjunto de duas leis federais russas (Lei de 06.07.2016 n° 374-FZ e Lei de 06.07.2016 n° 375-FZ) que entraram em vigor em 2016, alterando a lei antiterrorismo anterior e outras leis que regulam o contraterrorismo e medidas de segurança pública.

<sup>19</sup> *Serviço Federal de Supervisão de Comunicações, Tecnologia da Informação e Meios de Comunicação de Massa*, órgão russo responsável pela censura na mídia e nas telecomunicações, incluindo a mídia eletrônica, a comunicação de massa, a tecnologia da informação e telecomunicações.

<sup>20</sup> De acordo com a referida legislação, as empresas de Internet devem armazenar mensagens de texto, informações de voz, imagens, sons, vídeos e outras mensagens eletrônicas dos usuários.

Em 2019 entrou em vigor no país a *Lei da Internet Soberana* (Lei Federal de 05/01/2019 Nº 90-FZ)<sup>21</sup>, que altera várias legislações russas e, somadas a elas, amplia a já significativa capacidade das autoridades estatais de filtrar e bloquear automaticamente o conteúdo da Internet, não mais dependendo da cooperação das plataformas digitais para implementar o bloqueio. Nesse sentido, a referida lei exige que os provedores de serviços de Internet instalem equipamentos que possibilitem às autoridades russas “burlar” os provedores e bloquear automaticamente eventual conteúdo que o governo tenha proibido, bem como redirecionar o tráfego da Internet por conta própria (RÚSSIA, 2021c), sendo que pelo não cumprimento destas obrigações os provedores podem ser bloqueados no país.

Adicionalmente, em linhas gerais a lei (i) define as regras de roteamento de tráfego necessárias e organiza o controle sobre sua observância; (ii) regula a criação de uma infraestrutura necessária para garantir a operabilidade dos recursos de Internet russos no caso de as operadoras de telecomunicações russas não conseguirem se conectar a servidores raiz estrangeiros; (iii) estabelece para as autoridades em geral, operadoras de telecomunicações e proprietários de redes de tecnologia, a necessidade de realização de exercícios regulares para identificar ameaças e elaborar medidas para restaurar a operabilidade do segmento de Internet russo; e, (iv) cria a oportunidade para minimizar a transferência de dados que são trocados entre usuários russos de Internet para o exterior (RÚSSIA, 2021c).

Como um dos pontos mais críticos, a *Lei da Internet Soberana* prevê a transferência total do controle sobre as redes de comunicação online para uma agência governamental, desde o desligamento de redes em certas áreas da Rússia até o total desligamento do país da Internet “em caso de emergência” (HUMAN RIGHTS WATCH, 2020), fato que atesta as constantes investidas russas rumo ao isolamento.

---

<sup>21</sup> O conteúdo da referida lei pode ser consultado na íntegra em: <https://bit.ly/32isUp2>.

A *Lei Federal de 24 de fevereiro de 2021 n. 19-FZ*<sup>22</sup> introduziu novas e pesadas multas para as *Big Techs* que não removerem o conteúdo considerado pelo Estado como “ilegal” (RÚSSIA, 2021d). Na sequência, outra lei reforçou as obrigações das plataformas digitais de identificar e remover conteúdo banido e exigiu que elas atuem de maneira coordenada com o Roskomnadzor a respeito decisões de moderação de conteúdo (FREEDOM HOUSE, 2021, p. 14).

### 3.5 Índia: The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021<sup>23</sup>

As regras indianas de Tecnologia da Informação (Diretrizes de Intermediário e Código de Ética de Mídia Digital), atualizadas em fevereiro de 2021, compõem o arcabouço de leis nacionais que potencializam a batalha global pela regulamentação da Internet e implicam em restrições à liberdade de expressão.

Dentre suas principais disposições estão inclusas (i) a *due diligence* e novas obrigações para os intermediários de mídia social (art. 3º); (ii) um mecanismo expandido de reparação de queixas, por meio do qual os usuários podem reclamar diretamente às empresas sobre os conteúdos publicados (art. 10); (iii) a possibilidade de imediato bloqueio de informações contidas em determinados conteúdos em casos emergenciais (art. 16); e, (iv) lapsos temporais reduzidos para que as empresas possam responder a solicitações estatais de aplicação da lei, v.g., para a remoção de determinado conteúdo (art. 3º, I, d) (ÍNDIA, 2021).

Os *intermediários significativos de mídia social*, definidos como empresas com pelo menos cinco milhões de usuários (art. 2º, I, v), são obrigados a implantar ferramentas de moderação baseadas em inteligência

---

<sup>22</sup> O conteúdo da referida lei pode ser consultado na íntegra em: <https://bit.ly/3Ffn7Px>.

<sup>23</sup> O conteúdo da referida lei pode ser consultado na íntegra em: <https://bit.ly/3yCB3AG>.

artificial, a abrir escritórios no país e nomear três novos executivos locais (art. 4º), sendo que um diretor de conformidade (*Chief Compliance Officer*), v.g., deve cumprir as ordens de remoção de conteúdo emanadas de um tribunal, agência governamental ou qualquer outra autoridade competente dentro de um lapso temporal de 36 horas, sendo que esse indivíduo pode ser pessoalmente responsabilizado e condenado a penas de prisão de até sete anos, caso não cumpra as determinações jurídico-normativas (ÍNDIA, 2021).

Além disso, vale destacar que a lei indiana proíbe uma ampla gama de conteúdos que são vagamente definidos, incluindo o discurso que mina a ordem pública, a decência, a moralidade ou a soberania, a integridade e a segurança do país (art. 19, II) e os mecanismos de denúncia de usuários também podem ser utilizados de maneira abusiva por partidários do governo para promover a remoção de comentários críticos e contrários aos seus interesses (ÍNDIA, 2021).

É possível aferir que as novas normas indianas fornecem algumas melhorias para moderação de conteúdo das plataformas digitais, v.g., ao exigir que os intermediários significativos de mídias sociais notifiquem os usuários quando o seu conteúdo for removido, lhes forneçam uma justificativa clara para a decisão e uma via para apelação (art. 10) (ÍNDIA, 2021). No entanto, a ampliação das obrigações impostas às plataformas digitais de mídia social, conjugada com os requisitos de representação no país e o risco de responsabilidade criminal, certamente constituem fatores que podem restringir a vontade das empresas em resistir às solicitações de censura do Estado que não atendem aos padrões internacionais de proteção aos direitos humanos.

### 3.6 Reino Unido: Draft Online Safety Bill<sup>24</sup>

---

<sup>24</sup> O conteúdo do referido Projeto de Lei pode ser consultado na íntegra em: <https://bit.ly/3yEeOKE>.

O *Projeto de Lei de Segurança Online* do Reino Unido ainda não entrou em vigor até o presente momento. No marco deste Projeto, novas infrações criminais para usuários e executivos e multas substanciais são estabelecidas, bem como importantes mudanças que podem ter implicações globais são propostas, em busca de regulamentar as mídias sociais e a atividade das *Big Techs* no país.

Em termos de extensão, o Projeto alcança todas as empresas de tecnologia que permitem aos usuários postar seu próprio conteúdo ou interagir uns com os outros em suas plataformas, o que significa que grandes empresas como o *Facebook*, *Twitter*, *Instagram* e *YouTube* estarão sujeitos às suas disposições, além de sites de pornografia comercial como *OnlyFans* e dos mecanismos de busca como o Google.

Em linhas gerais, o Projeto impõe um “dever de cuidado” ou “dever de cautela” (*duties of care*) às grandes plataformas digitais quanto à remoção de conteúdo online e à proteção de crianças e adultos (art. 1º, 4º e ss.). Esse dever visa garantir a segurança dos usuários destas plataformas, de modo que os provedores de conteúdo devem assegurar que seus usuários não sejam expostos a um conteúdo *ilegal* (nos termos da legislação inglesa) ou *prejudicial* (embora legal), conceitos que ainda não estão claramente definidos. Definir o que é um material “legal, mas prejudicial” pode ser um dos principais desafios. Também são concedidos amplos poderes (inclusive de investigação) ao OFCOM, órgão regulador dos serviços de comunicação no Reino Unido, v.g., para regulamentar as atividades das *Big Techs*, emitindo códigos de prática relacionados a deveres específicos das grandes empresas de tecnologia (art. 29); para promover investigações (art. 75) e exercer poderes de execução (art. 80) (REINO UNIDO, 2021).

O dever de cuidado proposto no Projeto é dividido em três partes: (i) prevenir a proliferação de conteúdo e atividades ilegais, como pornografia infantil, material terrorista e crimes de ódio como o abuso racial; (ii) garantir que as crianças não sejam expostas a conteúdo prejudicial ou

impróprio; e, (iii) para as grandes empresas (descritas como serviços de “Categoria 1”), o dever de garantir que adultos sejam protegidos de conteúdo que, embora seja legal, é prejudicial (REINO UNIDO, 2021).

Atualmente, o Projeto exige que as empresas proprietárias de plataformas digitais apresentem detalhes de como seus serviços podem expor os usuários a um conteúdo prejudicial e como elas irão combater esse risco, versando inclusive sobre o potencial impacto prejudicial dos algoritmos em suas atividades (REINO UNIDO, 2021).

### 3.7 França: L'inconstitutionnalité de la Loi Avia (2020)<sup>25</sup>

Na França, a Lei n. 2020-766, de 24 de junho de 2020, apelidada de *Loi Avia*, que visa combater o conteúdo de incitação ao ódio na Internet, teve vários de seus dispositivos declarados inconstitucionais pelo Conselho Constitucional francês, em uma decisão cujos fundamentos podem servir de parâmetros tanto para outros tribunais nacionais como para cortes internacionais no tocante à proteção da liberdade de expressão em face das investidas estatais para regular a Internet.

Por meio da Decisão n. 2020-801 DC, de 18 de junho de 2020, o Conselho Constitucional anulou, além de várias outras disposições normativas, o dispositivo central da referida lei (art. 1º, II)<sup>26</sup>, que consistia

---

<sup>25</sup> O conteúdo da referida lei, bem como a decisão sobre a inconstitucionalidade de diversos dispositivos pelo Conselho Constitucional francês podem ser consultados na íntegra em: <https://bit.ly/3paDYO4>.

<sup>26</sup> Referido dispositivo criava um artigo 6, II, na Lei n. 2004-575, de 21 de junho de 2004, para a confiança na economia digital, cuja redação era a seguinte: “Art. 6- 2. – I. – *Sans préjudice des dispositions du 2 du I de l'article 6 de la présente loi, les opérateurs de plateforme en ligne au sens du I de l'article L. 111- 7 du code de la consommation qui proposent un service de communication au public en ligne reposant sur la mise en relation de plusieurs parties en vue du partage de contenus publics et dont l'activité sur le territoire français dépasse des seuils déterminés par décret sont tenus, au regard de l'intérêt général attaché au respect de la dignité humaine, de retirer ou de rendre inaccessible, dans un délai de vingt-quatre heures après notification par une ou plusieurs personnes, tout contenu contrevenant manifestement aux dispositions mentionnées aux cinquième, septième et huitième alinéas de l'article 24, à l'article 24 bis et aux troisième et quatrième alinéas de*

no dever de as mídias sociais remover o discurso de ódio “manifestamente ilegal” e vários outros tipos de conteúdo publicados em suas plataformas, em até 24 horas após serem notificadas, sob o risco de incorrer em multas pesadas de até 4% de sua receita global (FRANÇA, 2021a).

Dentre os fundamentos da decisão, o Conselho destacou que a liberdade de expressão é um direito humano que se encontra consagrado no art. 11 da Declaração dos Direitos do Homem e do Cidadão, de 1789 e no art. 34 da Constituição francesa, devendo ser protegido contra investidas arbitrárias. Além disso, ressaltou que dadas as grandes dificuldades das plataformas em avaliar o caráter de manifesta ilegalidade dos conteúdos veiculados e a possibilidade da existência de muitas denúncias (até mesmo infundadas), o prazo de 24 horas era particularmente curto para a tomada de decisões “jurídicas” muito técnicas. Afirmou ainda que a conduta exigida das plataformas é manifestamente ilegal pelo fato de não estar sujeito à intervenção prévia de um juiz nem a qualquer outra condição (FRANÇA, 2021a).

O que se nota é que o objetivo central da lei francesa era obrigar as plataformas online a assumir a maior parte da responsabilidade na remoção do discurso de ódio ilegal, fato que poderia levar as empresas a agir com muito mais cautela na análise de conteúdos, o que poderia implicar em remoções desnecessárias e atentatórias contra os direitos humanos. Além disso, a manutenção dos dispositivos normativos declarados inconstitucionais também consolidaria poderes de censura privada, já que relegava às empresas privadas de tecnologia a responsabilidade de determinar se um conteúdo é “manifestamente ilegal” ou não.

---

*l'article 33 de la loi du 29 juillet 1881 sur la liberté de la presse, aux articles 222- 33, 227- 23 et 421- 2- 5 du code pénal ainsi que, lorsque l'infraction porte sur un contenu à caractère pornographique, à l'article 227- 24 du même code.” (FRANÇA, 2021b)*

### 3.8 Brasil: Medida Provisória n. 1.068, de 6 de setembro de 2021<sup>27</sup>

O Brasil também tem experimentado tentativas de ingerência e controle das mídias sociais por parte do governo federal, como foi o caso da natimorta *MP das Redes Sociais* (Medida Provisória-MP n. 1.068, de 6 de setembro de 2021), que alterava significativamente o Marco Civil da Internet – MCI (Lei n. 12.965, de 23 de abril de 2014) e tinha o potencial de inviabilizar a moderação de perfis e conteúdo pelas plataformas digitais, notadamente aquele relacionado às notícias falsas (*fake news*) e constituía clara ofensa aos direitos fundamentais.

Dentre várias outras alterações, a referida MP incluía no MCI o art. 8º-A (*Dos direitos e das garantias dos usuários de redes sociais*), que previa, dentre outras coisas, a não exclusão, cancelamento ou suspensão, total ou parcial, de serviços e funcionalidades da conta ou do perfil, bem como a não exclusão, suspensão ou bloqueio da divulgação de conteúdo gerado pelo usuário, exceto por “justa causa”; e o art. 8º-B, segundo o qual a exclusão, o cancelamento ou a suspensão, total ou parcial, dos serviços e das funcionalidades da conta ou do perfil de usuário de redes sociais somente poderia ser realizado com “justa causa” e motivação (BRASIL, 2021).

Conforme se nota, a intenção da MP era fazer com que as empresas de mídia social somente pudessem restringir as contas e o conteúdo publicado pelos usuários em suas plataformas em circunstâncias bastante restritas (hipóteses de “justa causa”) ou para atender a ordens judiciais. O que se verifica é que a referida normativa efetivamente limitava a capacidade das empresas de fazer cumprir seus próprios termos de serviço e os esforços dos poderes instituídos no país para reduzir a desinformação e a disseminação

---

<sup>27</sup> O conteúdo da referida MP pode ser consultado na íntegra em: <https://bit.ly/3Fcqk2e>.

de notícias falsas que semeiam dúvidas sobre uma infinidade de assuntos, inclusive sobre os processos eleitorais.

O Partido Socialista Brasileiro (PSB) propôs no Supremo Tribunal Federal uma Ação Direta de Inconstitucionalidade (ADI n. 6.991/DF) com pedido de medida cautelar para determinar a imediata suspensão dos efeitos da referida MP em sua íntegra, petição que foi acatada pela Ministra Relatora Rosa Weber, em 14 de setembro de 2021. Na decisão a Ministra destacou

[...] a complexidade e a peculiaridade das diversas questões envolvidas na MP 1.068/2021. A propagação de *fake news* (notícias falsas), de discursos de ódio, de ataques às instituições e à própria democracia, bem como a regulamentação da retirada de conteúdos de redes sociais consubstanciam um dos maiores desafios contemporâneos à conformação dos direitos fundamentais (STF, 2021, p. 31-32).

Em 14 de setembro de 2021, o presidente do Senado e do Congresso Nacional rejeitou e devolveu a MP ao Executivo, o que fez com que a normativa perdesse o seu efeito legal, com o encerramento de sua tramitação no Congresso Nacional.

#### **4. Estratégias que podem ser utilizadas na batalha entre o bem e o mal na regulação e moderação de conteúdo**

As análises feitas na seção anterior deixaram evidente a existência de uma guerra global pela governança da Internet, guerra essa que se traduz em diversas batalhas que têm sido travadas entre os governos e as gigantes da tecnologia. A interferência estatal na regulação da Internet e mais especificamente das plataformas digitais de mídia social também restou evidenciada. Verificou-se que os governos ao editarem leis específicas voltadas à regulação da grande rede mundial e do uso da tecnologia, invariavelmente têm acabado por violar uma série de direitos humanos,

notadamente, a liberdade de expressão, o direito à informação e a privacidade.

Pelas legislações analisadas é possível se aferir um padrão de condutas que têm sido imposto pelos governos às *Big Techs*, que pode ser sintetizado da seguinte maneira, embora de maneira não exaustiva: (i) a necessidade de registro e nomeação de um representante legal da empresa para atuação no país e para fins de responsabilização; (ii) a constituição de uma filial no território do país; (iii) o estabelecimento de um órgão gestor estatal com amplos poderes de regulação para determinar condutas, fiscalizar e impor pesadas sanções administrativas (multas de valor elevado, suspensões e bloqueios de serviços e até mesmo o impedimento de atuar no país) àquelas empresas que descumprem a legislação local (fato com implicações globais); (iv) a edição de regras pouco claras, vagas e imprecisas relativamente à moderação de conteúdo; (v) a imposição de lapsos temporais exíguos para que as empresas tomem as medidas cabíveis para a remoção de conteúdo após terem sido notificadas; (vi) o dever de armazenamento dos dados e informações dos usuários em servidores situados no território do país; (vii) o fornecimento dos dados e informações dos usuários aos governos mediante ordem administrativa, sem que haja uma decisão judicial nesse sentido; e (viii) a obrigação administrativa de as empresas de tecnologia fornecer acesso aos sistemas eletrônicos e/ou dados eletrônicos às autoridades estatais, inclusive para fins penais.

Restou bastante claro o fato de que certas condutas normativamente impostas são no mínimo desproporcionais, quando não abusivas, discriminatórias e conseqüentemente atentatórias contra uma gama de direitos e liberdades hoje consagrados universalmente, bem como ofensivas aos regimes democráticos, de modo que refletir cuidadosamente e discutir o assunto de maneira comprometida tornou-se imperioso nesta quadra histórica das sociedades nacionais, profundamente marcadas e transformadas pela Internet e a tecnologia.

Dada a importância e a preocupação global em torno do assunto, organizações internacionais, organizações não governamentais e especialistas de todo globo passaram a desenvolver pesquisas e estudos aprofundados sobre a temática, com o fim de encontrar soluções que possam contribuir para o enfrentamento das questões suscitadas no âmbito das várias batalhas que estão sendo travadas, bem como mitigar os seus deletérios efeitos não apenas para Estados e empresas, mas sobretudo para o inquantificável número de pessoas que hoje utilizam a Internet e as mídias sociais ao redor do mundo e que são direta e indiretamente prejudicadas no exercício de seus direitos em razão destes conflitos.

Antes de prosseguir importa aqui fazer um adendo. Na presente pesquisa não se discute se as plataformas digitais de mídia social devem ou não moderar o conteúdo online. Para ser evidente que o quadro das relações e interações sociais globais atual revela essa necessidade. Tarleton Gillespie (2018, p. 5) está com a razão quando afirma que as

[...] plataformas devem, de uma forma ou de outra, moderar: tanto para proteger um usuário de outro, ou um grupo de seus antagonistas, quanto para remover o que é ofensivo, vil ou ilegal - bem como para apresentar sua melhor face aos novos usuários, para seus anunciantes e parceiros, e para o público em geral.

Portanto, a questão central não está no *se*, mas no *como* moderar e regulamentar, ou seja, em como deve se dar a atividade normativa estatal e a atuação das empresas para coibir a fala prejudicial nas plataformas digitais e ao mesmo tempo preservar a liberdade de expressão e outros direitos. E ainda nesse contexto serve o alerta: por estarem envolvidos altos interesses econômicos e políticos de ambos os lados, em certos casos até mesmo inconciliáveis, a tarefa não é nada fácil e sem dúvida precisa ser enfrentada de forma madura, responsável e coordenada, com ampla participação de todos os envolvidos (efetiva ou potencialmente) nas e pelas ações de regulamentar e moderar.

Com o fim de auxiliar na resolução dos dilemas de regulamentação e moderação de conteúdo online, em 2021 o Alto Comissariado das Nações Unidas para Direitos Humanos propôs cinco ações para consideração por parte de Estados e empresas (OHCHR, 2021, n.p.):

(i) que o foco de toda regulamentação estatal esteja na melhoria dos processos de moderação de conteúdo por parte das empresas de tecnologia, e não na adição de restrições específicas ao conteúdo, v.g., diante de questões complexas, as pessoas devem tomar as decisões e não os algoritmos;

(ii) que as restrições impostas pelos Estados estejam baseadas em leis não apenas nacionais, mas também em parâmetros normativos internacionais (vinculantes ou *soft law*), além da necessidade de ser claras, necessárias, proporcionais e não discriminatórias;

(iii) que as empresas atuem com transparência sobre como fazem a curadoria e moderam o conteúdo e como compartilham informações, e que também os Estados o sejam acerca de suas solicitações para restringir ou remover determinado conteúdo e nas solicitações de acesso a dados dos usuários;

(iv) que aos usuários sejam ofertadas oportunidades efetivas de apelar contra decisões que considerem injustas e que tribunais independentes tenham a palavra final sobre a legalidade ou não do conteúdo; e, por fim,

(v) que a sociedade civil e os especialistas estejam envolvidos na formulação e avaliação das regulamentações a serem editadas.

Parece estar fora de dúvida de que as ações propostas pela ONU, se levadas a sério e implementadas por governos e empresas, têm a aptidão de em muito contribuir no tratamento do tema em escala global, uma vez que oferecem significativas alternativas aos rígidos padrões que os governos têm adotado na regulamentação da Internet e, notadamente, em relação à forma como as empresas devem moderar o conteúdo online.

A Comissão Europeia está dando passos significativos para o enfrentamento do tema no âmbito da União Europeia (UE). Em 2020 a

Comissão apresentou ao Parlamento Europeu e ao Conselho Europeu duas iniciativas legislativas para atualizar as regras que regem os serviços digitais na UE: o *Digital Services Act (DSA)*<sup>28</sup> e o *Digital Markets Act (DMA)*<sup>29</sup>, que constituem um único conjunto de novas regras aplicáveis em toda comunidade europeia e que visa a criação de um espaço digital mais seguro e aberto a todas as pessoas.

Embora ainda não seja possível aferir a real dimensão e os impactos desse conjunto normativo, alguns elementos bastante positivos podem ser destacados, dentre eles, como o fato de estar baseado na linguagem dos direitos humanos, de conter requisitos de transparência claros para as plataformas digitais e, especialmente, de ter sido elaborado através de um amplo processo participativo (setor privado, usuários de serviços digitais, organizações da sociedade civil, autoridades nacionais, Academia, comunidade técnica, organizações internacionais e o público em geral).

Fazendo uma adaptação das regras do direito empresarial/comercial e civil para as entidades comerciais que atuam no universo online, o que se extrai é que as novas regras apelam para mais justiça, transparência e responsabilização pelos processos de moderação de conteúdo dos serviços digitais, buscando garantir que os direitos humanos sejam respeitados e que os indivíduos tenham acesso a um recurso independente de reparação judicial. Também há o estabelecimento de um mecanismo detalhado de “notificação e ação” para lidar com a veiculação de conteúdo ilegal, regras abrangentes sobre publicidade online, incluindo a publicidade direcionada, e a permissão para o desenvolvimento e uso dos chamados contratos inteligentes (EUR-Lex, 2021).

Diante das ações apresentadas pela ONU e dos projetos da Comissão Europeia, e com base em estudos realizados pelas próprias plataformas

---

<sup>28</sup> Mais informações sobre a referida proposta normativa podem ser consultadas em: <https://bit.ly/3mgQDwX>.

<sup>29</sup> Mais informações sobre a referida proposta normativa podem ser consultadas em: <https://bit.ly/3m1GiQj>.

digitais<sup>30</sup>, pela doutrina e por especialistas, alguns caminhos que podem ser trilhados por Estados e empresas ao lidar com as tormentosas questões oriundas da regulamentação e moderação de conteúdo:

(i) a ampla consideração pela gramática e linguagem dos direitos humanos consagrados no plano internacional (seja por meio de documentos vinculantes ou não), tanto na regulamentação como na moderação do conteúdo<sup>31</sup>; os reguladores e moderadores precisam levar em consideração os impactos que suas decisões podem ter sobre a liberdade de expressão e outros direitos;

(ii) o estabelecimento, pelas empresas, de canais dialógicos, interativos e com interfaces amigáveis para que os usuários possam relatar conteúdo ilegal ou prejudicial, bem como apelarem da remoção de determinado conteúdo ou da decisão de não remoção por parte da plataforma;

(iii) a determinação legal da criação de mecanismos de supervisão externa das políticas e decisões de aplicação normativa das empresas de tecnologia, bem como a exigência de apresentação de relatórios públicos periódicos à comunidade de usuários e clientes, onde se conste em linguagem clara e acessível, dentre outras informações, os padrões de moderação de conteúdo adotados e as medidas tomadas em face das denúncias feitas pelos usuários;

(iv) a estipulação legal de criação de mecanismos pelos quais as empresas de tecnologia possam fornecer aos governos e aos usuários de seus serviços, as informações de que precisam para julgar com precisão os seus esforços na moderação de conteúdo, fortalecendo-se assim a transparência, que constitui um dos vetores fundamentais no tratamento do tema;

---

<sup>30</sup> Nesse sentido, quanto aos caminhos apontados, a presente pesquisa se vale também das quatro questões-chaves e dos princípios formulados pelo Facebook, no estudo *Charting a Way Forward on Online Content Regulation*, disponível em: <https://bit.ly/3mHKWbJ>.

<sup>31</sup> Vale aqui lembrar que os *Princípios Orientadores da ONU sobre Negócios e Direitos Humanos* estipulam que todas as empresas têm a responsabilidade de respeitar os direitos humanos.

(v) a previsão legal de prazos apropriados para a adoção de medidas pelas empresas, sobretudo quanto à remoção de conteúdo, que sejam condizentes com a complexidade da atividade que deve ser desenvolvida para o atendimento das solicitações, sejam elas de usuários ou das autoridades estatais;

(vi) a edição de uma regulamentação capaz de incentivar as empresas a cumprir metas específicas, v.g., de manter o conteúdo ilegal ou prejudicial abaixo de certos limites previamente acordados; é preciso incentivar as plataformas de tecnologia a atuar de modo que possam equilibrar responsabilmente valores como segurança, privacidade e liberdade de expressão, e isso pode ser alcançado por meio do estabelecimento de estruturas domésticas da responsabilidade das empresas;

(vii) uma regulamentação que crie regras levando em consideração a complexidade do mundo digital, notadamente que reconheçam as preferências dos usuários, que possam ser aplicadas em escala, que permitam a flexibilidade entre idiomas, tendências e contextos, e que considere a variação existente entre os serviços de Internet ao redor do globo;

(viii) a edição de uma regulamentação nacional que respeite a natureza da Internet, ou seja, a sua escala global e transcultural, bem como a ampla gama de valores presentes nas comunicações transfronteiriças;

(ix) a criação legal de órgãos reguladores estatais capazes de desenvolver pesquisas e estudos voltados à compreensão das capacidades e limitações tecnológicas na moderação de conteúdo e a permissão para que as plataformas digitais tenham flexibilidade para inovar e, inclusive contribuir em com tais estudos;

(x) considerar na atividade regulatória a gravidade que enseja a remoção do conteúdo ilegal e a prevalência do conteúdo prejudicial, seu status perante as leis, bem como os esforços já em andamento para lidar com o conteúdo.

Por fim, vale destacar a opinião do Grupo de Trabalho Transatlântico de Alto Nível Sobre Moderação de Conteúdo Online e Liberdade de Expressão<sup>32</sup>, que considera a transparência, os mecanismos alternativos de resolução de litígios e as limitações dos algoritmos como alguns dos melhores caminhos a seguir no tema da regulação e moderação de conteúdo online (TWG, 2019).

## 5. Conclusão

Pela presente pesquisa se pôde aferir que a regulamentação da Internet e das plataformas digitais, notadamente das mídias sociais, e especialmente no tocante à moderação de conteúdo online, constitui um cenário de intensas e dramáticas batalhas que têm sido travadas na contemporaneidade entre os Estados nacionais e as Big Techs, cujos resultados têm sido altamente maléficos para os direitos humanos.

Aferiu-se que a ânsia regulatória dos Estados, reveladora da intenção de estender a sua soberania sobre a Internet e controlar a vida digital, tem acarretado sérios prejuízos a direitos que hoje se encontram universalmente consagrados, notadamente à liberdade de expressão, ao direito à informação e à privacidade de um sem-número de usuários que se encontram espalhados ao redor do globo.

Verificou-se que a digitalização da vida nas diversas sociedades contemporâneas tem trazido consigo novas questões e problemas que são de difícil resolução e exigem um esforço hercúleo para o seu enfrentamento. Se de um lado é de absoluta importância combater o mal, evidenciado na Internet por atuações extremadas de violência, pelo ódio, racismo, intolerância, disseminação de notícias falsas e enganação online, desinformação, crimes virtuais e outras formas pelas quais tem se

---

<sup>32</sup> Um projeto do *Annenberg Public Policy Center*, da Universidade da Pensilvânia, em parceria com *The Annenberg Foundation Trust*, em Sunnylands e o *Institute for Information Law*, da Universidade de Amsterdam. O grupo e suas atividades podem ser consultadas aqui: <https://www.ivir.nl/twg/>.

manifestado no mundo virtual; de outro, tornou-se igualmente imperiosa a proteção e a tutela efetiva dos direitos humanos não apenas da parte real, mas também da parte virtual da vida dos seres humanos, enquanto luta pela prevalência do bem no mundo digital.

Nesse contexto, o trabalho demonstrou que a liberdade de expressão e o direito à informação devem ser tutelados em toda a sua amplitude também na esfera digital, aferindo-se que a dupla dimensão desses direitos (substantiva e instrumental) é de observância obrigatória no âmbito da Internet, o que reclama uma atuação dos Estados e das empresas proprietárias de plataformas digitais que operam online, comprometida com os parâmetros internacionais de proteção dos direitos humanos.

Pela análise das várias e expansivas legislações nacionais que visam regulamentar a moderação de conteúdo nas plataformas de mídia social se verificou a existência de um padrão normativo estatal que invariavelmente importa em diversas ofensas aos direitos protegidos. E como forma de enfrentamento desse estado de coisas digital contemporâneo, foram apresentadas diversas ações que, uma vez implementadas, podem contribuir significativamente para a construção de uma Internet mais livre, segura, democrática e de respeito aos direitos humanos.

## Referências

- ARTICLE 19. **Legal Analysis**. Indonesia: Regulation of the Minister of Communication and Informatics Number 5 of 2020 on Private Electronic System Operators (Ministerial Regulation 5). Disponível em: <https://bit.ly/3yB5OpI>. Acesso em 10 dez. 2021.
- BALBINO, Daniel Abrantkoski; PEREIRA, Luciano Meneguetti. A liberdade de expressão e o discurso de ódio: Um estudo sobre a necessidade da educação em direitos humanos. *In: Educação em direitos humanos: construindo uma cultura de respeito aos direitos humanos*. São Paulo: Boreal, 2018, p. 14-35.
- BASTOS, Celso Ribeiro. **Curso de direito constitucional**. São Paulo: Celso Bastos Editora, 2002.
- BICKERT, Monika. **Charting a way forward on online content regulation**. Disponível em: <https://bit.ly/3mHKWbJ>. Acesso em 10 dez. 2021.
- BOBBIO, Norberto. **A era dos direitos**. Trad. Carlos Nelson Coutinho. Rio de Janeiro: Campus, 1992.
- BRASIL. **Medida Provisória n. 1.068, de 6 de setembro de 2021**. Disponível em: <https://bit.ly/3Fcqk2e>. Acesso em 10 dez. 2021.

EUR-LEX. **Proposal for a regulation of the European Parliament and of the Council on a single market for digital services (Digital Services Act) and amending Directive 2000/31/EC**. Disponível em: <https://bit.ly/3pfk30e>. Acesso em 10 dez. 2021.

FRANÇA. Conseil Constitutionnel. **Décision n° 2020-801 DC du 18 juin 2020**. Disponível em: <https://bit.ly/3yDmih4>. Acesso em 10 dez. 2021a.

FRANÇA. **Loi n. 2020-766 du 24 juin 2020 visant à lutter contre les contenus haineux sur internet**. Disponível em: <https://bit.ly/30DTybh>. Acesso em 10 dez. 2021b.

FREEDOM HOUSE. **Freedom on the Net 2021**. Disponível em: <https://bit.ly/3snPF5E>. Acesso em 10 dez. 2021.

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

HUMAN RIGHTS WATCH. **Russia: Growing Internet Isolation, Control, Censorship. Authorities Regulate Infrastructure, Block Content**. Disponível em: <https://bit.ly/32caRke>. Acesso em 10 dez. 2021.

IGI GLOBAL. **What is Digital Platform**. Disponível em: <https://bit.ly/3q9mH6P>. Acesso em 10 dez. 2021.

ÍNDIA. **The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021**. Disponível em: <https://bit.ly/3yCB3AG>. Acesso em 10 dez. 2021.

INDONÉSIA. **Peraturan Menteri Komunikasi dan Informatika Nomor 5**. Disponível em: <https://bit.ly/3F2NUOZ>. Acesso em 10 dez. 2021.

LAIDLAW, Emily B. **Regulating speech in cyberspace: gatekeepers, human rights and corporate responsibility**. Cambridge: Cambridge University Press, 2015. Epub.

MCGOLDRICK, Dominic. **The Limits of Freedom of Expression on Facebook and Social Networking Sites: A UK Perspective**. In: **Human Rights Law Review**, Volume 13, Issue 1, March 2013, p. 125–151. Disponível em: <https://doi.org/10.1093/hrlr/ngt005>. Acesso em 10 dez. 2021.

ÖĞRET, Özgür. **Turkish social media law consolidates news censorship under ‘right to be forgotten’**. Disponível em: <https://bit.ly/3p5WWF1>. Acesso em 10 dez. 2021.

OHCHR. OFFICE OF THE HIGH COMMISSIONER FOR HUMAN RIGHTS. **Moderating online content: fighting harm or silencing dissent?** Disponível em: <https://bit.ly/32gDAEn>. Acesso em 10 dez. 2021.

RAMOS, André de Carvalho; BUCCI, Daniela. **Direitos humanos, mídias sociais e democracia perspectivas de direito nacional e internacional**. In: NUNES, César Augusto R. et. al. (Orgs.). **Temas de direitos humanos do VI CIDH Coimbra 2021**. Campinas; Jundiaí: Brasília; Edições Brasil, 2021, p. 193-205.

REINO UNIDO. **Draft Online Safety Bill**. Disponível em: <https://bit.ly/3yEeOKE>. Acesso em 10 dez. 2021.

ROBERTS, Sarah T. **Behind the screen: content moderation in the shadows of social media**. New Haven and London: Yale University Press, 2019.

RÚSSIA. **Федеральный закон 01.05.2019 № 90-ФЗ “О внесении изменений в Федеральный закон “О связи” и Федеральный закон “Об информации, информационных технологиях и о защите информации”**. Disponível em: <https://bit.ly/32isUp2>. Acesso em 10 dez. 2021c.

RÚSSIA. **Федеральный закон 24.02.2021 № 19-ФЗ “О внесении изменений в Кодекс Российской Федерации об административных правонарушениях”**. Disponível em: <https://bit.ly/3Ffn7Px>. Acesso em 10 dez. 2021d.

- RÚSSIA. **Федеральный закон** 27.06.2018 № 155-ФЗ. “О внесении изменений в Кодекс Российской Федерации об административных правонарушениях”. Disponível em: <https://bit.ly/3yGqdd5>. Acesso em 10 dez. 2021a.
- RÚSSIA. **Федеральный закон** 27.07.2006 № 149-ФЗ. “Об информации, информационных технологиях и о защите информации”. Disponível em: <https://bit.ly/32e3Pfo>. Acesso em 10 dez. 2021b.
- SARLET, Ingo Wolfgang; MARINONI, Luiz Guilherme; MITIDIERO, Daniel. **Curso de direito constitucional**. 5. ed. São Paulo: Saraiva, 2016.
- SARLET, Ingo Wolfgang; MARINONI, Luiz Guilherme; MITIDIERO, Daniel. **Curso de direito constitucional**. 10. ed. São Paulo: Saraiva, 2021. Epub.
- SUPREMO TRIBUNAL FEDERAL. **ADI 6991 MC/DF**. Medida Cautelar na Ação Direta de Inconstitucionalidade. Relator(a): Min. Rosa Weber. Julgamento: 14/09/2021. Publicação: 16/09/2021. Disponível em: <https://bit.ly/32jF8NV>. Acesso em 10 dez. 2021
- TAVARES, André Ramos. **Curso de direito constitucional**. 18. ed. São Paulo, Saraiva, 2020.
- TRANSATLANTIC WORKING GROUP (TWG). **Co-Chairs Report No. 3. Working Papers**. The Bellagio Session, November 13-16, 2019, hosted by the Rockefeller Foundation Bellagio Center, Italy. Disponível em: <https://bit.ly/3J0BkSR>. Acesso em 10 dez. 2021.
- TRIBE, Laurence H. **American constitutional law**. 2. ed. Mineola/NY: The Foundation Press, 1988.
- TURQUIA. **Lei de Regulamentação de Transmissões na Internet e Combate aos Crimes Cometidos por meio Destas Transmissões**. Lei n. 5.651/2007. Disponível em: <https://bit.ly/3GSNsDp>. Acesso em 10 dez. 2021a.
- TURQUIA. **Lei para Alterar a Lei de Regulamentação de Transmissões na Internet e Combate aos Crimes Cometidos por meio Destas Transmissões**. Lei n. 7.253/2020. Disponível em: <https://bit.ly/3mbKAtC>. Acesso em 10 dez. 2021b.

Artigo recebido em: 26/03/2022.

Aceito para publicação em: 09/11/2022