Revista Jurídica vol. 04, nº. 66, Curitiba, 2021. pp. 163 - 196

DESINFORMAÇÃO, PÓS-VERDADE E DEMOCRACIA: UMA ANÁLISE NO CONTEXTO DO ESTADO DEMOCRÁTICO DE DIREITO

DISINFORMATION, POST-TRUTH AND DEMOCRACY: AN ANALYSIS IN THE CONTEXT OF THE DEMOCRATIC STATE OF LAW

FLÁVIO PANSIERI

Pós-Doutor em Direito pela Universidade de São Paulo – USP. Doutor em Direito pela Universidade Federal de Santa Catarina – UFSC. Mestre em Direito pela USP. Professor Adjunto da Pontifícia Universidade Católica do Paraná – PUCPR.

MARIELLA KRAUS

Mestre em Direito pela Universidade Federal de Santa Catarina – UFSC. Pós-graduada em Direito Constitucional pela Academia Brasileira de Direito Constitucional – ABDConst. Graduada em Direito pela Universidade Regional de Blumenau – FURB.

STEFANO ÁVILA PAVAN

Pós-graduado em Direito Constitucional pela Academia Brasileira de Direito Constitucional – ABDConst e em Direito Digital pelo Instituto de Tecnologia e Sociedade do Rio de Janeiro – ITS/RIO UERJ. Graduado em Direito pela Pontifícia Universidade Católica do Paraná – PUCPR.

RESUMO

Objetivo: Com as novas formas de comunicação possibilitadas com a internet, a desinformação se apresenta como uma nova realidade cada vez mais presente na

[Received/Recebido: Maio 23, 2021; Accepted/Aceito: Julho 23, 2021]

sociedade de diversos países, constatando-se um período de pós-verdade que reflete nos contextos democráticos de diversas maneiras. Assim, o presente artigo tem o objetivo de traçar um panorama a respeito da forma como a desinformação nos tempos atuais traz reflexos no Estado Democrático de Direito, notadamente à moderação de conteúdo em plataformas digitais.

Metodologia: Por meio do método de abordagem dedutivo e técnica de pesquisa bibliográfica.

Resultados: Conclui-se que é preciso traçar estratégias jurídicas para conviver nesta nova realidade posta e, ainda, que a imprecisão do termo *fake news* não traduz o problema da pós-verdade.

Contribuições: Refletir sobre as formas de moderação dos conteúdos de desinformação no contexto de pós-verdade e os reflexos disto no Estado Democrático de Direito.

Palavras-chave: desinformação; *fake news*; pós-verdade; democracia.

ABSTRACT

Objective: *With the new forms of communication made possible with the internet, misinformation presents itself as a new reality that is increasingly present in society in different countries, revealing a period of post-truth that reflects on democratic contexts in different ways. Thus, this article aims to draw an overview of how misinformation in current times, reflections on the Democratic Rule of Law, notably the moderation of content on digital platforms.*

Methodology: *Through the method of deductive approach and bibliographic research technique.*

Results: *It is concluded that it is necessary to trace legal strategies to live in this new reality, and also that the imprecision of the term fake news does not reflect the problem of post-truth.*

Contributions: *To reflect on the ways of moderating disinformation content in the context of post-truth and its reflexes on the Democratic Rule of Law.*

Keywords: *disinformation; fake news; post-truth; democracy.*

1 INTRODUÇÃO

O fenômeno da desinformação apresenta diversas nuances, particularmente relacionadas ao desenvolvimento da internet, a qual possibilita a qualquer indivíduo, desde que incluído digitalmente, consumir e produzir conteúdo, recebendo-o e divulgando-o para qualquer lugar do planeta, sem necessitar de intermediários.

Por sua própria infraestrutura, percebe-se o potencial democrático da internet, através de suas plataformas de trocas de mensagens, redes sociais, fóruns de discussão e outros acessos facilitados. O desafio posto às democracias contemporâneas com o fenômeno da desinformação – ao qual se convencionou atribuir o nome *fake news* – pode ser sintetizado como a "entropia informativa, a proliferação de visões por vezes distorcidas da realidade, o excesso de informação que circula e é partilhada nas redes sociais, muitas vezes falsa e difundida com o propósito de enganar ou confundir as mentes dos sujeitos" (PRIOR, 2019, p. 89) e, com isso, coloca novos desafios não apenas aos agentes do sistema midiático, como também à própria população.

A divulgação e difusão massivas de informações falsas, descontextualizadas, e enviesadas maliciosamente enseja a necessidade de se refletir acerca dos impactos da desinformação no âmbito político de Estados democráticos e a sua relação com a regulação do ambiente da internet, particularmente com relação ao período eleitoral, que por si só é tendencialmente mais conturbado.

Sendo assim, o presente artigo tem por objetivo, a partir de abordagem dedutiva e método bibliográfico, trazer alguns pressupostos para a compreensão do impacto da desinformação na democracia, especialmente na moderação do meio digital e estudar o tema.

De início, a proposta é contextualizar a chamada pós-verdade enquanto fenômeno social que impacta diretamente a convivência democrática e fornece o pano de fundo à difusão da desinformação. Na sequência, esclarecer acerca do equívoco em se tratar da desinformação apenas a partir do termo *fake news,* o qual é

insuficiente para compreender a dimensão do problema. E, por fim, será feita uma análise a respeito da moderação do conteúdo de desinformação e, ainda, dos reflexos jurídicos da desinformação e alguns argumentos ligados aos efeitos do combate à desinformação na democracia.

2 PÓS-VERDADE E SEUS FATORES DE INFLUÊNCIA

A criação e reprodução de boatos, mentiras e todo gênero de informações falsas ou imprecisas certamente não é um fenômeno recente, uma vez que consiste em característica da própria evolução da fala e da comunicação (KEYES, 2004, p. 19).

Neste sentido, é de se observar que a capacidade de iludir, enganar, esteve presente desde os primeiros grupos organizados de seres humanos, sendo atributos essenciais em momentos de caça, fuga de predadores, confusão de inimigos (KEYES, 2004, p. 20). No entanto, a persistência de mitos pode contribuir para uma degradação da verdade, onde acabam por valer mais as narrativas de cada pequeno círculo social e se degradarem as relações de confiança comunitárias, importantes para a convivência social (KEYES, 2004, p. 83).

O contexto em que se desenvolveu a chamada “pós-verdade” ocorre justamente desta forma, considerando a difusão que vêm tomando movimentos de questionamento a verdades antes tidas por incontestáveis. Segundo o *Oxford Dictionaries*, que elencou “pós-verdade” como palavra do ano em 2016, o termo significa circunstâncias em que os fatos objetivos são menos influentes em formar a opinião pública do que os apelos à emoção e à crença pessoal (WASHINGTON POST, 2016).

A escolha da palavra se deu em razão de um salto de 2.000% (dois mil por cento) no ano de 2015 nas buscas pelo termo nos meses que em que ocorreram a votação do referendo no Reino Unido pela saída da União Europeia (comumente

[Received/Recebido: Maio 23, 2021; Accepted/Aceito: Julho 23, 2021]

conhecido por Brexit) e a vitória de Donald Trump nas previas às eleições americanas do Partido Republicano (MCINTYRE, 2018, p. 1).

A pós-verdade sinaliza a passagem a um estado de legitimação de argumentos com base em seu apelo emocional e valores simbólicos e subjetivos, em oposição à verdade impessoal, o que demonstra a importância de como os indivíduos se conectam emocionalmente com a informação (KALPOKAS, 2019, p. 2).

Interessante observar que a pós-verdade não se trata simplesmente de mentiras, sendo um fenômeno mais complexo. O autor Ralph Keyes, em uma das primeiras obras sobre o tema, explica que na era da pós-verdade já não existem verdades nem mentiras, mas, sim, uma outra categoria de afirmações que não são exatamente verdadeiras, mas também não podem ser enquadradas como mentiras (KEYES, 2004, p. 15).

Foi dentro do contexto da pós-verdade que a campanha pelo Brexit, além do crescimento de campanhas de desinformação utilizadas por políticos em países como Hungria, Russia e Turquia (MCINTYRE, 2018, p. 5-6) que as *fake news* se apresentaram enquanto problema às democracias contemporâneas, das quais a brasileira não está à parte.

2.1 CRISE POLÍTICA E REPRESENTATIVIDADE DEMOCRÁTICA

Outro aspecto relacionado à pós-verdade é a anunciada crise de confiança nas instituições democráticas, a qual fornece a base social da pós verdade (D'ANCONA, 2018, p. 42). Dentre as diferentes razões para a chegada desta crise, pode-se destacar inicialmente o descontentamento evidenciado em diversos países com os sistemas de partidos políticos e com os resultados entregues pelos representantes.

Quanto à representação política, o autor Manuel Castells a desconstrói de início, declarando a democracia como construída em torno das relações de poder que a fundaram, e da sua adaptação à evolução destas relações. A representação

[Received/Recebido: Maio 23, 2021; Accepted/Aceito: Julho 23, 2021]

política não existiria de fato, senão enquanto os cidadãos pensem que estão sendo representados (CASTELLS, 2018, p. 10).

Nesse sentido, entender a crise de legitimidade institucional e da representação política como uma das bases da pós-verdade passa também por compreender o vínculo subjetivo entre o que os cidadãos pensam e querem e as ações daqueles que elegeram, ou seja, quando estas deixam de se corresponder, é gerado o sentimento majoritário de que os atores do sistema político não representam a população (CASTELLS, 2018, p. 10).

Um ponto a ser destacado é a ausência de uma real possibilidade de contestação pela população das decisões políticas que acabam por afetá-la (MOUNK, 2018, p. 66). Em diversos países com diferentes contextos políticos, estes elementos favoreceram a tomada de poder por líderes populistas – seja à esquerda e à direita, no espectro político – que contestam a legitimidade das instituições, a liberdade de imprensa e dirigem ataques a todos os que questionem a veracidade de suas afirmações, uma vez que estes se postam como representantes da voz do povo[1] (MOUNK, 2018, p. 42).

Ainda, um dos pontos mais relevantes desses movimentos de protesto e contestação às instituições a se destacar é que os mesmos não desconsideram a importância da democracia. Antes, buscam afirmar a possibilidade de uma real democracia (MOUNK, 2018, p. 52).

Assim, cabe ressaltar o quanto as mudanças tecnológicas das últimas décadas têm influenciado diretamente na percepção comum da política democrática e de seus atores, uma vez que há, segundo o autor Helder Prior:

> [...] uma descentralização dos processos comunicacionais, permitindo que os atores políticos dispensem a mediação dos agentes do campo do jornalismo na interação com os cidadãos, erigindo um novo modelo comunicacional assente nos processos horizontais da comunicação em rede. (PRIOR, 2019, p. 92)

[1] Nas palavras do autor: "A promessa de dar expressão à voz não adulterada do povo é a característica central do populismo". Tradução livre.

A partir desta descentralização, torna-se possível o contato direto entre os representantes (ou postulantes a cargos de representação) e seus representados (ou eleitores), sem freios e contrapesos feitos através dos meios tradicionais de comunicação (PRIOR, 2019, p. 92) e do sistema político como um todo.

Desta maneira, as informações compartilhadas entre grupos de interesse políticos semelhantes, e com forte descrença nos meios políticos tradicionais, tendem a circular sem contestação, sendo esse um campo fértil ao desenvolvimento da pós-verdade.

2.2 MIDIATIZAÇÃO E A ERA DA EXPERIÊNCIA

A midiatização, a título de exemplo, pode se dar na expansão crescente do consumo de mídias digitais, além da onipresença de telas (sejam computadores, celulares e televisão) tornando o entretenimento, a política e a comunicação indistintos entre si na maior parte do tempo (KALPOKAS, 2019, p. 54).

Tal expansão é atestável, segundo dados produzidos na Pesquisa Brasileira de Mídia 2016, através da qual se avaliou, somente quanto ao uso da internet, o tempo médio de acesso diário dos brasileiros, considerando-se tanto o meio de semana quanto o final de semana, fica acima de quatro horas e trinta minutos (BRASIL, 2016).

Em sentido semelhante, relatório produzido em 2017 no âmbito da Conferência das Nações Unidas sobre Comércio e Desenvolvimento (UNCTAD), evidenciou que, em 2015, o Brasil tinha 120 milhões de usuários de internet, ficando atrás apenas dos Estados Unidos, Índia e China (NAÇÕES UNIDAS BRASIL, 2017).

Por fim, relevante enquanto demonstração da influência cada vez mais significativa das mídias digitais no dia a dia, cabe mencionar a pesquisa realizada pelo Instituto Reuters, realizada em parceria com a Universidade de Oxford, em que foram aplicados questionários no início do ano de 2018 em diversos países para se entender como se consomem notícias. Quanto ao Brasil, a pesquisa apontou que

66% (sessenta e seis por cento) dos brasileiros consomem notícias através de redes sociais (NEWMAN, 2018).

As pesquisas destacadas fornecem indícios também acerca de outro aspecto ligado à midiatização, que vem acarretando mudanças não só sociais, mas também relacionadas à construção do indivíduo, cuja existência no mundo é cada vez mais atrelada a uma representação estabelecida digitalmente – a partir de como o indivíduo quer ser e se mostrar nas redes – e não mais somente à sua identidade física (KALPOKAS, 2019, p. 59).

A percepção daquilo que ocorre, afirma o autor Prior (2019, p. 86), passa a ser condicionada àquilo que os meios de comunicação possibilitam ser experimentado como real, como o que de fato está ocorrendo. Ainda, em outro aspecto ligado à midiatização, Kalpokas (2019, p. 60) aponta que a identidade individual se constrói sob essa nova ótica da maciça presença das mídias sócias através do acúmulo de dados pessoais de acesso e comportamento dentro dos sites, e que são controladas por empresas privadas que os utilizam como parte de seus modelos de negócios.

Considerando a intangibilidade das comunicações através das redes, as interações entre os indivíduos – que migram cada vez mais ao meio digital – se fazem através de trocas afetivas, ou seja, do engajamento emocional entre pessoas distintas sobre interesses em comum. A partir dessa premissa, criam-se discursivamente *nós* ligados entre si por suas semelhanças e afetos, em oposição ao *outro*, discursivamente ininteligível (KALPOKAS, 2019, p. 64-65).

No mesmo sentido afirma o autor Matthew D'Ancona, ressaltando de forma negativa a ligação emocional entre os indivíduos como fator de criação de vínculos comunitários, em oposição à verdade enquanto valor em si mesmo (D'ANCONA, 2018, p. 38), no que o complementa o autor Cass Sunstein:

> (...) As emoções podem obstruir o caminho da busca pela verdade. As pessoas não processam as informações com neutralidade. Suas pressuposições afetam suas reações. *Assimilação tendenciosa* se refere ao fato de que as pessoas assimilam novas informações de maneira

> tendenciosa; aqueles que acreditam em boatos falsos não abrem mão de suas crenças com facilidade, especialmente quando têm forte envolvimento emocional com elas. Pode ser muito difícil mudar o que as pessoas pensam, mesmo que se lhe apresentem os fatos. (SUSTEIN, 2010, p. 12)

A esse respeito, Kalpokas trata das redes sociais, as quais o autor entende como particularmente eficazes em criar contra esferas públicas para a agregação de grupos que por quaisquer motivos, se sintam marginalizados ou ignorados pela tradicional esfera política (KALPOKAS, 2019, p. 102).

Ressalta o autor que as redes sociais talvez sejam o ambiente em que a midiatização se demonstra da forma mais evidente, uma vez que neste meio o sucesso na disseminação de determinada afirmação de verdade dependerá do quanto esta atenda aos desejos e necessidades emocionais dos públicos relevantes (KALPOKAS, 2019, p. 102), mesmo se não seja confirmada por fatos verificáveis.

3 IMPRECISÃO DO TERMO *FAKE NEWS*

Uma simples tradução literal de *fake news* trataria o fenômeno como notícias falsas. Sendo assim, a problemática envolvendo as *fake news* se resolveria com o emprego de práticas destinadas a coibir a difusão de notícias falsas, particularmente aquelas compartilhadas pela internet.

A utilização da tradução literal do termo *fake news* faz presumir que o problema está unicamente ligado a dois pontos: **(i)** notícias, ou seja, relatos caracteristicamente jornalísticos e que se assemelhem ao que tradicionalmente se conhece por uma notícia; **(ii)** falsas, entendendo-se por falso para os fins do presente trabalho aquilo que é contrário ao verdadeiro, ao que realmente ocorre.

No entanto, a presunção destacada não considera algumas nuances do problema, como por exemplo, a problemática de se divulgarem informações reais (ou verdadeiras), porém maliciosamente utilizadas fora de contexto, a divulgação de

informações sem intenção de causar dano, mesmo que falsas, além de uma série de outras particularidades.

Para melhor elucidar a complexidade do tema, há de se destacar a distinção feita pela jornalista americana Claire Wardle (2017), em que a mesma separa entre sete categorias de conteúdo problemático na internet e que estão relacionadas com o tema das *fake news*, quais sejam: **(1)** sátira ou paródia – informação divulgada sem intenção de causar danos, porém com potencial de confundir; **(2)** conexões falsas – ocorrem quando o conteúdo de determinada notícia não é corroborado por sua manchete, imagens ou legendas; **(3)** conteúdo enganoso – quando uma informação é utilizada de forma enganosa contra determinado assunto ou indivíduo; **(4)** contexto falso – quando informações verdadeiras são retiradas de contexto, empregando-lhes outro sentido distinto do original; **(5)** conteúdo impostor – quando fontes verdadeiras e confiáveis tem seus nomes utilizados para o compartilhamento de afirmações que não fizeram; **(6)** conteúdo manipulado – quando imagens ou informações verdadeiras são manipuladas com o intuito de enganar; **(7)** conteúdo fabricado – quando um conteúdo é completamente falso, criado com o intuito de causar danos e enganar o público.

A partir das distinções feitas pela jornalista é de fácil constatação que a simples tradução do termo *fake news* não elucida o debate, especialmente ao se buscarem formas de regulação adequada para o problema, particularmente com o interesse de se preservarem valores democráticos como a liberdade jornalística e o direito fundamental à liberdade de expressão.

Além destas definições, os autores Mendes, Doneda e Bachur (2018, p. 143) mencionam relatório produzido pelo Conselho da Europa acerca do fenômeno “desordem informacional”. O relatório intitulado “Information Disorder: Toward an interdisciplinary framework for research and policymaking” parte do pressuposto que muito do discurso envolvendo o fenômeno das *fake news* mistura três noções: misinformação, desinformação e malinformação (WARDLE; DERAKHSHAN, 2018, p. 20).

A desinformação seria aquela informação falsa criada deliberadamente com o intuito de prejudicar determinado grupo social, indivíduo, organização ou um mesmo um país. Já a misinformação seria a informação falsa, porém criada sem a intenção de gerar qualquer tipo de prejuízo. Por fim, a malinformação, que mesmo possuindo bases reais, se trata de informação utilizada para infligir danos a pessoas e organizações (WARDLE; DERAKHSHAN, 2018, p. 20).

Apresentadas as nuances e complexidades na definição de *fake news,* restam claros os riscos que esta carrega à democracia, como a integridade dos procedimentos eleitorais, e mesmo a valores que conformam políticas públicas em diferentes setores (COMISSÃO EUROPEIA, 2018, p. 7), uma vez que o público formaria sua opinião cada vez mais sobre assuntos de interesse compartilhado com base em informações falsas ou enganosas.

Sobre este ponto se coloca a presente pesquisa, buscando-se compreender, a partir do enorme potencial de difusão da desinformação através do uso de diferentes tecnologias, a moderação do conteúdo falso disseminado e, ainda, quais seriam os impactos disso no Estado Democrático de Direito.

4 DESINFORMAÇÃO E O DIREITO NO BRASIL

A contextualização da pós-verdade auxilia a compreender a atual conjuntura político-democrática de contestação às instituições e mudança no comportamento dos indivíduos a partir da presença cada vez mais intensada tecnologia e das redes sociais no diálogo político, de modo que os veículos midiáticos tradicionais perdem seu papel de protagonismo na formação de opinião e passam a ser vistos como enviesados, não refletindo uma visão objetiva dos fatos[2].

[2] O autor Lee Mcintyre trata de forma detalhada a relação entre a pós-verdade e o fenômeno das *fake news*, destacando que a noção de uma mídia imparcial e objetiva é na realidade algo bastante recente. Ver MCINTYRE, 2018. p. 89-122.

[Received/Recebido: Maio 23, 2021; Accepted/Aceito: Julho 23, 2021]

A dinâmica normalmente associada às *fake news* é a divulgação de informações mentirosas (ou fraudulentas), destinadas a lesar a imagem de determinado ente ou pessoa, além de criar confusão no público. O termo faz crer, assim, que caminha no sentido de "englobar informações que, muitas vezes, aparentam ser fruto de apuração jornalística, mas que são dotadas intencionalmente de conteúdos falsos ou manipulados, visando-se a obter como regra benefício econômico e/ou político" (TEFFÉ; SOUZA, 2019, p. 526).

O conteúdo falso ou manipulado, criado com intuito de obtenção de benefícios ilegítimos e que lese a imagem de terceiros ou de instituições democráticas, deveria, então, sob essa lógica, ser removido de onde estivesse publicado, priorizando-se os conteúdos verdadeiros e que não atingissem negativamente a imagem de quem quer que fosse.

O processo de desinformação passa, atualmente, pela desconstrução da credibilidade dos veículos oficiais de comunicação, públicos ou privados. Isto faz com que se abra um espaço para que outros agentes da sociedade tomem esta legitimidade para si, ou seja, para além deste fato da conduta humana de aceitar uma desinformação da pós-verdade – que se adeque ao seu ponto de vista opinativo – entende que deve ser replicado, independentemente da verificação ou constatação da veracidade daquela informação.

A profusão de projetos de lei envolvendo *fake news* atualmente (GRIGORI, 2018), propondo seja a criminalização da criação e/ou a divulgação de notícias falsas, ou a imposição de multas pelos mesmos motivos demonstra o receio criado pelo fenômeno.

Em essência, o temor causado pelas *fake news* se interliga diretamente com a mudança nos meios de comunicação – como ressaltado quanto ao fenômeno da midiatização – e à perda de confiança nos canais midiáticos tradicionais, em que é visto um incremento cada vez maior no poder dos indivíduos (quando inseridos digitalmente) em se tornarem produtores e editores de conteúdo, participando ativamente na criação e difusão de informação:

[Received/Recebido: Maio 23, 2021; Accepted/Aceito: Julho 23, 2021]

> (...) essa explosão de novos canais de interação, especialmente através das mídias sociais, tem criado novas esferas de comunicação social imunes a qualquer cultura ou *ethos* forte, similares à ética de imprensa ou a ética jornalística comumente encontrada na maior parte dos países democráticos. Os paradigmas jurídicos, morais e éticos relativos à comunicação pública e a mídia se desenvolveram e foram compartilhadas lentamente durante décadas em que a prática jurídica, a educação jornalística institucionalizada (especialmente pelas faculdades de jornalismo) e debates públicos ocorreram. Tais práticas fixaram padrões de civilidade, compromisso com a verdade e responsabilidade que de muitas maneiras ainda regulam o *ethos* profissional na mídia dos países democráticos (LIPPMANN, 1997, p. 203). Não há uma "cultura do compromisso com a verdade" similar já em funcionamento nas mídias sociais. (MACEDO JUNIOR, 2018, p. 133)

Vê-se, então, que a questão e os possíveis problemas gerados pela disseminação de *fake news* não é resolvida de forma tão simplista, uma vez que a simples remoção do conteúdo supostamente fraudulento e criação de novos tipos penais, além da previsão de multas para a punição dos eventuais responsáveis não abrange a complexidade do problema. A difusão da informação através das redes sociais, como mencionado por Macedo Júnior, acontece essencialmente de modo alheio aos paradigmas construídos sobre a comunicação pública.

Nesse sentido, o fenômeno das *fake news* em sua relação com o Direito enseja reflexões acerca dos impactos das mesmas em diversos âmbitos, passando pela responsabilidade civil, penal, administrativa tanto dos criadores dos conteúdos problemáticos como dos próprios provedores de acesso à internet, os quais estão sujeitos à disciplina do Marco Civil da Internet (BRASIL, 2014).

Ainda, trazendo a questão para o âmbito constitucional e a proteção de direitos fundamentais, há de se pensar sobre as possíveis tensões envolvendo ordens de remoção de conteúdo em páginas da internet, especialmente porque determinado conteúdo tratado por *fake news* pode simplesmente refletir uma opinião pessoal ou mesmo se tratar de fato ainda não comprovado, mas sob investigação.

A Constituição Federal de 1988 traz em seu artigo 5º, o inciso IV que dispõe que "é livre a manifestação do pensamento, sendo vedado o anonimato" e, em seu inciso IX, que "é livre a expressão da atividade intelectual, artística, científica e de comunicação, independentemente de censura ou licença", ressaltando a liberdade

[Received/Recebido: Maio 23, 2021; Accepted/Aceito: Julho 23, 2021]

constitucionalmente garantida a todos aqueles sob a guarida do ordenamento jurídico brasileiro de livremente expressarem suas opiniões.

Segue no mesmo sentido a liberdade de imprensa, elencada no artigo 220 da Constituição Federal, o qual menciona expressamente em seu § 2º a vedação à censura:

> **Art. 220**. A manifestação do pensamento, a criação, a expressão e a informação, sob qualquer forma, processo ou veículo não sofrerão qualquer restrição, observado o disposto nesta Constituição. **§ 1º** Nenhuma lei conterá dispositivo que possa constituir embaraço à plena liberdade de informação jornalística em qualquer veículo de comunicação social, observado o disposto no art. 5º, IV, V, X, XIII e XIV. **§ 2º** É vedada toda e qualquer censura de natureza política, ideológica e artística. (BRASIL, 1988)

É possível estabelecer a partir da leitura dos dispositivos constitucionais a robusta proteção concedida à livre expressão do pensamento, sob suas variadas formas, pelo ordenamento jurídico brasileiro e "visam afastar a interferência indesejada de terceiros – Estado e particulares – na livre difusão de um complexo de conteúdos que abrange a informação jornalística" (GIACCHETA, 2018, p. 28).

Qualquer iniciativa que busque enfrentar o fenômeno das *fake news* através da remoção de conteúdo da internet, necessariamente deverá resguardar as liberdades de expressão e informação, constitucionalmente asseguradas. Neste sentido, está a vedação a qualquer forma de filtragem ou controle prévio de conteúdo, seja por parte do Estado ou por particulares, estes representados pelos provedores de aplicações de internet.

A centralidade conferida à liberdade de expressão é refletida pelo Marco Civil da Internet, o qual a entre seus fundamentos: "**Art. 3º** A disciplina do uso da internet no Brasil tem os seguintes princípios: **I -** garantia da liberdade de expressão, comunicação e manifestação de pensamento, nos termos da Constituição Federal; (...)" (BRASIL, 2014).

O estabelecimento da liberdade de expressão entre os princípios regentes do uso da internet no Brasil se reflete na obrigação estabelecida para os provedores de

[Received/Recebido: Maio 23, 2021; Accepted/Aceito: Julho 23, 2021]

aplicações de internet no artigo 19 do Marco Civil, os quais somente podem ser compelidos a remover conteúdo de suas redes após ordem judicial específica para tanto:

> **Art. 19.** Com o intuito de assegurar a liberdade de expressão e impedir a censura, o provedor de aplicações de internet somente poderá ser responsabilizado civilmente por danos decorrentes de conteúdo gerado por terceiros se, após ordem judicial específica, não tomar as providências para, no âmbito e nos limites técnicos do seu serviço e dentro do prazo assinalado, tornar indisponível o conteúdo apontado como infringente, ressalvadas as disposições legais em contrário. (BRASIL, 2014)

A jurisprudência do Superior Tribunal de Justiça está consolidada justamente nesse sentido, destacando que o controle prévio de conteúdo criado por usuários na Internet implicaria em censura. Veja-se, por exemplo, decisão no Recurso Especial nº 1.342.640/SP:

> (...) Conforme entendimento desta Corte, o controle editorial prévio do conteúdo das informações se equipara à quebra do sigilo da correspondência e das comunicações, vedada pelo art. 5º, XII, da CF/88. Não bastasse isso, a avaliação prévia do conteúdo de todas as informações inseridas na web eliminaria um dos maiores atrativos da internet, que é a transmissão de dados em tempo real. Nesse sentido, vale mencionar a lição de Carlos Affonso Pereira de Souza vê "meios tecnológicos para revisar todas as páginas de um provedor", mas ressalva que esse procedimento causaria "uma descomunal perda na eficiência do serviço prestado, quando não vier a impossibilitar a própria disponibilização do serviço" (A responsabilidade civil dos provedores pelos atos de seus usuários na internet. *In* Manual de direito eletrônico e internet. São Paulo: Aduaneiras, 2006, p. 651). No mesmo sentido opina Paulo Nader, que considera inviável impor essa conduta aos provedores, "pois tornaria extremamente complexa a organização de meios para a obtenção dos resultados exigidos, além de criar pequenos órgãos de censura" (Curso de direito civil. vol. VII, 3ª ed. Rio de Janeiro: Forense, 2010, p. 385). Em outras palavras, exigir dos provedores de conteúdo o monitoramento das informações que veiculam traria enorme retrocesso ao mundo virtual, a ponto de inviabilizar serviços que hoje estão amplamente difundidos no cotidiano de milhares de pessoas. A medida, portanto, teria impacto social extremamente negativo. (BRASIL, 2017)

A decisão em comento reflete o paradigma constitucional de vedação à censura e filtragem prévia de conteúdo, não sendo permitido, sob a justificativa de se

dificultar a difusão de conteúdo ofensivo ou ilícito nas redes, a remoção destes sem ordem judicial expressa.

No entanto, cabe destacar que a controvérsia não está de todo resolvida, uma vez que a constitucionalidade da determinação e necessidade de prévia e específica ordem judicial de exclusão de conteúdo para a responsabilização dos provedores, conforme o artigo 19 do Marco Civil da Internet, é objeto de repercussão geral no Supremo Tribunal Federal, no tema nº 987 (BRASIL, 2018), ainda pendente de julgamento quando da redação do presente artigo.

Em sentido semelhante, tramita o Projeto de Lei do Senado de nº 471 de 2018 trazendo, entre outras providências de cunho penal e eleitoral, a previsão de alteração do Marco Civil da Internet para se obrigar os provedores a remover, em até 24 horas, conteúdo denunciado como falso. Assim está redigido o projeto de lei:

> **Art. 18-A.** O provedor de aplicações de internet que disponibilize conteúdo gerado por terceiros adotará medidas efetivas e transparentes para combater a publicação e a disseminação de perfis e notícias falsos. (...) **§ 2º** As denúncias serão tratadas de forma diligente, cabendo ao provedor: **I –** remover ou o bloquear, no prazo de até vinte e quatro horas do recebimento da denúncia, o conteúdo que não atenda à política de uso da aplicação; (...) (BRASIL, 2018)

O mesmo projeto define o que seriam notícias falsas: “IX – notícia falsa: texto não ficcional que, consideradas as características de sua veiculação, possua o potencial de ludibriar o receptor em relação à veracidade do fato.” (BRASIL, 2018). Em que pese a relevante consideração dos danos possíveis pelas *fake news*, também é certo que retirar do Judiciário a baliza para a determinação de remoção de conteúdo é igualmente preocupante.

Ao se atribuir o poder de agentes privados, mediante simples notificação, para determinar e decidir o que deva ser considerado verdadeiro é extremamente perigoso e vai contra as diretrizes de como devem funcionar Estados Democráticos. Da mesma forma, atribuir igual poder ao Estado, através de seus agentes, sem a baliza do Judiciário, pode ser igualmente preocupante, correndo-se o risco de ser

produzido um efeito silenciador dos discursos, de modo que os provedores, temendo eventual responsabilização, passassem a exercer eles próprios a censura do conteúdo publicado em suas plataformas (MACEDO JÚNIOR, 2018, p. 140), prejudicando em última instância o debate democrático e a formação de opinião.

5 MODERAÇÃO DE CONTEÚDO

5.1 DESINFORMAÇÃO E REMOÇÃO DE CONTEÚDO

A disseminação de boatos, *fake news* e *deep fakes* em geral são relacionadas ao fenômeno da desinformação, destacando-se, seja em pesquisas acadêmicas ou em matérias jornalísticas, a rapidez e intensidade com que se espalham nas redes sociais conteúdos falsos ou descontextualizados, onde os indivíduos são direcionados para lugares na rede que apenas reforçam suas opiniões prévias, afastando qualquer contestação e reforçando a polarização política.

No intuito de se conceberem respostas a este dilema das redes, aparentemente sem solução, é sintomática a profusão de projetos de lei envolvendo o combate às *fake news*, ora pela via da criminalização, ora através da imposição de multas aos criadores e disseminadores de conteúdo problemático, entre diversas outras estratégias regulatórias.

A esse respeito, o Projeto de Lei 2.630 de 2020 tem ganhado a maior repercussão entre os existentes (HAJE, 2020), o qual traz diversas obrigações a serem cumpridas pelos provedores de conteúdo para auxiliarem às autoridades no combate a desinformação e conteúdo ilícito em geral na Internet.

Se, por um, lado, é constantemente ressaltada a centralidade dos provedores de aplicação e plataformas de redes sociais na construção da esfera pública no meio digital (LEVY, 2020), os procedimentos adotados por estes nessa construção não são evidenciados da mesma forma, especialmente com relação aos

procedimentos de governança adotados para se evitarem abusos ao mesmo tempo em que se mitigam os impactos de conteúdos nocivos (MARANHÃO, 2020).

Nesse contexto, segundo Hannah Bloch-Wehba, vem ganhando destaque global entre as estratégias ligadas a moderação de conteúdo, incentivos a que os provedores se utilizem de ferramentas destinadas a prevenirem a disseminação de conteúdo ilícito – incluídos aqueles tratados como desinformação – antes até que o mesmo seja visto ou distribuído (BLOCH-WEHBA, 2020).

O lugar ocupado neste debate pelas plataformas de redes sociais e provedores de conteúdo em geral – os *intermediários* – é bem sintetizado por Martin Eifert ao tratar do chamado discurso de ódio e que se aplica de forma bastante semelhante à desinformação:

> Caso os intermediários ajam com demasiada contenção ou moderação, eles se confrontam com a acusação de tolerar discursos de ódio. Caso eles retirem e apaguem conteúdos de forma demasiadamente ampla, então eles se confrontam com a acusação de uma repressão de opiniões e da liberdade de expressão. (...) No geral (...), parece que há, desde o começo, certo viés em favor do apagamento ou retirada. (EIFERT, 2018, p. 80)

Já Cass Sunstein ressalta que, apesar dos riscos de um efeito inibitório à livre expressão pessoal, o que necessariamente enseja atenção adequada a se evitar uma regulação equivocada quanto à disseminação de boatos nas redes, este efeito também poderia auxiliar a enfraquecer as falsidades particularmente danosas e destrutivas (SUNSTEIN, 2010, p. 102), afirma o autor que: "(…) as influências sociais e a assimilação tendenciosa garantem que os boatos falsos se espalhem e ganhem raízes. Uma sociedade sem qualquer efeito inibitório, imposto pelas normas sociais e pelas leis, seria um lugar excepcionalmente torpe".

Se por um lado é possível vislumbrar os efeitos nocivos e particularmente perigosos da desinformação não só para a democracia enquanto ideal, mas efetivamente para vidas reais (LOPEZ, 2016), a estruturação de regimes de responsabilidade envolvendo os provedores de aplicação e plataformas de redes

sociais demanda maior clareza sobre os procedimentos adotadas na moderação de conteúdo, particularmente aqui quando exercido de forma automatizada.

5.2 MODERAÇÃO AUTOMATIZADA DE CONTEÚDO

É inegável o papel representado pelas plataformas e empresas de tecnologia, como Google, Facebook, Twitter, as quais possuem um potencial de interferir na liberdade de expressão e informação maior do que muitos Estados, exercendo imenso poder sobre a arquitetura internet, sendo responsáveis, em grande medida, pelas causas e efeitos da desinformação em massa (MACEDO JUNIOR, 2018, p. 135).

Em pesquisa realizada por Soroush Vosoughi, Deb Roy e Sinan Aral foram identificadas duas frentes de enfrentamento às *fake news*[3], sendo elas: **(i)** iniciativas destinadas a capacitar indivíduos a detectar *fake news* que venham a encontrar; **(ii)** iniciativas buscando mudanças estruturais com o objetivo de evitar a exposição de indivíduos às *fake news* (VOSOUGHI; ROY; ARAL, 2018, p. 359).

A moderação de conteúdo se insere nesta segunda categoria, na medida em que trata da estruturação da esfera pública no meio digital, a partir dos intermediários – aqui as redes sociais ganham papel de destaque – onde estes efetivamente exercem o poder de determinar aquilo que é aceitável ou não de ser dito na internet.

Este poder é exercido em larga escala, como apontam Juliano Maranhão *et al*:

> Apenas no primeiro quarto de 2020, o Facebook removeu ou aplicou medidas sobre cerca de 1,8 bilhão de contas inautênticas e conteúdos envolvendo nudez e atividade sexual, violência, organizações perigosas (terrorismo e ódio), hate speech, drogas e armas de fogo, pornografia infantil, suicídio e auto-mutilação. O último relatório de transparência do Twitter revela suspensão de 244.188 contas por exploração de menores (91% usando tecnologia) e 115.861 contas por conteúdo violento (87% com

[3] Apesar das inúmeras pesquisas de relevo apontando o equívoco na utilização da expressão *fake news*, aqui esta é utilizada de forma intercambiável com a expressão desinformação. Sobre as questões de nomenclatura envolvendo desinformação ver: WARDLE; DERAKHSHAN, 2018. p. 20.

> uso de ferramentas computacionais proprietárias). Segundo o relatório do primeiro trimestre de 2020, moderadores do YouTube removeram manualmente 399.422 vídeos, ao passo que 5.711.586 vídeos foram removidos por ferramentas automatizadas. O relatório de transparência do Google, por sua vez, indica que mais de quatro bilhões de URLs tiveram sua deslistagem requisitada por usuários. (MARANHÃO, 2020)

Não é equivocado pensar na existência de um otimismo com estatísticas dessa natureza, pois fornecem uma impressão tanto à sociedade em geral como às entidades estatais de que os provedores e plataformas estão efetivamente fazendo algo para atacar conteúdos problemáticos.

Ainda, há de se considerar a pressão para que os intermediários desenvolvam tecnologias cada vez mais efetivas, precisas e rápidas no auxílio da aplicação da lei (BLOCH-WEHBA, 2020), especialmente diante de obrigações legais para remoção de conteúdo a partir de simples notificação e em curto prazo, a exemplo da Lei Alemã para a Melhoria da Aplicação da Lei nas Redes Sociais, a qual prevê prazo de vinte e quatro horas para a retirada, em alguns casos (EIFERT, 2018, p. 81).

Adota-se a definição de Grimmelmann quanto à moderação de conteúdo, sendo ela "o conjunto de mecanismos de governança que estruturam a participação em uma plataforma para facilitar cooperação e prevenir abusos" (2015 apud MARANHÃO, 2020), sendo que o uso de sistemas automatizados para se realizar essa governança é historicamente atrelado quando problemas de escala dificultam a moderação realizada somente de forma manual (GORWA; BINNS; KATZENBACH, 2020, p. 3)[4].

Partindo da definição de Grimmelmann, Gorwa, Binns e Katzenbach definem moderação algorítmica como "sistemas que classificam conteúdo gerado por usuários com base em correspondência ou previsão, e que levam a uma decisão e resultado de governança (p. ex. remoção, geobloqueio, derrubada de conta)" (GORWA; BINNS; KATZENBACH, 2020, p. 3). Essencialmente, a moderação

[4] Os autores citam como exemplos nesse sentido as tecnologias de filtros de *spam* e *bots* na plataforma Wikipedia utilizados para impedir vandalismo e monitorar as páginas.

algorítmica se apresenta como complemento ou mesmo alternativa à moderação manual feita por seres humanos, que é insuficiente diante do problema de escala mencionado.

Se as plataformas de redes sociais e provedores de conteúdo em geral deveriam ou não utilizar métodos de moderação algorítmica é um debate de certa forma inócuo, pois estas já o fazem, em maior ou menor medida. Interessa aqui elucidar quais desafios de governança e regulação existem nesse âmbito e algumas iniciativas já existentes para enfrentá-los.

Há, contrariamente a certo otimismo na moderação de conteúdo automatizada, severo risco de que a infraestrutura das redes fique cada vez mais sujeita a censura e controle:

> (...) Vista por essa lente, a virada em direção à automação não é nem um limite no poder das companhias de tecnologia nem uma garantia de que estas agirão de forma mais efetiva ou neutra. Em vez disso, a crescente confiança na automação aumentará o risco de que tanto as plataformas quanto governos experimentem cooptação e captura. (BLOCH-WEHBA, 2020, p. 8)

É necessário diferenciar a moderação automatizada em caso de conteúdo violento, ilícito, como exploração sexual, exposição de menores, etc, da moderação com intuito de enclausurar a capacidade de manifestação que pretende discutir temas polêmicos.

Destaca-se, por isso, que não se pretende discutir a moderação de conteúdo ilícito, mas apenas a moderação que não pode ser considerada como aval ou possiblidade de censura de discursos por parte das redes sociais. Afinal, sem a possibilidade de debate e diálogo no espaço virtual há prejuízo da liberdade de expressão, pois é um espaço privado com características de público.

Compreender a complexidade da estruturação do debate público através da internet e a forma de regulação mais adequada a esse meio exige observar alguns matizes. Em especial, deve-se refletir seriamente que estes espaços, justamente porque, por mais que sejam privados em sua gestão, possuem características de

espaço público de debate. Por conta disso, não se deve acreditar que possa haver uma censura, prévia ou posteriori, por parte destes veículos, afinal não possuem legitimidade para tanto.

Isto porque, aceitar censura prévia pela empresa privada seria enclausurar os discursos aceitáveis e, assim, não se trataria mais de uma "pós-verdade", mas de uma "verdade única", ou seja, aquela aceita e definida pela empresa privada a ser postada no mundo digital. Este é o grande debate que permeia o tema, pois seria uma moderação do privado sobre o público, ou seja, seria conceder uma legitimidade jurisdicional às pessoas jurídicas de direito privados em detrimento da atividade do Estado exercida por meio do Judiciário – ponto basilar de um Estado Democrático de Direito.

Esta automatização de processos que debatem alguns controles não pode ser considerada um espaço livre para a atuação das empresas privadas, pois desta maneira estariam no poder para escolher quais os debates são permitidos. Repisa-se que não se tratam de conteúdos ilícitos, mas, sim, de debates sobre temas complexos e polêmicos, por isso tais medidas não podem ser comparadas.

Por isso, a problemática do controle desta moderação do conteúdo considerado válido perpassa o debate sobre o legitimado para tal controle. Há uma grande diferença entre o discurso que pode ser reconhecido como ilícito – e este, sim, investigado e expurgado do mundo digital – e o discurso que discute sobre a licitude ou ilicitude de algo. Se a sociedade não puder mais discutir o que é lícito ou ilícito, o debate não será mais possível e liberdade não mais haverá, afinal faz parte do ônus da democracia conviver com opiniões divergentes no âmbito da liberdade de expressão.

Também se faz relevante falar sobre os métodos de garantia de maior transparência à moderação de conteúdo realizada pelos provedores de aplicação, uma vez que não há garantia de segurança e confiabilidade pela simples automatização desses processos.

5.3 TRANSPARÊNCIA NA MODERAÇÃO DE CONTEÚDO

Diante dos desafios postos em relação a governança na moderação algorítmica, algumas salvaguardas podem ser implementadas, seja buscando garantir maior confiabilidade nos parâmetros decisórios utilizados para a remoção de conteúdo previamente ao controle judicial, seja possibilitando mecanismos de apelo e transparência sobre as razões pelas quais determinado conteúdo retirado viola os termos de conduta de uma plataforma, por exemplo.

Veja-se, por exemplo, no projeto de Lei 2.630 de 2020, na atual redação de seu artigo 4º, III, onde há expressa menção a transparência:

> **Art. 4º** Esta Lei tem como objetivos: (...) **III –** a busca por maior transparência das práticas de moderação de conteúdos postados por terceiros em redes sociais, com a garantia do contraditório e da ampla defesa; (BRASIL, 2020)

Na sequência, o artigo 13 do mesmo projeto de lei:

> **Art. 13.** Os provedores de redes sociais devem produzir relatórios trimestrais de transparência, disponibilizados em seus sítios eletrônicos, em português, para informar os procedimentos e as decisões de tratamento de conteúdos gerados por terceiros no Brasil, bem como as medidas empregadas para o cumprimento desta Lei. (BRASIL, 2020)

O dispositivo segue em seu parágrafo primeiro e respectivos incisos estabelecendo requisitos mínimos que devem constar destes relatórios, com particular ênfase na publicação de números agregados sobre as postagens e conteúdos removidos, sendo este um parâmetro incompleto, como destacam Maranhão *et al*, uma vez que há uma concentração somente no que ocorre após a remoção do conteúdo (MARANHÃO, 2020).

Aqui se faz possível uma melhor caracterização do problema, pois a mera divulgação de relatórios de transparência sobre moderação de conteúdo não supera

o fato de que se trata de uma perspectiva limitada, refletindo a caracterização do conteúdo problemático dada pelas próprias plataformas (MARANHÃO, 2020).

Se por um lado, a abordagem da autorregulação exercida pelas plataformas de redes sociais e provedores de aplicação para decidirem internamente questões ligadas a moderação e remoção de conteúdo traz o benefício da celeridade na decisão (eis que é feita sem intervenção judicial) o paradigma da transparência deve ser melhor qualificado, uma vez que, segundo Hanna Bloch Wehba, tratando da regulação alemã – o chamado *NetzDG* – destaca:

> A regulação requer das plataformas que documentem seus esforços para derrubar conteúdo que está sujeito a uma específica reclamação, mas não captura de forma completa as decisões das companhias para implantar mecanismos automatizados para evitar reclamações em primeiro lugar. (BLOCH-WEHBA, 2020, p. 55)

Nesse sentido, o relatório produzido no âmbito do parlamento europeu e intitulado *Regulating disinformation with artificial intelligence,* é essencial na compreensão do que se busca quanto a melhoria na transparência da moderação de conteúdo automatizada em relação à desinformação.

A partir da análise de diversas iniciativas e opções de políticas regulatórias no âmbito da União Europeia e mesmo do setor privado, os autores do relatório elencam boas práticas já existentes e oportunidades de melhoria tanto em termos de regulação estatal direta como na corregulação e na autorregulação pelos entes privados.

Com base nas políticas de conteúdo das plataformas de Youtube, Facebook e Twitter, é dada ênfase sobre a necessidade de transparência no processo decisório da remoção e no processo de reconhecimento de conteúdo como sendo desinformação:

> Mais controverso ainda do que o ranqueamento algorítmico é o reconhecimento e remoção automatizada de conteúdo. Não está claro com que frequência e sob quais circunstâncias a filtragem ou bloqueio *ex ante* acontece nas plataformas. Parte é acionada por máquinas, mas não é claro

> como a ilegalidade do conteúdo ou sua violação das diretrizes comunitárias é determinada, nem quais são as salvaguardas postas para prevenir censura excessiva. Transparência na frequência e categorias de filtragem de conteúdo está ausente para públicos externos, nem parece haver possibilidade de recurso. (MARSDEN; MEYER, 2019, p. 45)

Em sequência, para se pensarem em possíveis soluções ao problema, o relatório aponta aos chamados Princípios de Santa Clara, sendo estes um norte para se avaliar a qualidade da transparência dada pelas plataformas e provedores de conteúdo em geral sobre suas políticas de moderação. Veja-se que neste ponto passamos da simples ideia de garantia de transparência na moderação algorítmica para uma reflexão sobre a qualidade da transparência garantida pelos intermediários.

Os Princípios de Santa Clara sobre transparência e prestação de contas na moderação de conteúdo foram elaborados em 2018, na Segunda Conferência de Moderação de Conteúdo em escala, onde organizações civis, especialistas e defensores da liberdade de expressão em rede definiram um conjunto de parâmetros mínimos que companhias deveriam adotar para garantir que a "aplicação de suas diretrizes de conteúdo seja justa, imparcial, proporcional e respeite os direitos dos usuários" (ACLU FOUNDATION OF NORTHERN CALIFORNIA, 2018).

O primeiro dos princípios diz respeito aos números, sugerindo-se às plataformas que publiquem, trimestralmente, os números de postagens removidas e contas permanentemente ou temporariamente suspensas em razão de violações às suas diretrizes de conteúdo.

Estas informações devem ser desmembradas em distintas categorias, como por exemplo, número de postagens e contas marcadas e número de postagens e contas posteriormente removidas ou suspensas de acordo com a origem da marcação, como governo, usuários, detecção automatizada[5].

Segundo, ao informarem os usuários sobre a remoção de conteúdo de sua autoria ou responsabilidade, as plataformas deveriam fazer constar, no mínimo, as

[5] Cabe comentar que o Projeto de Lei 2.630 de 2020, na atual redação do § 1º de seu artigo 13, e seus incisos, aparenta estar em consonância com a intenção deste primeiro princípio, eis que são elencadas distintas categorias para os números que devem constar dos relatórios das plataformas.

[Received/Recebido: Maio 23, 2021; Accepted/Aceito: Julho 23, 2021]

seguintes informações: a URL, link ou qualquer informação suficiente a indicar o que foi removido, a cláusula específica das diretrizes comunitárias violada pelo conteúdo removido, como o conteúdo foi detectado e removido e uma explicação sobre como o usuário poderia apelar da decisão. Ainda, recomenda-se que as plataformas forneçam explicação sobre como a detecção automatizada de conteúdo é utilizada em cada categoria de conteúdo proibido. Por fim, o terceiro princípio diz respeito à possibilidade do usuário recorrer da decisão sobre a remoção de conteúdo, sendo que as plataformas devem garantir oportunidade razoável para um recurso temporâneo de qualquer remoção de conteúdo ou suspensão de conta.

A oportunidade razoável deve ser entendida segundo este princípio a partir de três parâmetros: **(i)** a possibilidade de revisão da decisão por seres humanos individualmente ou um painel, que não esteja envolvido na decisão inicial – a qual determinou a remoção ou suspensão[6]; **(ii)** a oportunidade ao usuário para que este apresente informações adicionais a serem consideradas na revisão; **(iii)** notificação dos resultados da revisão, em termos que o usuário possa compreender os fundamentos da decisão.

Avaliados individualmente cada um dos princípios, é importante destacar que seus elaboradores os consideram apenas como pontos de partida para diálogos e pesquisas mais aprofundadas sobre o assunto (ACLU FOUNDATION OF NORTHERN CALIFORNIA, 2018). Resta acompanhar se estas novas estruturas de governança criam salvaguardas mais ou menos poderosas contra remoção indevida, censura e vigilância (BLOCH-WEHBA, 2020, p. 62).

6 CONSIDERAÇÕES FINAIS

Na linha do defendido por Cass Sunstein, ressalte-se que, apesar dos riscos de um efeito inibitório à livre expressão pessoal, o que necessariamente enseja

[6] Apesar das inegáveis limitações quantitativas sobre esse formato de revisão, cabe citar o Oversight Board do Facebook, como um exemplo de tentativa nesse sentido. A respeito: LEMOS, 2020.

atenção adequada a se evitar uma regulação equivocada quanto a disseminação de boatos nas redes, este efeito também poderia auxiliar a enfraquecer as falsidades particularmente danosas e destrutivas. Isto porque as influências sociais e a assimilação tendenciosa garantem que os boatos falsos se espalhem, e se não houver qualquer efeito inibitório legislativo, viveremos sob grave risco.

Igualmente arriscado usar o argumento democrático, como aponta o autor Gross, tanto como fundamento para a retirada de *fake news* das redes como expressão de ceticismo a respeito de iniciativas envolvendo a retirada de conteúdo.

O exercício de ponderação entre os direitos fundamentais da liberdade de expressão e a liberdade de informação, além da inviolabilidade da imagem, não é algo novo, e o fenômeno das *fake news* não deve ser analisado fora desse contexto. A necessidade é que se construam estratégias envolvendo todos os atores interessados, ou seja, Estado, provedores de conteúdo e a sociedade civil.

As iniciativas destinadas a capacitar os indivíduos a avaliarem cuidadosamente as informações com que se deparam nas redes, consultando as fontes e evitando de replicar informações contendo falsidades também poderia ser uma saída. Neste âmbito, se inserem as chamadas agências de checagem, sendo alguns exemplos brasileiros nesse sentido a Agência Lupa[7], Agência Pública[8] e a Aos Fatos[9].

A título de exemplo de iniciativas adotadas esse âmbito, a Agência Aos Fatos mantém acervo com diversos manuais instruindo a como se checar a veracidade de informações em diversos âmbitos, desde redes sociais a declarações políticas. Mesmo as empresas de tecnologia, como Google e Facebook, por exemplo, já anunciaram medidas de enfrentamento à desinformação, conforme aponta o autor Rais.

De outro lado, estão as estratégias voltadas a gerar mudanças estruturais visando evitar que os indivíduos sejam expostos à desinformação. A esse respeito,

[7] Disponível em: https://piaui.folha.uol.com.br/lupa/
[8] Disponível em: https://apublica.org/
[9] Disponível em: https://aosfatos.org/

Rais destaca quanto a subdivisão em duas grandes áreas, quais sejam, mudanças estruturais carreadas pelas próprias empresas de tecnologia, diante de alterações em suas políticas de funcionamento[10], e as mudanças estruturais efetivadas diante de uma atuação jurisdicional.

Particularmente, vê-se a necessidade de um aparato conceitual mais sofisticado, segundo o autor Macedo Júnior, de modo que deve ser esclarecida socialmente a imprecisão do termo *fake news*, bem como o relevante e problemático papel da tecnologia no desenvolvimento do fenômeno, para assim ser possível um tratamento jurisdicional adequado sobre a matéria.

Neste contexto, o enfrentamento da disseminação da desinformação através da internet enseja diversas abordagens, sendo uma delas a moderação de conteúdo diretamente pelos intermediários, a qual vem ganhando destaque à medida que as plataformas de redes sociais e provedores de conteúdo são cada vez mais incentivadas a prevenirem a disseminação de conteúdo ilícito.

A moderação automatizada de conteúdo, ou moderação algorítmica, é utilizada por diversos provedores de conteúdo e tende a ser incrementada cada vez mais para enfrentar problemas de escala como para atender a exigências regulatórias estatais com relação à celeridade na remoção de conteúdo ilícito.

Considerando a ausência de clareza sobre os procedimentos adotados na moderação algorítmica, desde a detecção até a remoção ou suspensão de conteúdo ilícito, são elencados alguns princípios destinados a garantir maior transparência e confiabilidade na prestação de contas das plataformas sobre suas políticas comunitárias.

O fenômeno da desinformação em rede não é estático e acompanha o desenvolvimento da tecnologia, de tal modo que os desafios regulatórios atuais poderão ser outros futuramente, com diferentes desdobramentos. A moderação

[10] Aqui há de se fazer a ressalva quanto ao enorme poder das companhias de tecnologia a respeito do conteúdo exposto nas redes, e, ao mesmo tempo, o baixo incentivo que estas teriam a expor seus usuários a informações que lhes desagradem, embora sejam verdadeiras, *furando* então os filtros-bolha.

algorítmica não é única, sendo que existem diferentes abordagens técnicas e diferentes etapas ligadas a filtragem, de-priorização e bloqueio de conteúdo, além de inúmeras outras, cada qual com particularidades e desafios específicos e que exigiriam estudos dedicados.

Para a garantia de maior transparência com relação à moderação algorítmica, a elaboração e adoção de princípios não vinculantes pelas plataformas exige maior compreensão sobre os seus limites e modos de fiscalização – seja estatal ou pelos próprios usuários.

Ainda, resta entender se e de que formas o combate à desinformação pode ser potencializado ou inibido pela utilização dos algoritmos, bem como os riscos envolvidos para todos os atores interessados. Levando em consideração que este combate através da moderação deve-se dar dentro dos limites constitucionais, em respeito à separação dos Poderes e à democracia.

Isto porque, o controle do discurso não pode se dar por empresas privadas ou pelo Estado sem transparência sob o risco de colocar em cheque o princípio constitucional da liberdade de expressão e também da própria democracia. Afinal, quem controlaria o órgão controlador dos discursos? O limiar deste controle não pode criminalizar o discurso quando se tratar de debate, seja de conservadores ou progressistas, pois isto ensejaria um controle prévio de discursos, atuação típicas de regimes autoritários e ditatoriais.

Segundo Carlos Drummond de Andrade: “a conquista da democracia é algo que faz tanta poeira, que por medo da bagunça, preferimos, normalmente, optar pela arrumação”. Neste caso, a era da desinformação e da pós-verdade pode aparentar “uma bagunça” que requer arrumação, contudo, se a arrumação for sob o controle das empresas privadas ou do Estado, mas sem transparência, corremos o risco de expor a própria existência do Estado Democrático de Direito, a controle de grupos políticos e econômicos, que não necessariamente estarão comprometidos com a República.

REFERÊNCIAS

ACLU FOUNDATION OF NORTHERN CALIFORNIA, *Center of Democracy and Technology, Electronic Frontier Foundation, New America's Open Technology Institute* et.al. **Santa Clara Principles on Transparency and Accountability in Content Moderation**, Disponível em: https://santaclaraprinciples.org/. Acesso em 05 de dezembro de 2020.

AGÊNCIA LUPA. Disponível em: https://piaui.folha.uol.com.br/lupa/ Acesso em 10 de novembro de 2019.

AGÊNCIA PÚBLICA. Disponível em: https://apublica.org/ Acesso em 10 de novembro de 2019.

AOS FATOS. Disponível em: https://aosfatos.org/. Acesso em 10 de novembro de 2019.

BLOCH-WEHBA, Hannah. *Automation in Moderation. Texas A&M University School of Law Legal Studies Research Paper* No. 20-33. **Cornell International Law Journal**. New York, 2020, p. 3 No prelo.

BRASIL. **Constituição da República Federativa do Brasil de 1988**.

BRASIL. **Lei 12.965 de 24 de abril de 2014**. Estabelece princípios, garantias, direitos e deveres para o uso da Internet no Brasil.

BRASIL. Senado Federal. **Projeto de Lei nº 2.630 de 2020**. Institui a Lei Brasileira de Liberdade, Responsabilidade e Transparência na Internet.

BRASIL. Senado Federal. **Projeto de Lei do Senado nº 471 de 2018**. Altera o Decreto-Lei nº 2.848, de 7 de dezembro de 1940, a Lei nº 4.737, de 15 de julho de 1965, e a Lei nº 12.965, de 23 de abril de 2014, para dispor sobre a definição das infrações penal, eleitoral e civil de criar ou divulgar notícia falsa, e cominar as respectivas penas

BRASIL. Superior Tribunal de Justiça. **Recurso Especial 1.342.640/SP**. Relatora Ministra Nancy Andrighi, 3ª Turma. Julgado: 07 de fevereiro de 2017. DJE: 24 de fevereiro de 2017.

BRASIL. Supremo Tribunal Federal. **Tema 987**. Discussão sobre a constitucionalidade do art. 19 da Lei n. 12.965/2014 (Marco Civil da Internet) que determina a necessidade de prévia e específica ordem judicial de exclusão de conteúdo para a responsabilização civil de provedor de internet, websites e gestores

de aplicativos de redes sociais por danos decorrentes de atos ilícitos praticados por terceiros. Relator Ministro Dias Toffoli.

BRASIL. Presidência da República. Secretaria Especial de Comunicação Social. **Pesquisa brasileira de mídia 2016**: hábitos de consumo de mídia pela população brasileira. Brasília: Secom, 2016.

CASTELLS, Manuel. **Ruptura**: A crise da democracia liberal. Rio de Janeiro: Zahar, 2018.

COMISSÃO EUROPEIA. **Uma abordagem multidimensional para a desinformação**: relatório do Grupo Independente de Alto Nível sobre fake news e desinformação online. União Europeia, 2018.

CONJUR. **Moraes manda Crusoé tirar do ar notícia que associa Toffoli a delação de Odebrecht**. Disponível em: https://www.conjur.com.br/2019-abr-15/moraes-manda-revista-tirar-ar-noticia-ligando-toffoli-odebrecht. Acesso em 12 de maio de 2019.

D'ANCONA, Matthew. **Pós-verdade**: a nova guerra contra os fatos em tempos de fake-news. Tradução: Carlos Szlakj. Barueri: Faro Editorial, 2018.

EIFERT, Martin. Lei Alemã para a Melhoria da Aplicação da Lei nas Redes Sociais (NetzDG) e a Regulação da Plataforma. In. ***Fake News* e regulação**. ABBOUD, Georges; NERY JUNIOR, Nelson; CAMPOS, Ricardo (Coords). São Paulo: Thomson Reuters, 2018. p. 80.

GIACCHETA, André Zonaro. Atuação e responsabilidade dos provedores diante das fake news. In: ***Fake News***: a conexão entre a desinformação e o direito. RAIS, Diogo (Coord). São Paulo: Thomson Reuters, 2018.

GORWA, Robert; BINNS, Reuben; KATZENBACH, Christian. *Algorithmic content moderation: Technical and political challenges in the automation of platform governance*. **Big Data & Society**, 2020.

GRIGORI, Pedro. **20 projetos de lei no Congresso pretendem criminalizar fake news**. A Pública. Disponível em: https://apublica.org/2018/05/20-projetos-de-lei-no-congresso-pretendem-criminalizar-fake-news. Acesso em 23 de novembro de 2019.

GROSS, Clarissa Piterman. *Fake News* E Democracia: Discutindo O Status Normativo Do Falso E A Liberdade De Expressão. In: **Fake News**: a conexão entre a desinformação e o direito. RAIS, Diogo (Coord). São Paulo: Thomson Reuters, 2018. p. 153-173.

HAJE, Lara. Deputados querem aprimorar projeto contra fake news aprovado pelo Senado. **Agência Câmara de Notícias.** 2020. Disponível em: https://www.camara.leg.br/noticias/675696-deputados-querem-aprimorar-projeto-contra-fake-news-aprovado-pelo-senado

KALPOKAS, Ignas. ***A Political Theory of Post-Truth***. Cham, Switzerland: Palgrave Macmillan, 2019.

KEYES, Ralph. ***The Post Truth Era***: Dishonesty and Deception in Contemporary Life. New York: St. Martin's Press, 2004.

LEMOS, Ronaldo. O Oversight Board do Facebook. **Folha de São Paulo**. 2020 Disponível em: https://www1.folha.uol.com.br/colunas/ronaldolemos/2020/05/o-oversight-board-do-facebook.shtml?origin=folha. Acesso em 05 de dezembro de 2020.

LEVY, Steven. Mark Zuckerberg Is an Arbiter of Truth – Whether He Likes It or Not. **Wired**. Disponível em: https://www.wired.com/story/mark-zuckerberg-is-an-arbiter-of-truth-whether-he-likes-it-or-not/. Acesso em 08 de novembro de 2020.

LOPEZ, German. Pizzagate, *the fake news conspiracy theory that led a gunman to DC's Comet Ping Pong*, explained. **Vox.** Disponível em https://www.vox.com/policy-and-politics/2016/12/5/13842258/pizzagate-comet-ping-pong-fake-news. Acesso em 15 de novembro de 2020.

MACEDO JÚNIOR, Ronaldo Porto. **Fake news e as novas ameaças à liberdade de expressão**. In: Fake News e regulação. ABBOUD, Georges; NERY JUNIOR, Nelson; CAMPOS, Ricardo (Coords). São Paulo: Thomson Reuters, 2018.

MARSDEN, Chris. MEYER, Trisha. ***Regulating disinformation with artificial intelligence***: Effects of disinformation initiatives on freedom of expression and media pluralism. European Union: Brussels, 2019. Disponível em: https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-researc/168076277c Acesso em 19 de novembro de 2020.

MARANHÃO, Juliano. et. al. Como regular a moderação privada de conteúdo nos novos espaços públicos? **Consultor Jurídico**. Disponível em: https://www.conjur.com.br/2020-set-01/direito-digitalcomo-regular-moderacao-privada-conteudo-novos-espacos-publicos. Acesso em 08 de novembro de 2020.

MCINTYRE, Lee C. **Post-truth**. Cambridge, MA: MIT Press, 2018.

MENDES, Laura, Schertel; DONENA, Danilo; BACHUR, João Paulo. **Manipulação da informação, *fake news* e os novos desafios da democracia.** Cadernos

Adenauer xix (2018), nº4 Fake news e as eleições 2018. Rio de Janeiro: Fundação Konrad Adenauer, dezembro 2018. p. 129-156.

MOUNK, Yascha. ***The people vs democracy****: why our freedom is in danger and how to save it.* Cambridge, Massachusetts: Harvard University Press, 2018.

NAÇÕES UNIDAS BRASIL. **Brasil é o quarto país com mais usuários de Internet do mundo, diz relatório da ONU**. 2017. Disponível em: https://nacoesunidas.org/brasil-e-o-quarto-pais-com-mais-usuarios-de-internet-do-mundo-diz-relatorio-da-onu/ Acesso em 14 de setembro de 2019.

NEWMAN, Nic., et. al. ***Reuters Institute Digital News Report 2018***. Disponível em: http://media.digitalnewsreport.org/wp-content/uploads/2018/06/digital-news-report-2018.pdf. Acesso em 09 de novembro de 2019.

PRIOR, Hélder. Mentira e política na era da pós-verdade: *fake news*, desinformação e factos alternativos. In: **Comunicação digital**: Media, práticas e consumos. Lisboa: NIP-C@M & UAL. LOPES, Paula; REIS, Bruno (Coords.). pp. 75-97. Disponível em <http://hdl.handle.net/11144/3976>. Acesso em 16 de outubro de 2019.

RAIS, Diogo. *Fake News* e eleições. In: ***Fake News***: a conexão entre a desinformação e o direito. São Paulo: Thomson Reuters, 2018. p. 105-131.

SUNSTEIN, Cass. **A verdade sobre os boatos**: como se espalham e porque acreditamos neles. Trad. Marcio Hack. Rio de Janeiro: Elsevier, 2010. 139 p.

SUNSTEIN, Cass*. #**Republic**: divided democracy in the age of social media. Princeton: Princeton University Press*, 2017.

SUPREMO TRIBUNAL FEDERAL. **Inquérito 4.781/DF**. Rel. Ministro Alexandre de Moraes. Data da publicação: 18 de abril de 2019.

TEFFÉ, Chiara Spadaccini de; SOUZA, Carlos Affonso Pereira de. Fake news: como garantir liberdades e conter notícias falsas na internet? In: MENEZES, Joyceane Bezerra de; TEPEDINO, Gustavo (Coord.). **Autonomia privada, liberdade existencial e direitos fundamentais**. Belo Horizonte. Fórum, 2019.

VOSOUGHI, Soroush; ROY, Deb; ARAL, Sinan*. **The spread of true and false news online**. Science.* 359. 1146-1151. 2018. Disponível em: 10.1126/science.aap9559. Acesso em 27 de outubro de 2019.

WARDLE, Claire*. **Fake news. It's complicated**.* Disponível em https://firstdraftnews.org/fake-news-complicated/ Acesso em: 18 de setembro de 2018.

WARDLE, Claire; DERAKHSHAN, Hossein. ***Information disorder****: Toward an interdisciplinary framework for research and policy making. Council of Europe: Strasbourg*, 2018. Disponível em: <https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-researc/168076277c>. Acesso em 05 de novembro de 2019.

WASHINGTON POST. ***Post-truth' named 2016 word of the year by Oxford Dictionaries****.* Disponível em: https://www.washingtonpost.com/news/the-fix/wp/2016/11/16/post-truth-named-2016-word-of-the-year-by-oxford-dictionaries/ Acesso em 09 de novembro de 2019.