Fake news, liberdade de expressão e moderação nas redes sociais

Introdução: os impactos perniciosos das fake news

A utilização de fake news tem seríssimos impactos não somente no que tange à crise de confiança verificada nos dias que correm, inclusive e principalmente quanto às instituições públicas [1] ; resulta, muitas vezes, em seríssimos e irreversíveis resultados práticos, podendo chegar, até mesmo, ao

da dos cidadãos.

Tal assertiva é corroborada por estudos empíricos sobre o assunto: em experimento levado a efeito pelo *Centre for Countering Digital Hate* do *King's College London* , restou demonstrado que 60% das pessoas que acreditam nas informações mentirosas sobre o coronavírus estão mais sujeitas a descumprirem as medidas de isolamento social. (Allington; Dhavan. 2020)

No mesmo sentido, o estudo "*More than words: leader's speech and risky behavior during a pandemic*", publicado na *Cambridge Working Papers in Economics* da Faculdade de Economia da Universidade de Cambridge [2], segundo o qual existe uma direta relação de causalidade entre as declarações e atitudes negacionistas do presidente brasileiro e o afrouxamento das medidas de isolamento social. (Ajzenman; Cavalcanti; Mata, 2020)

Digna de registro, ainda, pesquisa do *Reuters Institute* de Oxford, segundo a qual políticos, celebridades e pessoas públicas em geral são responsáveis pelo espraiamento de 70% das notícias relacionadas à Covid-19 nas redes sociais — destas, cerca de 20% são falsas (Brenen, Simon, Howard, Nielsen. 2020)

Diante de tão inquietante cenário, ganha extrema relevância a discussão desencadeada a partir da publicação, pelo chefe do Executivo federal, da MP 1068, às vésperas do feriado de 7 de setembro do corrente ano.

Chama atenção, em especial, o dispositivo proposto com relação à limitação da possibilidade de moderação de conteúdo pelas redes sociais, com a virtual de pronta e unilateral retirada do ar de notícias falsas.

Não obstante a condição natimorta da iniciativa (o Senado rapidamente rejeitou a medida provisória em questão, subtraindo-lhe, assim, qualquer chance de subsistência), a discussão permanece na pauta do dia: inconformado com a absoluta e imediata inviabilização de sua intenção, o governo, prontamente, se encarregou de estruturar projeto de lei, mais uma vez expressando intenção de limitação da possibilidade de moderação de conteúdo, com a criação de uma estrutura formal e burocrática voltada a garantir a plena circulação de notícias falsas, ainda que tão somente até a manifestação do Judiciário.

Trata-se do PL 2831/2021, que, a partir da inserção do artigo 8º-C no Marco Civil da Internet, condiciona a exclusão de conteúdo ou perfil a prévia manifestação judicial, em até 24 horas (inovação com relação à polêmica MP 1068) — salvo as hipóteses de pornografia infantil, terrorismo, ou *de potencial dano significativo ou efeitos de difícil reversão.*

A primeira grande preocupação gerada nesse âmbito decorre do fato de que, em se tratando de internet, e da velocidade da propagação das assertivas por seu meio disparadas, poucas horas — por vezes até minutos — são suficientes para o espraiamento das fake news — em um processo de dificílima reversão (a qual chega a ser por vezes absolutamente inviável).

O rol trazido pelo projeto de lei não contempla as fake news entre as hipóteses de pronta exclusão — e, não obstante a possibilidade da abertura representada pela derradeira situação excepcional elencada (potencial dano significativo ou efeitos de difícil reversão), inquestionável representar uma ameaça à segurança da sociedade e à própria democracia.

Realmente, a grande discussão está em quem terá a capacidade de, satisfatoriamente, analisar a incidência dessas condições — em especial a terceira, marcadamente aberta e com elementos de marcante subjetividade.

Mais que isso: quem goza de legitimidade — tanto técnica quanto democrática — para levar a efeito esse tipo de análise?

Trata-se de brecha passível de aplicação com vistas à extirpação das perniciosas fake news — mas que cria inquestionável e indesejável insegurança jurídica.

Conforme esclarece o Parecer n° 01/2018 do Conselho de Comunicação Social do Congresso Nacional, até abril de 2018 havia 14 projetos de lei sobre o assunto.

Nesse particular, importante lembrar a tramitação — marcada por intensas discussões, inclusive por meio de audiências públicas, do PL 2630/2020, igualmente voltado à alteração da Lei n° 12.965/14 (Marco Civil da Internet), o qual prevê expressamente a possibilidade de exclusão unilateral e imediata de conteúdos capazes de causar dano imediato ou de difícil reparação, comprometer a segurança da informação ou do usuário, resultar em violação a direitos de crianças e adolescentes, na prática de crimes tipificados na Lei 7716/1989, ou em grave comprometimento da usabilidade, integridade ou estabilidade da aplicação.

Tal ponto, absolutamente fundamental da discussão, não vem passando desapercebido em termos globais, em que vários documentos e indicadores foram construídos.

Diretrizes globais sobre o tema: tendências
Interessante referência é o estudo levado a efeito pela Universidade de Palermo, Argentina, segundo o qual a garantia constante do artigo 13 da Convenção Americana de Direitos Humanos (Pacto de San Jose da Costa Rica) com relação à liberdade de expressão e pensamento não pode ser sujeita a censura prévia, traduzindo potencial impeditivo à moderação privada de assuntos e perfis na internet.

Ocorre que a convenção estabelece, ainda, que a lei deve proibir toda propaganda a favor da guerra, bem como toda apologia ao ódio nacional, racial ou religioso que constitua incitamento à discriminação, à hostilidade, ao crime ou à violência.

A conciliação entre os dois parâmetros está intimamente ligada à responsabilidade legalmente atribuída ao intermediário (no caso, as redes sociais), havendo a possibilidade de graduação que vai desde a absoluta ausência de responsabilidade pelo conteúdo veiculado, em que não há sequer motivação para uma sua interferência, dependendo qualquer exclusão de conteúdo da determinação advinda de órgãos oficiais (modelo adotado nos Estados Unidos); a consagração de uma imunidade condicional, em que o moderador tem o dever de observar procedimentos pré-definidos, seja de notificação da parte interessada e retirada do conteúdo (adotado por União Europeia, Cingapura, Gana, Uganda e África do Sul), ou aviso e notificação sobre pleito de terceiro voltado à derrubada do conteúdo (modelo canadense); ou, ainda, a responsabilidade integral, em que os intermediários respondem por todo o conteúdo divulgado em suas plataformas, sendo-lhe permitido, portanto, o monitoramento e pronta extirpação de conteúdos com potencial de gerar responsabilização (modelo tailandês).

Os denominados Princípios de Manila, criados na conferência *RightsCon*, realizada em 2015 nas Filipinas com vistas ao estabelecimento de garantias e parâmetros quanto à responsabilidade de intermediários e liberdade de expressão da rede, indica como modelos mais recomendáveis a primeira e segunda alternativas — em que não há a pronta e incontinente exclusão de conteúdo/perfil pelo moderador.

Englobam, nesse sentido, seis diretrizes fundamentais: 1) os intermediários devem ser protegidos por lei com relação aos conteúdos produzidos por terceiros; 2) a remoção de conteúdo deve necessariamente decorrer de determinação judicial; 3) os pedidos de restrição de conteúdos devem ser claros, não ambíguos e observar o devido processo legal; 4) normas, ordens e práticas de restrição devem observar testes de necessidade e proporcionalidade, assim como o devido processo legal; 5) a transparência e rendição de contas quanto às políticas e práticas de restrição de conteúdos é fundamental.

Dignos de registro, também, os princípios formulados pela *Eletronic Frontier Foundation* em conjunto com organizações como artigo 19 e Derechos Digitales com vistas à orientar a atividade legislativa e de intermediários quanto ao tema: 1) os intermediários não devem responder por conteúdos produzidos por terceiros; 2) a remoção de conteúdo deve estar condicionada a uma ordem judicial; 3) pedidos de restrição de conteúdo devem ser claros, não ambíguos e seguir o devido processo; 4) normas e práticas de restrição de conteúdo devem seguir os testes de necessidade e proporcionalidade e o devido processo legal; 5) transparência e prestação de contas devem ser integradas em leis e em políticas de práticas de restrição de conteúdos.

Para além desses parâmetros, há, ainda, os Princípios de Santa Clara, resultantes da primeira conferência *Content Moderation at Scale*, realizada na Califórnia em 2018, e voltados à promoção e garantia da transparência por meio da publicização de: 1) números: as empresas devem publicar trimestralmente, em formato aberto e acessível relatórios com os números de postagens removidas e contas permanente ou temporariamente suspensas em razão de violações de diretrizes de conteúdo. Esses dados devem ser fornecidos em um relatório regular; 2) aviso: necessidade de notificar os usuários cujo conteúdo tenha sido removido ou cuja conta tenha sido suspensa sobre o motivo da retirada ou suspensão, devendo as comunicações permanecerem disponíveis em plataforma duradoura e acessível, inclusive, aos denunciantes (mesmo no caso de suspensão/encerramento de conta); 3) recurso: possibilidade de apresentação de recursos contra remoção de conteúdo ou suspensão de conta, com eventuais processos de revisão externa independente.

À parte das questões de ordem teórica e orientativa, o cenário brasileiro atual evidencia a importância e recorrência da interferência dos moderadores para a preservação da "normalidade" da rede: os relatórios trimestrais disponibilizados pelo YouTube evidenciam a remoção, entre janeiro e junho deste ano, de 6.417.950 canais, resultando na extirpação de 126.309.949 vídeos, em razão de veicularem: 1) *spam*, conteúdo enganoso ou golpes (90,3%); 2) nudez ou conteúdo sexual (4,3%); 3) assédio e *bullying* virtual (1,9%).

Mundialmente, os países com maior índice de remoção de conteúdo nesse período foram Índia, Estados Unidos, Brasil, Indonésia e Rússia [3].

Conclusões

De se lembrar que a moderação de conteúdos é não raramente levada a efeito de forma automatizada, por algoritmos — que não são necessariamente neutros, podendo apresentar vieses altamente prejudiciais à isonomia, inclusão e, em última análise, à própria democracia (discriminação algorítmica).

Tal atividade se torna ainda mais problemática quando considerado o fato de ser desempenhada por robôs de propriedade de pouquíssimos conglomerados econômicos com ideologia e objetivos próprios — e voltados fundamentalmente ao lucro.

Foi o que evidenciou o mais recente escândalo envolvendo o Facebook, em que uma ex-funcionária apresentou documentos indicadores de uma postura de crescimento econômico em detrimento da segurança da rede e dos usuários, com tratamentos diferenciados a celebridades e relativização de mecanismos de segurança quanto a ataques à democracia e à própria segurança da sociedade [4].

É absolutamente fundamental o desenvolvimento e aplicação de critérios transparentes e objetivos de remoção algorítmica de notícias, de forma a garantir, mais que tratamento equânime e impessoal a todos os indivíduos/ideologias, o prévio conhecimento e potencial controle dos motivos e circunstâncias ensejadores da remoção de conteúdo.

Imprescindível, ainda, o envolvimento da sociedade civil com vistas ao combate e impulsionamento de notícias capazes de desmentir as falsas — estratégia comprovadamente eficaz, conforme se vê da pesquisa levada a efeito pela Fundação Getúlio Vargas com relação às inverdades propagadas na rede quanto à morte da vereadora Marielle Franco e a reação voltada à sua desconstrução, que atingiu vulto e alcance muitíssimo maiores a partir de uma mobilização popular concisa e efetiva [5].

A maravilhosa ágora representada pela internet e pelas redes sociais não pode ser manipulada em favor de interesses escusos, em manobras voltadas à captação do poder e instrumentalização da sociedade em prol de interesses escusos.

Mais uma vez, o posicionamento — e a proatividade — da sociedade serão fundamentais.

Referências bibliográficas

AJZENMAN; Nicolás; CAVALCANTI, Tiago; DA MATA, Daniel. More than words: leader's speech and risky behavior during a pandemic. Cambridge-INET Working Paper Series No: 2020/19. Cambridge Working Papers in Economics: 2034. Disponível em https://www.econ.cam.ac.uk/research-files/repec/cam/pdf/cwpe2034.pdf. Acesso em 02/10/2021.

ALLINGTON, Daniel; DHAVAN, Nayana. The relationship between conspiracy beliefs and compliance with public health guidance with regard to COVID-19.
London: Centre for Countering Digital Hate. 2020. 6p. Disponível em https://kclpure.kcl.ac.uk/portal/files/127048253/Allington_and_Dhavan_2020.pdf. Acesso em 03/05/2020.

BRENEN, J. Scott; SIMON, Felix; HOWARD, Philip N.; NIELSEN, Rasmus Klein. Types, sources and claims of COVID-19 misinformation. 2020. Disponível em https://reutersinstitute.politics.ox.ac.uk/types-sources-and-claims-covid-19-misinformation. Acesso em 02/05/2020.

Fundação Getúlio Vargas. Diretoria de Análise de Políticas Públicas. Reação a boatos superou a difusão de informações contra Marielle no Twitter, aponta estudo da FGV DAPP. Disponível em http://dapp.fgv.br/reacao-boatos-superou-difusao-de-informacoes-contra-marielle-no-twitter-aponta-estudo-da-fgv-dapp/. Acesso em 02/10/2021.

GOOGLE. Transparency Report. Cumprimento das diretrizes da comunidade do YouTube. Disponível em: https://transparencyreport.google.com/youtube-policy/removals?hl=pt_BR. Acesso em 05/10/2021.

LATINOBARÔMETRO. Informe 2018. Disponível em file:///C:/Users/escob/Downloads/F00008421-INFORME_2018_LATINOBAROMETRO.pdf. Acesso em 15/09/2020.

Princípios de Manila sobre Responsabilidade dos Intermediários. Práticas recomendadas para limitar a responsabilidade dos intermediários pelos conteúdos de terceiros e promover liberdade de expressão e inovação. 2015. Disponível em https://www.eff.org/files/2015/07/02/manila_principles_1.0_pt.pdf . Acesso em 06/108/2021.

Universidad de Palermo. Faculdad de Derecho. Centro de Estudios en Libertad de Expresión y Acceso a la Información. Content Moderation and private censorship: standards drawn from the jurisprudence of the Inter-American Human Rights system. Dez/2017. Disponível em https://www.ohchr.org/Documents/Issues/Opinion/ContentRegulation/CELE.pdf. Acesso em 05/10/2021.

[1] Conforme bem ilustra a mais recente pesquisa do Latinobarômetro, segundo a qual o ano de 2018 é qualificado como um *annus horribilis* para a democracia latino-americana. Disponível em https://www.latinobarometro.org/latNewsShowMore.jsp?evYEAR=2018&evMONTH=-1.

[2] https://econpapers.repec.org/paper/camcamdae/2034.htm.

[3] GOOGLE. Transparency Report. Cumprimento das diretrizes da comunidade do YouTube. Disponível em: https://transparencyreport.google.com/youtube-policy/removals?hl=pt_BR. Acesso em 05/10/2021.

[4] Os documentos em questão foram apresentados por Frances Haugen ao Wall Street Journal, que os publicou, havendo a matéria desencadeado o seu depoimento perante o Senado americano, com vistas à apuração dos fatos.

[5] Fundação Getúlio Vargas. Diretoria de Análise de Políticas Públicas. Reação a boatos superou a difusão de informações contra Marielle no Twitter, aponta estudo da FGV DAPP. Disponível em http://dapp.fgv.br/reacao-boatos-superou-difusao-de-informacoes-contra-marielle-no-twitter-aponta-estudo-da-fgv-dapp/. Acesso em 02/10/2021.

Date Created

17/10/2021