DEMOCRACIA, MÍDIAS SOCIAIS E LIBERDADE DE EXPRESSÃO: ÓDIO, MENTIRAS E A BUSCA DA VERDADE POSSÍVEL[1]

Luís Roberto Barroso

Ministro do Supremo Tribunal Federal. Professor Titular da Universidade do Estado do Rio de Janeiro e do Centro Universitário de Brasília. Mestre pela Yale Law School. Doutor e Livre-Docente pela Universidade do Estado do Rio de Janeiro (1990). *Senior Fellow* na Harvard Kennedy School. Ex-Presidente do Tribunal Superior Eleitoral (2020-2022). *E-mail* institucional: gabmlrb@stf.jus.br.

Luna van Brussel Barroso

Mestre pela Yale Law School. Doutoranda na Universidade de São Paulo. Mestre em Direito Público pela Universidade do Estado do Rio de Janeiro. Graduada em Direito pela Fundação Getúlio Vargas. *E-mail* institucional: lbarroso@bfbm.com.br.

Resumo: Este artigo é uma reflexão crítica sobre o impacto da revolução digital e da internet em três tópicos que moldam o mundo contemporâneo: democracia, mídias sociais e liberdade de expressão. O Capítulo I estabelece alguns pressupostos históricos e conceituais sobre a democracia constitucional e discute o momento atual de recessão democrática, bem como o uso das plataformas digitais como estratégia de poder. O Capítulo II discute as plataformas de mídia social e a forma como revolucionaram a comunicação interpessoal e social, democratizando o acesso ao conhecimento e à informação, mas também levando a uma disseminação exponencial de desinformação, discursos de ódio e teorias conspiratórias. O Capítulo III propõe um quadro regulatório para a disciplina das plataformas digitais, sensível à importância de encontrar o equilíbrio certo com o direito fundamental contraposto, que é a liberdade de expressão, essencial para a dignidade humana, para a busca da verdade possível e para a democracia. O Capítulo IV destaca o papel da sociedade e a importância da educação midiática para criar um ambiente livre, mas positivo e construtivo na internet. Por fim, o Capítulo V comenta brevemente novos desenvolvimentos importantes quando o artigo já se encontrava em vias de publicação.

Palavras-chave: Democracia. Plataformas digitais. Mídias sociais. Liberdade de expressão. Autorregulação regulada. Constitucionalismo. Recessão democrática.

Abstract: This Essay is a critical reflection on the impact of the digital revolution and the internet on three topics that shape the contemporary world: democracy, social media, and freedom of expression. Chapter I establishes historical and conceptual assumptions about constitutional democracy and discusses the role

[1] Este texto foi publicado originariamente no *Chicago International Law Journal*, outono 2023, sob o título *Democracy, social media, and freedom of expression: hate, lies, and the search for the possible truth*. Tradução para o português feita com a colaboração de Matheus Verano.

of digital platforms in the current moment of democratic recession. Chapter II discusses how, while social media platforms have revolutionized interpersonal and social communication and democratized access to knowledge and information, they also have led to an exponential spread of mis- and disinformation, hate speech, and conspiracy theories. Chapter III proposes a framework that balances regulation of digital platforms with the countervailing fundamental right to freedom of expression, a right that is essential for human dignity, the search for the possible truth, and democracy. Chapter IV highlights the role of society and the importance of media education in the creation of a free, but positive and constructive, environment on the internet. Finally, Chapter V comments on new developments occurred after the article was ready for printing.

Keywords: Democracy. Digital platforms. Social media. Freedom of expression. Coregulation. Constitutionalism. Democratic recession.

Sumário: **I** Democracia e populismo autoritário – **II** Internet, mídias sociais e liberdade de expressão – **III** Um modelo regulatório para as redes sociais – **IV** O papel da sociedade – **V** Novos desenvolvimentos sobre o tema – **VI** Conclusão – Referências

Table of contents: **I** Democracy and Authoritarian Populism – **II** Internet, Social Media, and Freedom of Expression – **III** A Framework for the Regulation of Social Media – **IV** The role of society – **V** New developments – **VI** Conclusion

I Democracia e populismo autoritário

A democracia constitucional foi a ideologia que prevaleceu no século XX, na maior parte do planeta, superando os projetos alternativos que se apresentaram: comunismo, fascismo, nazismo, regimes militares e fundamentalismo religioso. O constitucionalismo democrático gira em torno de duas ideias principais que se fundiram no final do século XX. O *constitucionalismo*, herdeiro das revoluções liberais na Inglaterra, Estados Unidos da América e França, expressa as ideias de poder limitado, Estado de direito e respeito aos direitos fundamentais. A *democracia*, por sua vez, é o regime de soberania popular, eleições livres e justas e governo da maioria. Em muitos países, a democracia só se consolidou verdadeiramente ao longo do século XX, com o sufrágio universal garantido pelo fim das restrições à participação política baseada em condição social, religião, raça, sexo ou nível de educação.[2]

As democracias contemporâneas são feitas de votos, direitos e razões. Elas não se limitam à integridade dos processos eleitorais, mas exigem, também, o respeito pelos direitos fundamentais de todos os cidadãos e um debate público permanente que informa e legitima as decisões políticas.[3] Para garantir a proteção desses três elementos essenciais, a maioria dos regimes democráticos inclui

[2] BARROSO, Luís Roberto. O constitucionalismo democrático ou neoconstitucionalismo como ideologia vitoriosa do século XX. *Revista Publicum*, v. 4, 2018. p. 14.

[3] DWORKIN, Ronald. *Is democracy possible here?* Princeton: Princeton University Press, 2008. p. 12; DWORKIN, Ronald. *Taking rights seriously*. Cambridge: Harvard University Press, 1997. p. 181.

em sua estrutura constitucional uma suprema corte ou um tribunal constitucional com jurisdição para arbitrar as tensões inevitáveis que surgem entre democracia e constitucionalismo, ou seja, entre soberania popular e valores constitucionais.[4] Tais tribunais são, em última análise, as instituições responsáveis por proteger os direitos fundamentais e as regras do jogo democrático contra qualquer tentativa de abuso de poder por parte da maioria. Experiências recentes na Hungria, Polônia, Turquia, Venezuela e Nicarágua mostram que, quando falham em cumprir esse papel, a democracia entra em colapso ou sofre grandes retrocessos.[5]

Nos últimos anos, vários eventos desafiaram a prevalência do constitucionalismo democrático em muitas partes do mundo. Esse fenômeno tem sido caracterizado como recessão democrática,[6] retrocesso democrático,[7] constitucionalismo abusivo,[8] autoritarismo competitivo,[9] democracia iliberal,[10] legalismo autocrático,[11] entre outros. Mesmo democracias consolidadas enfrentaram momentos de turbulência e descrédito institucional,[12] à medida que o mundo testemunhou a ascensão de uma onda populista autoritária, antipluralista e anti-institucional que representa séria ameaça à democracia.

Populismo pode ser de direita ou de esquerda,[13] mas a onda recente tem sido caracterizada pela prevalência do extremismo de direita, frequentemente racista, xenófobo, misógino e homofóbico.[14] Enquanto no passado existia uma Internacional

---

4 BARROSO, Luís Roberto. O constitucionalismo democrático ou neoconstitucionalismo como ideologia vitoriosa do século XX. *Revista Publicum*, v. 4, 2018. p. 14.

5 ISSACHAROFF, Samuel. *Fragile democracies*: contested power in the era of constitutional courts. Cambridge: Cambridge University Press, 2015. p. 1.

6 DIAMOND, Larry. Facing up to the democratic recession. *Journal of Democracy*, v. 26, 2015. p. 141.

7 HUQ, Aziz; GINSBURG, Tom. How to lose a constitutional democracy. *UCLA Law Review*, v. 65, 2018. p. 91.

8 LANDAU, David. Abusive constitutionalism. *U.C. Davis Law Review*, v. 47, 2013. p. 189.

9 LEVITSKY, Steven; WAY, Lucan A. The rise of competitive authoritarianism. *Journal of Democracy*, v. 13, 2002. p. 51.

10 Aparentemente, o termo foi utilizado pela primeira vez por ZAKARIA, Fareed. The rise of illiberal democracies. *Foreign Affairs*, v. 76, n. 22, 1997.

11 SCHEPPELE, Kim Lane. Autocratic legalism. *University of Chicago Law Review*, v. 85, 2018. p. 545.

12 BALZ, Dan. A year after Jan. 6, are the guardrails that protect democracy real or illusory? *The Washington Post*, Washington, 6 jan. 2022. Disponível em: https://www.washingtonpost.com/politics/democracy-january-6/2022/01/06/2a1fc41e-6db4-11ec-a5d2-7712163262f0_story.html. Acesso em: 5 maio 2023; BREXIT: Reaction from around the UK. *BBC*, Londres, 24 jun. 2016. Disponível em: https://www.bbc.com/news/uk-politics-eu-referendum-36619444. Acesso em: 5 maio 2023.

13 MUDDE, Cas. *The populist zeitgeist*. Government and opposition. Cambridge: Cambridge University Press, 2004. v. 39. p. 541-544.

14 MUDDE, Cas. *The populist zeitgeist*. Government and opposition. Cambridge: Cambridge University Press, 2004. v. 39. p. 541; 544. Para uma discussão geral sobre o extremismo de direita na Índia, veja: SIYECH, Mohammed Sinan. An introduction to right-wing extremism in India. *New Eng. J. Pub. Pol.*, v. 1, 2021. p. 33. Para traçar a história do "Hindutva" e constatar que se tornou *mainstream* desde 2014 sob Modi, v. LEIDIG, Eviane. Hindutva as a variant of right-wing extremism. *Patterns of Prejudice*, v. 54, n. 3, p. 215-237, 2020. Para uma discussão do extremismo de direita no Brasil sob Bolsonaro, veja GOLDSTEIN, Ariel. Brazil leads the third wave of the Latin American far right. *C-REX – Center for Research on Extremism*, 1º mar. 2021.

Comunista, hoje é a extrema direita que tem uma grande rede global.[15] A marca do populismo de direita é a divisão da sociedade em nós – o povo puro, decente e conservador – e eles – as elites corruptas, liberais e cosmopolitas. O populismo autoritário decorre dos desvãos da democracia, das promessas não cumpridas de oportunidade e prosperidade para todos.[16] São muitos os fatores que levam a essa frustração democrática, dos quais se destacam três: *políticos* – as pessoas não se sentem representadas pelos sistemas eleitorais existentes, sentindo-se sem voz ou relevância –; *sociais* – pobreza, estagnação ou decréscimo de renda e aumento da desigualdade –; *cultural-identitários* – uma reação conservadora à agenda progressista de direitos humanos que prevaleceu nas últimas décadas, com a proteção dos direitos fundamentais de mulheres, afrodescendentes, minorias religiosas, *gays*, populações indígenas e meio ambiente.[17]

O populismo extremista autoritário adota, muitas vezes, estratégias semelhantes em diferentes partes do mundo, incluindo: a) comunicação direta com apoiadores, mais recentemente por meio das redes sociais; b) contorno ou cooptação das instituições intermediárias que fazem a interface entre o povo e o governo, como o Legislativo, a imprensa e a sociedade civil; e c) ataques às supremas cortes e aos tribunais constitucionais, bem como tentativas de capturá-los por meio da nomeação de juízes submissos.[18] Como o título do presente artigo sugere, uma das principais preocupações nessa temática é o uso de campanhas de desinformação, discursos de ódio, crimes contra a honra, mentiras e teorias conspiratórias para avançar esses objetivos antidemocráticos. Essas táticas ameaçam a democracia e as eleições livres e justas, porque enganam os eleitores, violam direitos fundamentais,

Disponível em: https://www.sv.uio.no/c-rex/english/news-and-events/right-now/2021/brazil-leads-the-third-wave-of-the-latin-american-.html. Acesso em: 5 maio 2023. Para uma discussão do extremismo de direita nos Estados Unidos sob Trump, veja JONES, Seth G. The rise of far-right extremism in the United States. *Center for Strategic & International Studies*, nov. 2018. Disponível em: https://www.csis.org/analysis/rise-far-right-extremism-united-states. Acesso em: 5 maio 2023.

15 FAUSTO, Sergio. O desafio democrático. *Revista Piauí*, v. 8, 2022. p. 191.

16 KUO, Ming-Sung. Against instantaneous democracy. *International Journal of Constitutional Law*, v. 17, p. 554-575, 2019. Disponível em: https://doi.org/10.1093/icon/moz029. Acesso em: 5 maio 2023. V. tb., ECPS – EUROPEAN CENTER FOR POPULISM STUDIES. *Digital Populism*. Disponível em: https://www.populismstudies.org/Vocabulary/digital-populism/. Acesso em: 5 maio 2023.

17 Sobre o assunto, v.: BARROSO, Luís Roberto. Technological revolution, democratic recession and climate change: the limits of law in a changing world. *International Journal of Constitutional Law*, v. 18, 2020. p. 334-349.

18 Para o uso das mídias sociais, v.: ENGESSER, Sven *et al*. Populism and social media: how politicians spread a fragmented ideology. *Information, Communication & Society*, v. 20, 2017. p. 1109. Sobre ataques à imprensa, v. WPFD 2021: attacks on press freedom growing bolder amid rising authoritarianism. *International Press Institute*, 30 abr. 2021. Disponível em: https://ipi.media/wpfd-2021-attacks-on-press-freedom-growing-bolder-amid-rising-authoritarianism/. Acesso em: 5 maio 2023. Para ataques ao Judiciário, v.: DICHO, Michael; LOGVINENKO, Igor. Authoritarian populism, courts and democratic erosion. *Just Security,* 11 fev. 2021. Disponível em: https://www.justsecurity.org/74624/authoritarian-populism-courts-and-democratic-erosion/. Acesso em: 5 maio 2023.

silenciam minorias e distorcem o debate público, minando os valores que justificam a proteção especial da liberdade de expressão. A "decadência da verdade" e a "polarização dos fatos" desacreditam as instituições e, consequentemente, fomentam a desconfiança na democracia.[19]

II Internet, mídias sociais e liberdade de expressão[20]

O mundo vive sob a égide da terceira revolução industrial, também conhecida como a revolução tecnológica ou digital.[21] Algumas de suas principais características são a massificação de computadores pessoais, a universalização dos telefones celulares inteligentes e, acima de tudo, a internet, conectando bilhões de pessoas no planeta. Um dos principais subprodutos da revolução digital e da internet foi o surgimento de plataformas de mídias sociais como o *Facebook*, *Instagram*, *YouTube*, *TikTok* e aplicativos de mensagens como o *WhatsApp* e *Telegram*. Vivemos em um mundo de *apps*, algoritmos, inteligência artificial e inovação em ritmo acelerado, onde nada parece realmente novo por muito tempo. Esse é o cenário em que se desenrola a narrativa a seguir.

1 O impacto da internet

A internet revolucionou o mundo da comunicação interpessoal e social, expandiu exponencialmente o acesso à informação e ao conhecimento e criou uma esfera pública em que qualquer um pode expressar ideias, opiniões e disseminar fatos.[22] Antes da internet, a participação no debate público dependia, principalmente, da imprensa profissional,[23] que investigava fatos, seguia padrões da técnica e da

[19] JACKSON, Vicki C. Knowledge institutions in constitutional democracies: reflections on the "press". *The Jorunal of Meida Law*, v. 14, 2022. p. 275. Disponível em: https://doi.org/10.1080/17577632.2022.2142733. Acesso em: 5 maio 2023.

[20] BARROSO, Luna van Brussel. *Liberdade de expressão e democracia na era digital*: o impacto das mídias sociais no mundo contemporâneo. Belo Horizonte: Fórum, 2022.

[21] A primeira revolução industrial é simbolizada pelo uso do vapor como fonte de energia, a partir do meio do século XVIII. A segunda teve início com o uso da eletricidade e a invenção do motor de combustão interna, na virada do século XIX para o XX. E já se fala da quarta revolução industrial, fruto da fusão de tecnologias, que está afetando as fronteiras entre as esferas física, digital e biológica. Sobre este último ponto, v. SCHWAB, Klaus. *A Quarta Revolução Industrial*. Tradução de Cássio Leite Vieira. São Paulo: Edipro, 2018. v. 1.

[22] MAGARIAN, Gregory P. A internet e as mídias sociais. *In*: STONE, Adrienne; SCHAUER, Frederick. *Liberdade de expressão*. Oxford: Oxford University Press, 2021. p. 350-368.

[23] WU, Tim. Is the first amendment obsolete? *In*: POZEN, David E. (Ed.). *The perilous public square*. N. York: Columbia University Press, 2020. *E-book Kindle*.

ética jornalística[24] e era responsável por danos se publicasse informações falsas, deliberadamente ou por negligência.[25] Havia controle editorial e responsabilidade civil relativamente à qualidade e à veracidade do que era publicado. Isso não significa que fosse um mundo perfeito. O número de meios de comunicação é limitado e nem sempre plural, empresas jornalísticas têm seus próprios interesses e nem todas distinguem com o cuidado necessário fato de opinião. Ainda assim, havia um grau mais refinado de controle sobre o que se tornava público, bem como consequências negativas pela publicação de notícias falsas ou discursos de ódio.

A internet, com o surgimento de *sites*, *blogs* pessoais e redes sociais, revolucionou esse universo. Criou comunidades *on-line* para disseminação de textos, imagens, vídeos e *links* gerados pelo usuário, publicados sem controle editorial e sem custo. Tais inovações amplificaram o número de pessoas que participam do debate público, diversificaram as fontes de informação e aumentaram exponencialmente o acesso a elas.[26] Essa nova realidade deu voz às minorias, à sociedade civil, aos políticos, aos agentes públicos, aos influenciadores digitais e permitiu que as demandas por igualdade e democracia adquirissem dimensões globais. Tudo isso representou uma poderosa contribuição para o dinamismo político e a resistência ao autoritarismo, e estimulou a criatividade, o conhecimento científico e as trocas comerciais.[27] Cada vez mais, as comunicações políticas, sociais e culturais relevantes ocorrem através desse meio.

No entanto, o surgimento das redes sociais também levou a um aumento exponencial na disseminação de discurso abusivo e criminoso. Embora essas plataformas não tenham criado desinformação, discursos de ódio ou discursos que atacam a democracia, a capacidade de publicar livremente, sem controle editorial e com pouca ou nenhuma responsabilidade, aumentou o uso dessas táticas. Além disso, e mais fundamentalmente, os modelos de negócio das plataformas agravaram o problema pela utilização de algoritmos que controlam e distribuem conteúdo *on-line*.

[24] A ética jornalística inclui a distinção entre fato e opinião, verificação da veracidade do que é publicado, não ter interesse próprio no assunto relatado, ouvir o outro lado e retificar erros. Para um exemplo de carta internacional de ética jornalística, v. GLOBAL Charter of Ethics for Journalists. *The International Federation of Journalists*, jun. 2019. Disponível em: https://perma.cc/7A2C-JD2S. Acesso em: 5 maio 2023.

[25] *E.g.*, *New York Times Co. v. Sullivan*, 376 U.S. 254, 1964.

[26] BALKIN, Jack M. Free speech is a triangle. *Columbia Law Review*, v. 118, n. 7, p. 2011-2056, 2018. Disponível em: https://columbialawreview.org/wp-content/uploads/2018/11/Balkin-FREE_SPEECH_IS_A_TRIANGLE.pdf. Acesso em: 5 maio 2023.

[27] MAGARIAN, Gregory P. The internet and social media. *In*: STONE, Adrienne; SCHAUER, Frederick (Ed.). *Freedom of speech*. Oxford: Oxford University Press, 2021. p. 350-368.

2 O papel dos algoritmos

A capacidade de participar e de ser ouvido no discurso público *on-line* é atualmente definida pelos algoritmos de moderação de conteúdo das grandes empresas de tecnologia. Embora as plataformas digitais tenham se apresentado inicialmente como espaços neutros, em que os usuários poderiam publicar livremente, elas na verdade desempenham funções legislativas, executivas e judiciais, pois (i) instituem unilateralmente as regras de discurso em seus termos e condições, (ii) definem, por seus algoritmos, como o conteúdo é distribuído e moderado e, por fim, (iii) decidem como essas regras são aplicadas.[28]

Especificamente, as plataformas digitais dependem de algoritmos para duas funções diferentes: recomendar e moderar conteúdo.[29] Primeiramente, um aspecto fundamental do serviço que oferecem envolve a curadoria do conteúdo disponível, de modo a proporcionar a cada usuário uma experiência personalizada e aumentar o tempo gasto *on-line*. Elas recorrem a algoritmos de *deep learning* que monitoram cada ação na plataforma, extraem dados e preveem qual conteúdo manterá um usuário específico engajado e ativo, com base em sua atividade anterior ou de usuários semelhantes.[30] A transição de um mundo de escassez de informação para um mundo de abundância de informação gerou uma concorrência acirrada pela *atenção* do usuário – esse, sim, o recurso escasso na era digital.[31] Portanto, o poder de modificar o ambiente informacional de uma pessoa tem um impacto direto no seu comportamento e nas suas crenças. E como os sistemas de IA podem rastrear o histórico *on-line* de um indivíduo, eles podem adaptar mensagens específicas para maximizar o impacto. Mais importante ainda, eles monitoram como o usuário interage com a mensagem personalizada, utilizando esse *feedback* para influenciar a segmentação de conteúdo futuro, tornando-se cada vez mais eficazes na moldagem de comportamentos.[32] Dado que os seres humanos se envolvem mais com conteúdo polarizador e provocativo, esses algoritmos acabam por provocar emoções fortes,

[28] KADRI, Thomas E.; KLONICK, Kate. Facebook v. Sullivan: public figures and newsworthiness in online speech. *Southern California Law Review*, v. 93, p. 37-99, 2019. p. 94. Disponível em: https://scholarship.law.stjohns.edu/faculty_publications/292/. Acesso em: 5 maio 2023.

[29] ELKIN-KOREN, Niva; PEREL, Maayan. Speech contestation by design: democratizing speech governance by AI. *Florida State University Law Review* [forthcoming]. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4129341https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4129341. Acesso em: 5 maio 2023.

[30] MESEROLE, Chris. How do recommender systems work on digital platforms? *Brookings*, 21 set. 2022. Disponível em: https://www.brookings.edu/techstream/how-do-recommender-systems-work-on-digital-platforms-social-media-recommendation-algorithms/. Acesso em: 5 maio 2023.

[31] SHAFFER, Kris. *Data versus democracy*: how big data algorithms shape opinions and alter the course of history. Colorado: Apress, 2019. p. 11-15.

[32] RUSSELL, Stuart. *Human compatible*: artificial intelligence and the problem of control. N. York: Penguin Books, 2019.

incluindo raiva.[33] O poder de organizar o conteúdo *on-line*, portanto, tem impactos diretos sobre a liberdade de expressão, o pluralismo e a democracia.[34]

Além dos sistemas de recomendação, as plataformas também dependem de algoritmos para a moderação de conteúdo, que consiste na prática de classificar o conteúdo para verificar se viola os padrões da comunidade.[35] Como mencionado, o crescimento das redes sociais e seu uso por pessoas ao redor do mundo permitiram a propagação da ignorância, mentiras e a prática de crimes de diferentes naturezas com pouco custo e quase nenhuma responsabilização, ameaçando a estabilidade até mesmo de democracias duradouras. Nesse cenário, tornou-se inevitável a criação e imposição de termos e condições que definem os valores e normas que cada plataforma deseja para sua comunidade digital e que pautarão a moderação do discurso.[36] Mas a quantidade potencialmente infinita de conteúdo publicado *on-line* significa que esse controle não pode ser exercido exclusivamente por seres humanos.

Algoritmos de moderação de conteúdo otimizam a varredura do material publicado *on-line* para identificar violações dos padrões da comunidade ou termos de serviço em escala e aplicar medidas que variam desde remoção até redução/amplificação do alcance ou inclusão de esclarecimentos ou referências a informações alternativas. As plataformas frequentemente dependem de dois modelos algorítmicos para moderação de conteúdo. O primeiro é o modelo de *detecção de reprodução*, que usa o *hashing*, uma tecnologia que atribui um ID único a textos, imagens e vídeos, para identificar reproduções idênticas de conteúdo previamente rotulado como indesejado.[37] O segundo sistema, o *modelo preditivo*, usa técnicas

[33] V. SHAFFER, Kris. *Data versus democracy*: how big data algorithms shape opinions and alter the course of history. Colorado: Apress, 2019. p. 11-15.

[34] Mais recentemente, com o avanço da neurociência, as plataformas aprimoraram sua capacidade de manipular e mudar nossas emoções, sentimentos e, consequentemente, nosso comportamento de acordo não com nossos próprios interesses, mas com os deles (ou daqueles a quem vendem este serviço). Nesse contexto, já se fala em um novo direito fundamental à liberdade cognitiva, à autodeterminação mental ou ao direito ao livre arbítrio.

[35] A moderação de conteúdo refere-se a "sistemas que classificam o conteúdo gerado pelo usuário com base em correspondência ou previsão, resultando em uma decisão e governança (por exemplo, remoção, bloqueio geográfico, suspensão de conta)" (ORWA, Robert; BINNS, Reuben; KATZENBACH, Christian. Moderação de conteúdo algorítmico: desafios técnicos e políticos na automação da governança de plataformas. *Big Data & Society*, v. 7, p. 1-15, 2020. Disponível em: https://journals.sagepub.com/doi/full/10.1177/2053951719897945. Acesso em: 7 maio 2023).

[36] BALKIN, Jack M. Free speech in the algorithmic society: big data, private governance, and new school speech regulation. *University of California*, Davis, v. 51, p. 1149-1210, 2018. Disponível em: https://lawreview.law.ucdavis.edu/issues/51/3/Essays/51-3_Balkin.pdf. Acesso em: 7 maio 2023.

[37] THAKUR, Dhanaraj; LLANSÓ, Emma. Do you see what I see? Capabilities and limits of automated multimedia content analysis. *Center for Democracy & Technology*, Washington, 20 maio 2021. Disponível em: https://cdt.org/insights/do-you-see-what-i-see-capabilities-and-limits-of-automated-multimedia-content-analysis/. Acesso em: 7 maio 2023.

de *machine learning* para identificar potenciais ilegalidades em conteúdo novo e não classificado.[38] O *machine learning* é um subtipo de inteligência artificial que depende de algoritmos treinados em vez de programados, capazes de aprender a partir de dados sem codificação explícita.[39] Embora úteis, ambos os modelos têm limitações.

O modelo de detecção de reprodução é ineficiente para conteúdo como discurso de ódio e desinformação, em que o potencial de novas e diferentes publicações é praticamente ilimitado e os usuários podem fazer alterações deliberadas para evitar a detecção.[40] O *modelo preditivo*, por sua vez, ainda é limitado em sua capacidade de lidar com situações às quais não foi exposto durante o treinamento, principalmente por uma incapacidade de entender significados e levar em conta considerações contextuais que influenciam a legitimidade do discurso.[41] Além disso, os algoritmos de *machine learning* também dependem de dados coletados do mundo real e podem incorporar preconceitos ou vieses, levando a aplicações assimétricas do filtro. E como os conjuntos de dados de treinamento são muito grandes, é difícil auditá-los para detectar essas falhas.

[38] THAKUR, Dhanaraj; LLANSÓ, Emma. Do you see what I see? Capabilities and limits of automated multimedia content analysis. *Center for Democracy & Technology*, Washington, 20 maio 2021. Disponível em: https://cdt.org/insights/do-you-see-what-i-see-capabilities-and-limits-of-automated-multimedia-content-analysis/. Acesso em: 7 maio 2023.

[39] WOOLDRIDGE, Michael. *A brief history of artificial intelligence*: what it is, where we are, and where we are going. New York: Flatiron Book, jan 2021.

[40] No entanto, essa tecnologia tem sido eficaz no combate à pornografia infantil, que muitas vezes envolve a reprodução de imagens repetidas, dada a dificuldade de produzir esse conteúdo do zero. As empresas de tecnologia mantêm um banco de dados compartilhado e, portanto, são capazes de lidar com esse material com relativa eficiência. Essa tecnologia também é frequentemente usada para conteúdo terrorista e de direitos autorais (BUCKMAN, Ian. Hashing it out: how an automated crackdown on child pornography is shaping the Fourth Amendment. *Berkeley Journal of Criminal Law*, Berkeley, 13 abr. 2021. Disponível em: https://www.bjcl.org/blog/hashing-it-out-how-an-automated-crackdown-on-child-pornography-is-shaping-the-fourth-amendment/. Acesso em: 7 maio 2023; FUSSEL, Sidney. Why the New Zealand shooting video keeps circulating. *The Atlantic*, 21 mar. 2019. Disponível em: https://www.theatlantic.com/technology/archive/2019/03/facebook-youtube-new-zealand-tragedy-video/585418/. Acesso em: 7 maio 2023).

[41] A compreensão da linguagem natural é prejudicada pela ambiguidade da linguagem, dependência contextual de palavras não imediatamente próximas, referências, metáforas e regras de semântica geral. A compreensão da linguagem de fato requer conhecimento ilimitado de senso comum sobre o mundo real, que os humanos possuem e é impossível de codificar (LARSON, Erik J. *The myth of artificial intelligence*: why computers can't think the way we do. [s.l.]: Belknap Press, abr. 2021). Um caso decidido pelo Conselho de Supervisão do Facebook ilustra o ponto: o filtro preditivo da empresa para combater pornografia removeu imagens de uma campanha de conscientização sobre câncer de mama, um conteúdo claramente legítimo e que não deveria ser alvo do algoritmo. No entanto, com base no treinamento prévio, o algoritmo removeu a publicação porque detectou pornografia e não conseguiu levar em consideração o contexto de que se tratava de uma campanha de saúde legítima (Facebook Oversight Board, Case 2020-004-IG-UA, Breast Cancer Symptoms and nudity. Disponível em: https://www.oversightboard.com/decision/IG-7THR3SI1. Acesso em: 7 maio 2023).

Apesar dessas limitações, os algoritmos continuarão a ser um recurso crucial no monitoramento de conteúdo, dada a escala das atividades *on-line*.[42] Somente nos últimos dois meses de 2020, o *Facebook* aplicou alguma medida de moderação de conteúdo a 105 milhões de publicações, e o *Instagram*, a 35 milhões. O *YouTube* tem 500 horas de vídeo carregadas por minuto e removeu mais de 9,3 milhões de vídeos. No primeiro semestre de 2020, o *Twitter* analisou reclamações relacionadas a 12,4 milhões de contas em potencial violação de suas regras e removeu 1,9 milhão.[43] Portanto, o monitoramento humano é impossível, e os algoritmos são uma ferramenta necessária para reduzir a disseminação de conteúdo ilícito e prejudicial. Responsabilizar as plataformas por erros ocasionais nesses sistemas criaria incentivos errados para abandonar os algoritmos na moderação de conteúdo, com a consequência negativa de aumentar significativamente a propagação do discurso indesejado. Por outro lado, reivindicações genéricas para que as plataformas implementem algoritmos para otimizar a moderação de conteúdo, ou leis que imponham prazos muito curtos para responder a solicitações de remoção enviadas pelos usuários, podem criar pressão excessiva para o uso desses sistemas imprecisos em uma escala maior. Reconhecer as limitações dessa tecnologia é fundamental para uma regulamentação precisa.

3 Algumas consequências indesejáveis

Um dos impactos mais marcantes deste novo ambiente informacional é o aumento exponencial na escala das comunicações sociais e na circulação de notícias. Ao redor do mundo, jornais, publicações impressas e estações de rádio têm alguns milhares de leitores e ouvintes.[44] A televisão atinge milhões de espectadores, embora diluídos em dezenas ou centenas de canais. Por outro lado, o

[42] DOUEK, Evelyn. Governing online speech. *Columbia Law Review*, v. 121, n. 3, 2021. Disponível em: https://columbialawreview.org/wp-content/uploads/2021/04/Douek-Governing_Online_Speech-from_Posts_As-Trumps_To_Proportionality_And_Probability.pdf. Acesso em: 7 maio 2023.

[43] DOUEK, Evelyn. Governing online speech. *Columbia Law Review*, v. 121, n. 3, 2021. Disponível em: https://columbialawreview.org/wp-content/uploads/2021/04/Douek-Governing_Online_Speech-from_Posts_As-Trumps_To_Proportionality_And_Probability.pdf. Acesso em: 7 maio 2023.

[44] MINOW, Martha. *Saving the press*: why the Constitution calls for government action to preserve freedom of speech. Oxford: Oxford University Press, 2021. p. 20. Por exemplo, o jornal mais vendido do mundo, *The New York Times*, encerrou o ano de 2022 com cerca de 10 milhões de assinantes, entre digitais e impressos (https://www.nytimes.com/2022/11/02/business/media/nyt-q3-2022-earnings.html). A revista *The Economist* teve aproximadamente 1,5 milhão em dados de 2019 (https://en.wikipedia.org/wiki/The_Economist). Em todo o mundo, são raras as publicações que atingem um milhão de assinantes (THESE are the most popular paid ubscription news websites. *Wordl Econ. F.*, 29 abr. 2021. Disponível em: https://perma.cc/L2MK-VPNX).

Facebook tem cerca de 3 bilhões de usuários ativos.[45] O *YouTube* tem 2,5 bilhões de contas.[46] O *WhatsApp*, mais de 2 bilhões.[47] Os números são desconcertantes. No entanto, como já assinalado, assim como democratizou o acesso ao conhecimento, à informação e ao espaço público, a revolução digital também introduziu consequências negativas que devem ser abordadas. São elas:

a) *o aumento da circulação de desinformação*, mentiras deliberadas, discursos de ódio, teorias conspiratórias, ataques à democracia e comportamentos inautênticos, potencializados por algoritmos de recomendação que otimizam o engajamento do usuário e algoritmos de moderação de conteúdo que ainda são incapazes de identificar adequadamente conteúdo indesejável;

b) *a tribalização da vida*, com a formação de câmaras de eco onde grupos falam apenas para si mesmos, reforçando o viés de confirmação,[48] tornando o discurso progressivamente mais radical e contribuindo para a polarização e intolerância;

c) *uma crise global no modelo de negócios da imprensa profissional*. Embora as plataformas de mídia social tenham se tornado uma das principais fontes de informação, elas não produzem seu próprio conteúdo. Elas contratam engenheiros, não repórteres, e seu interesse é o engajamento, não as notícias.[49] No entanto, com a migração da maior parte da publicidade para plataformas tecnológicas, a imprensa sofreu com a falta de receita, o que forçou centenas de publicações, nacionais e locais, a fechar as portas ou demitir jornalistas.[50] Mas imprensa livre e forte é vital para uma sociedade aberta e livre.

A imprensa profissional, tradicional e institucional é mais do que um negócio privado. Ela serve ao interesse público na busca pela verdade possível em um mundo plural e na disseminação de notícias, opiniões e ideias, condições indispensáveis para o exercício informado da cidadania. O conhecimento e a verdade – nunca absolutos, mas sinceramente buscados – são elementos essenciais para o funcionamento de uma democracia constitucional. Os cidadãos precisam compartilhar

[45] FACEBOOK statistics and trends. *Datareportal*, 19 fev. 2023. Disponível em: https://datareportal.com/essential-facebook-stats. Acesso em: 8 maio 2023.

[46] YOUTUBE User Statistic. *Global Media Insight*, 27 fev. 2023. Disponível em: https://www.globalmediainsight.com/blog/youtube-users-statistics/. Acesso em: 8 maio 2023.

[47] WHATSAPP 2023 user statistics: how many people use WhatsApp? *Backlinko*, 5 jan. 2023. Disponível em: https://backlinko.com/whatsapp-users. Acesso em: 8 maio 2023.

[48] Viés de confirmação (*confirmation bias*) é um obstáculo ao bom pensamento, pois busca apenas informações que correspondem ao que alguém já acredita.

[49] MINOW, Martha. *Saving the press*: why the Constitution calls for government action to preserve freedom of speech. Oxford: Oxford University Press, 2021. p. 49.

[50] MINOW, Martha. *Saving the press*: why the Constitution calls for government action to preserve freedom of speech. Oxford: Oxford University Press, 2021. p. 3; 11.

um conjunto mínimo de fatos objetivos comuns a partir dos quais formam os seus próprios juízos de valor. Se eles não puderem aceitar os mesmos fatos, o debate público se torna impossível. Intolerância e violência são produtos da incapacidade de se comunicar. Daí a importância das "instituições do conhecimento", como universidades, entidades de pesquisa e imprensa institucional. Sintomaticamente, em diferentes partes do mundo, a imprensa é um dos poucos negócios privados especificamente mencionados na Constituição. Apesar de sua importância para a sociedade e para a democracia, pesquisas revelam o declínio no prestígio do ensino superior e da imprensa.[51] Isso é preocupante.

No início da revolução digital, havia a crença de que a internet deveria ser um espaço livre, aberto e não regulado, tanto do ponto de vista econômico e comercial, quanto da perspectiva da liberdade de expressão. Com o tempo, surgiram preocupações de diferentes ordens, e a necessidade de regulação da internet gradualmente se tornou um consenso, com abordagens propostas em diferentes áreas,[52] incluindo: a) *econômica*, por meio de legislação antitruste, proteção ao consumidor, tributação justa e respeito aos direitos autorais; b) *privacidade*, por meio de leis que restringem a coleta de dados do usuário sem consentimento, para direcionamento de conteúdo ou comercialização; e c) *combate aos comportamentos inautênticos, controle de conteúdo e regras de responsabilidade da plataforma.*

Encontrar o equilíbrio adequado entre a indispensável preservação da liberdade de expressão, de um lado, e a repressão do conteúdo ilegal nas redes sociais, de outro, é um dos problemas mais complexos de nossa geração. A liberdade de expressão é um direito fundamental incorporado em praticamente todas as constituições contemporâneas e, em muitos países, é considerada uma liberdade preferencial, que deve prevalecer *prima facie* quando em confronto com outros valores. Várias razões procuram justificar a sua proteção especial, incluindo: (i) *a busca pela verdade possível* em uma sociedade aberta e plural; (ii) como *elemento essencial para a democracia*, pois permite a livre circulação de ideias, informações e pontos de vista que informam a opinião pública e o voto; e (iii) como *elemento essencial da dignidade humana*, permitindo a expressão da personalidade de cada pessoa.

[51] Sobre a importância do papel da imprensa como instituição de interesse público e sua "relação crucial" com a democracia, v. MINOW, Martha. *Saving the press*: why the Constitution calls for government action to preserve freedom of speech. Oxford: Oxford University Press, 2021. p. 35. Sobre a imprensa como uma "instituição de conhecimento", a ideia de "imprensa institucional" e dados sobre a perda de prestígio de jornais e estações de televisão, v. JACKSON, Vicki C. Knowledge institutions in constitutional democracies: reflections on the "press". *The Jorunal of Meida Law*, v. 14, 2022. p. 280 e ss. Disponível em: https://doi.org/10.1080/17577632.2022.2142733. Acesso em: 5 maio 2023.

[52] BALKIN, Jack M. How to regulate (and not regulate) social media. *Journal of Free Speech Law*, v. 71, 2021; *Knight Institute Occasional Paper Series*, n. 1, March 2020; *Yale Law School, Public Law Research Paper Forthcoming*, 20 nov. 2019. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3484114. Acesso em: 7 maio 2023.

A regulação das plataformas digitais não pode comprometer esses valores. Pelo contrário, deve visar à sua proteção e fortalecimento. No entanto, na era digital, esses mesmos valores que historicamente justificaram a proteção reforçada da liberdade de expressão agora podem justificar a sua regulação. Como o Secretário-Geral da ONU, António Guterres, registrou com propriedade: "A capacidade de promover desinformação em larga escala e minar fatos cientificamente estabelecidos é um risco existencial para a humanidade".[53]

Dois aspectos do modelo de negócio da internet são particularmente problemáticos. O primeiro é que, embora o acesso à maioria das plataformas e aplicativos tecnológicos seja gratuito, os usuários pagam pelo acesso com sua privacidade.[54] Como Lawrence Lessig observou, assistimos à televisão, mas a internet assiste a nós. Tudo o que fazemos *on-line* é monitorado e monetizado. Os dados são o novo ouro.[55] O segundo aspecto é que os algoritmos são programados para maximizar o tempo gasto *on-line*, o que muitas vezes leva à amplificação de conteúdo provocativo, radical e agressivo. Isso compromete a liberdade de expressão, porque, ao visar ao engajamento, os algoritmos sacrificam a busca pela verdade – com a ampla circulação de *fake news* –, a democracia – com ataques às instituições e defesa de golpes e autoritarismo – e a dignidade humana – com ofensas, ameaças, racismo e outros. A busca por atenção e engajamento para obter receita nem sempre é compatível com os valores que sustentam a proteção da liberdade de expressão.

III Um modelo regulatório para as redes sociais

Os modelos de regulação de plataformas podem ser amplamente classificados em três categorias. A primeira é a (a) *regulação estatal ou governamental*, por meio de legislação e regras que criam um arcabouço obrigatório e abrangente; (b) *autorregulação*, por meio de regras elaboradas pelas próprias plataformas e materializadas em seus termos de uso; e (c) *autorregulação regulada ou corregulação*, por meio de padrões fixados pelo Estado, mas com flexibilidade das plataformas em materializá-los e implementá-los. Este artigo defende o terceiro modelo, com uma combinação adequada de responsabilidades governamentais e privadas. O cumprimento das regras deve ser supervisionado por um comitê independente,

[53] A GLOBAL dialogue to guide regulation worldwide. *Unesco*, 2023. Disponível em: https://www.unesco.org/en/internet-conference. Acesso em: 8 maio 2023.

[54] BEYER, R. Can we fix what's wrong with social media? *Yale Law Report*, verão 2022.

[55] LESSIG, Lawrence. *They don't represent us*: reclaiming our democracy. Providence: Dey Street Books, 2019. p. 105.

com minoria de representantes do governo e maioria de representantes do setor empresarial, academia, entidades de tecnologia, usuários e sociedade civil.

O quadro regulatório deve visar à redução da assimetria de informações entre as plataformas e os usuários, salvaguardar o direito fundamental à liberdade de expressão de intervenções privadas ou estatais indevidas e proteger e fortalecer a democracia. As limitações técnicas atuais dos algoritmos de moderação de conteúdo exploradas acima e a discordância substancial sobre o que deve ser considerado ilegal ou prejudicial trazem uma implicação inevitável: o objetivo da regulamentação deve ser o de encontrar um modelo capaz de otimizar o equilíbrio entre os direitos fundamentais dos usuários e das plataformas, reconhecendo que sempre haverá casos em que o consenso é inatingível. O foco da regulamentação deve ser o desenvolvimento de procedimentos adequados para a moderação de conteúdo, capazes de minimizar erros e legitimar decisões, mesmo quando alguém discorda do resultado substantivo.[56] Com essas premissas como pano de fundo, a proposta de regulação formulada aqui é dividida em três níveis: (i) o modelo apropriado de responsabilidade intermediária para conteúdo gerado pelo usuário; (ii) deveres procedimentais para a moderação de conteúdo; e (iii) deveres mínimos para moderar conteúdo que represente ameaças concretas à democracia e/ou à liberdade de expressão em si.

1 Responsabilidade intermediária por conteúdo gerado pelo usuário

Existem três regimes principais de responsabilidade da plataforma pelo conteúdo de terceiros. Nos modelos de *responsabilidade objetiva*, as plataformas são responsáveis por todas as postagens geradas pelos usuários. Como as plataformas não têm controle editorial sobre o que é postado e não têm condições materiais de supervisionar milhões de postagens feitas diariamente, esse regime seria potencialmente destrutivo e, por isso, não foi adotado por nenhuma democracia. No modelo de *responsabilidade subjetiva após notificação extrajudicial*, a responsabilidade das plataformas surgiria se elas não agissem para remover o conteúdo após uma notificação extrajudicial dos usuários. Por fim, na *responsabilidade subjetiva*

[56] DOUEK, Evelyn. Governing online speech. *Columbia Law Review*, v. 121, n. 3, 2021. p. 791. Disponível em: https://columbialawreview.org/wp-content/uploads/2021/04/Douek-Governing_Online_Speech-from_Posts_As-Trumps_To_Proportionality_And_Probability.pdf. Acesso em: 7 maio 2023; ZITTRAIN, Jonathan. Answering impossible questions: content governance in an age of disinformation. *Harvard Kennedy School – Misinformation Review*, 4 jan. 2020. Disponível em: https://misinforeview.hks.harvard.edu/article/content-governance-in-an-age-of-disinformation/. Acesso em: 8 maio 2023.

após decisão judicial, as plataformas seriam responsáveis pelo conteúdo postado pelos usuários somente em caso de não conformidade com uma ordem judicial de remoção do conteúdo. Este último modelo foi adotado no Brasil com o Marco Civil da Internet. A única exceção na legislação brasileira a essa regra geral é a chamada *pornografia de vingança*:[57] se houver violação da intimidade resultante da divulgação, sem consentimento dos participantes, de imagens, vídeos ou outros materiais contendo nudez privada ou atos sexuais privados, a notificação extrajudicial é suficiente para criar uma obrigação de remoção do conteúdo sob pena de responsabilidade.

Em nossa opinião, a regra geral prevista no modelo brasileiro, embora possa comportar exceções, é a que equilibra mais adequadamente os direitos fundamentais envolvidos.[58] Como mencionado, nos casos mais complexos relacionados à liberdade de expressão, as pessoas vão discordar sobre a legalidade do discurso. Regras que responsabilizam as plataformas por não remover o conteúdo após uma simples notificação do usuário criam incentivos para a remoção excessiva de qualquer conteúdo potencialmente controverso, restringindo excessivamente a liberdade de expressão dos usuários. Ou seja: haveria um incentivo para remover todo o conteúdo que ofereça risco de ser considerado ilícito pelos tribunais para evitar a responsabilidade,[59] criando um ambiente de autocensura.

No entanto, esse regime de responsabilidade deve coexistir com uma estrutura regulatória mais ampla impondo princípios, limites e deveres à moderação de conteúdo pelas plataformas digitais, tanto para aumentar sua legitimidade na aplicação de seus próprios termos e condições, quanto para minimizar os impactos potencialmente devastadores de discursos ilícitos ou prejudiciais.

---

[57] "Art. 21. O provedor de aplicações de internet que disponibilize conteúdo gerado por terceiros será responsabilizado subsidiariamente pela violação da intimidade decorrente da divulgação, sem autorização de seus participantes, de imagens, de vídeos ou de outros materiais contendo cenas de nudez ou de atos sexuais de caráter privado quando, após o recebimento de notificação pelo participante ou seu representante legal, deixar de promover, de forma diligente, no âmbito e nos limites técnicos do seu serviço, a indisponibilização desse conteúdo. Parágrafo único. A notificação prevista no caput deverá conter, sob pena de nulidade, elementos que permitam a identificação específica do material apontado como violador da intimidade do participante e a verificação da legitimidade para apresentação do pedido".

[58] Em pronunciamento na Conferência Global da Unesco "Por uma Internet de Confiança", em 23.2.2023, o primeiro autor defendeu as ideias a seguir. No caso de comportamentos criminosos, as plataformas devem remover os conteúdos ilícitos de ofício, isto é, independentemente de provocação. Em casos de clara violação de direitos, como compartilhamento de fotos íntimas sem autorização e violação de direitos autorais, entre outras, as plataformas devem remover o conteúdo imediatamente após a notificação da parte interessada. Nos demais casos, sobretudo onde possa haver dúvida razoável, a remoção deve se dar após a primeira ordem judicial.

[59] BALKIN, Jack M. Free speech is a triangle. *Columbia Law Review*, v. 118, n. 7, p. 2011-2056, 2018. Disponível em: https://columbialawreview.org/wp-content/uploads/2018/11/Balkin-FREE_SPEECH_IS_A_TRIANGLE.pdf. Acesso em: 5 maio 2023.

2 Regras para moderação de conteúdo pelas plataformas

As plataformas têm liberdade de iniciativa e de expressão para definir suas próprias regras e decidir o tipo de ambiente que desejam criar, bem como moderar conteúdo prejudicial que poderia afastar os usuários. No entanto, porque esses algoritmos de moderação de conteúdo são os novos governantes da esfera pública[60] e definem a capacidade de participar e ser ouvido no discurso público *on-line*, as plataformas devem atender a deveres procedimentais mínimos de transparência, auditoria, devido processo e isonomia.

a. Transparência e auditoria

As medidas de transparência e auditoria têm como principal objetivo garantir que as plataformas sejam responsabilizáveis (*accountable*) pelas decisões de moderação de conteúdo e pelos impactos de seus algoritmos. Elas fornecem aos usuários um maior entendimento e conhecimento sobre a intensidade com que as plataformas regulam o discurso, e dão aos órgãos de supervisão e aos pesquisadores informações para entender as ameaças advindas dos serviços digitais e o papel das plataformas em amplificá-las ou minimizá-las.

Impulsionado pelas demandas da sociedade civil, várias plataformas digitais já publicam relatórios de transparência. No entanto, a falta de normas vinculativas significa que esses relatórios têm lacunas relevantes, inexistindo verificação independente das informações fornecidas,[61] tampouco padronização entre as plataformas, o que impede a análise comparativa.[62] Nesse contexto, iniciativas regulatórias que imponham requisitos e padrões mínimos são cruciais para tornar a supervisão mais eficaz. Por outro lado, critérios de transparência excessivamente amplos podem forçar as plataformas a adotarem regras de moderação de conteúdo mais simples para reduzir custos, com impacto negativo na precisão da moderação de conteúdo ou na qualidade da experiência do usuário.[63] Uma abordagem escalonada para a transparência, em que certas informações são públicas e outras informações

[60] KLONICK, Kate. The new governors: the people, rules, and processes governing online speech. *Harvard Law Review*, v. 131, p. 1598-1670, 2018. Disponível em: https://harvardlawreview.org/2018/04/the-new-governors-the-people-rules-and-processes-governing-online-speech/. Acesso em: 7 maio 2023.

[61] HUMAN RIGHTS COMMITTEE. *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*. UN Doc A/HRC/32/38. 11 maio 2016. Disponível em: https://undocs.org/en/A/HRC/32/38. Acesso em: 8 maio 2023.

[62] LEERSSEN, Paddy. The soap box as a black box: regulating transparency in social media recommender systems. *European Journal of Law and Technology*, v. 11, 2020. Disponível em: https://ssrn.com/abstract=3544009. Acesso em: 7 maio 2023.

[63] KELLER, Daphne. Some humility about transparency. *The Center for Internet and Society Blog*, 19 mar. 2021. Disponível em: https://cyberlaw.stanford.edu/blog/2021/03/some-humility-about-transparency. Acesso em: 7 maio 2023.

são limitadas a órgãos de supervisão ou pesquisadores previamente qualificados, garante proteção adequada a interesses contrapostos, como privacidade do usuário e confidencialidade empresarial.[64] O *Digital Services Act*, aprovado pela União Europeia em 16.11.2022, contém disposições robustas de transparência que, no geral, estão alinhadas com essas considerações.[65]

As informações que devem ser publicamente fornecidas incluem, entre outras coisas, termos de uso claros e inequívocos, as sanções disponíveis para lidar com violações (remoção, redução de amplificação, esclarecimentos, suspensão de conta etc.) e a divisão de trabalho entre algoritmos e humanos. Mais importante ainda, os relatórios públicos de transparência devem incluir informações sobre a precisão das medidas de moderação automatizada e o número de ações de moderação de conteúdo desagregadas por tipo (remoção, bloqueio, exclusão de conta etc.).[66] Também deve haver obrigações de transparência para pesquisadores, dando-lhes acesso a informações e estatísticas cruciais, incluindo o conteúdo analisado para as decisões de moderação de conteúdo.[67]

Embora valiosos, os requisitos de transparência são insuficientes para promover a responsabilização adequada porque dependem de usuários e pesquisadores para monitorar ativamente a conduta da plataforma e pressupõem que eles tenham o poder de chamar a atenção para falhas e promover mudanças.[68] A auditoria algorítmica por terceiros é, portanto, um complemento importante para garantir que esses modelos satisfaçam padrões legais, éticos e de segurança, assim como para deixar claras as ponderações feitas, como entre a segurança do usuário e a

[64] MACCARTHY, Mark. Transparency requirements for digital social media platforms: recommendations for policy makers and industry. *Transatlantic Working Group*, 24 jun. 2020. Disponível em: https://ssrn.com/abstract=3615726. DOI: http://dx.doi.org/10.2139/ssrn.3615726. Acesso em: 8 maio 2023.

[65] O Ato de Serviços Digitais – DSA (promulgado juntamente com o Ato de Mercados Digitais – DMA) foi aprovado pelo Parlamento europeu em 5.7.2022, e em 4.10.2022 o Conselho europeu deu sua aprovação final à regulamentação. O DSA aumenta a transparência e a responsabilidade das plataformas, fornecendo, por exemplo, a obrigação de "informações claras sobre moderação de conteúdo ou o uso de algoritmos para recomendar conteúdo (os chamados sistemas de recomendação); os usuários poderão contestar decisões de moderação de conteúdo" (Disponível em: https://www.europarl.europa.eu/news/en/press-room/20220701IPR34364/digital-services-landmark-rules-adopted-for-a-safer-open-online-environment).

[66] MACCARTHY, Mark. Transparency requirements for digital social media platforms: recommendations for policy makers and industry. *Transatlantic Working Group*, 24 jun. 2020. Disponível em: https://ssrn.com/abstract=3615726. DOI: http://dx.doi.org/10.2139/ssrn.3615726. Acesso em: 8 maio 2023.

[67] Nesse sentido, o professor da Universidade de Stanford, Nathaniel Persily, apresentou recentemente um projeto de lei ao Congresso americano propondo um modelo para conduzir pesquisas sobre os impactos das comunicações digitais de maneira que proteja a privacidade do usuário. O projeto exige que as plataformas digitais compartilhem dados com pesquisadores previamente autorizados pela Comissão Federal de Comércio (FTC), e divulguem publicamente certos dados sobre conteúdo, algoritmos e publicidade (Disponível em: https://www.coons.senate.gov/imo/media/doc/text_pata_117.pdf. Acesso em: 8 maio 2023).

[68] NAHMIAS, Yifat; PEREL, Maayan. The oversight of content moderation by AI: impact assessment and their limitations. *Harvard Journal on Legislation*, v. 58, 2021. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3565025. Acesso em: 8 maio 2023.

liberdade de expressão.[69] Como ponto de partida, as auditorias de algoritmos devem considerar questões como sua precisão, qualquer viés ou discriminação potencial incorporada nos dados e em que medida as mecânicas internas são explicáveis para humanos.[70] O *Digital Services Act* contém uma proposta semelhante.[71]

O mercado de auditoria algorítmica ainda é emergente e cheio de incertezas. Ao tentar navegar nesse cenário, os reguladores devem: (i) definir com que frequência as auditorias devem ocorrer; (ii) desenvolver padrões e melhores práticas para os procedimentos de auditoria; (iii) obrigar a divulgação específica de determinados dados para assegurar que os auditores terão acesso aos dados necessários; e (iv) definir como os danos identificados devem ser abordados.[72]

b. Devido processo legal e razoabilidade (*fairness*)

Para garantir o devido processo legal, as plataformas devem informar aos usuários afetados pelas decisões de moderação de conteúdo qual a cláusula dos termos de uso supostamente violada, além de oferecer um sistema interno de recursos contra essas decisões. As plataformas também devem criar sistemas que permitam a denúncia fundamentada de conteúdo ou contas por outros usuários, e notificar os usuários denunciantes da decisão tomada.

Quanto à razoabilidade (i.e., critérios básicos de justiça das decisões), as plataformas devem garantir que as regras sejam aplicadas de maneira igualitária a todos os usuários. Embora seja admissível que as plataformas adotem critérios diferentes para pessoas públicas ou informações de interesse público, essas exceções devem estar claras nos termos de uso. Esse problema tem sido objeto de controvérsia entre o Comitê de Supervisão do *Facebook* e a empresa.[73]

---

[69] AUDITING Algorithms: the existing landscape, role of regulator and future outlook. *Digital Regulation Cooperation Forum*, 23 set. 2022. Disponível em: https://www.gov.uk/government/publications/auditing-algorithms-the-existing-landscape-role-of-regulators-and-future-outlook. Acesso em: 7 maio 2023.

[70] KOSHIYAMA, Adriano; KAZIM, Emre; TRELEAVEN, Philip. Algorithm auditing: managing the legal, ethical, and technological risks of artificial intelligence, machine learning, and associated algorithms. *IEEE Transactions on Technology and Society*, v. 3, p. 128-142, abr. 2022. Disponível em: https://ieeexplore.ieee.org/document/9755237. Acesso em: 8 maio 2023.

[71] No Artigo 37, o DSA estabelece que plataformas digitais de determinado tamanho devem ser responsáveis, por meio de auditoria independente anual, pelo cumprimento das obrigações estabelecidas na regulamentação e por quaisquer compromissos assumidos de acordo com códigos de conduta e protocolos de crise.

[72] AUDITING Algorithms: the existing landscape, role of regulator and future outlook. *Digital Regulation Cooperation Forum*, 23 set. 2022. Disponível em: https://www.gov.uk/government/publications/auditing-algorithms-the-existing-landscape-role-of-regulators-and-future-outlook. Acesso em: 7 maio 2023.

[73] Em um relatório de transparência publicado ao final de seu primeiro ano de operação, o Facebook Oversight Board (FOB) destacou a inadequação das explicações apresentadas pelo Meta sobre a operação de um sistema conhecido como *cross-check*, que aparentemente dava a alguns usuários maior liberdade na plataforma. Em janeiro de 2022, o Meta explicou que o sistema *cross-check* concede um grau adicional de revisão a determinados conteúdos que os sistemas internos marcam como violando os termos de

Devido à enorme quantidade de conteúdo publicado nas plataformas e à inevitabilidade do uso de mecanismos automatizados para moderação de conteúdo, as plataformas não devem ser responsabilizadas por uma violação desses deveres em casos específicos, mas somente quando a análise revelar uma falha sistemática no cumprimento.[74]

c. Deveres mínimos para moderar conteúdo ilícito

O quadro regulamentar também deve conter obrigações específicas para lidar com certos tipos de discurso especialmente prejudiciais. As seguintes categorias são consideradas pertencentes a este grupo: (a) desinformação, (b) discurso de ódio, (c) ataques antidemocráticos, (d) *cyberbullying*, (e) terrorismo e (f) pornografia infantil. É certo que definir e identificar o discurso incluído nessas categorias – exceto no caso da pornografia infantil, naturalmente – é uma tarefa difícil e amplamente subjetiva. Precisamente por esse motivo, as plataformas devem ser livres para definir como os conceitos serão operacionalizados, desde que guiados pelas normas internacionais de direitos humanos e de maneira transparente. Isso não significa que todas as plataformas chegarão às mesmas definições nem aos mesmos resultados substantivos em casos concretos, por valorações diferentes e pela impossibilidade de consenso. No entanto, a obrigação de observar parâmetros internacionais de direitos humanos reduz a discricionariedade das empresas, ao mesmo tempo em que permite a diversidade de políticas entre elas. Após definir essas categorias, as plataformas devem estabelecer mecanismos que permitam aos usuários denunciar violações.

Além disso, as plataformas também devem desenvolver mecanismos para lidar com comportamentos inautênticos coordenados, que envolvem o uso de sistemas automatizados ou meios enganosos para amplificar artificialmente mensagens falsas ou perigosas, usando *bots*, perfis falsos, *trolls* e provocadores.[75] Por exemplo: se

uso da plataforma. O Meta submeteu uma consulta à FOB sobre como melhorar o funcionamento desse sistema e a FOB fez recomendações relevantes. Mais informações em: https://www.oversightboard.com/news/501654971916288-oversight-board-publishes-policy-advisory-opinion-on-meta-s-cross-check-program/.

74 DOUEK, Evelyn. Content moderation as systems thinking. *Harvard Law Review*, v. 2, p. 136, 2022. Disponível em: https://harvardlawreview.org/2022/12/content-moderation-as-systems-thinking/. Acesso em: 8 maio 2023.

75 O Facebook define comportamento coordenado inautêntico como "o uso de múltiplos recursos do Facebook ou do Instagram, trabalhando em conjunto para se engajar em comportamento inautêntico, onde o uso de contas falsas é central para a operação". Comportamento inautêntico é definido como "o uso de recursos do Facebook ou do Instagram (contas, Páginas, Grupos ou Eventos), para enganar as pessoas ou o Facebook: (i) Sobre a identidade, propósito ou origem da entidade que eles representam; (ii) Sobre a popularidade do conteúdo ou recursos do Facebook ou do Instagram; (iii) Sobre o propósito de uma audiência ou comunidade; (iv) Sobre a fonte ou origem do conteúdo; ou (v) Para evitar a aplicação das nossas Normas da Comunidade" (Disponível em: https://transparency.fb.com/policies/community-standards/inauthentic-behavior/).

uma pessoa publicar uma postagem dizendo que querosene é bom para curar a Covid-19 e essa mensagem alcançar seus vinte seguidores, é ruim, mas o efeito é limitado. Contudo, se essa mensagem for amplificada para milhares de usuários, haverá um problema de saúde pública. Ou, em outro exemplo, se a mensagem falsa de que as eleições foram fraudadas alcançar milhões de pessoas, há um risco democrático devido à perda de credibilidade nas instituições.

O papel dos órgãos de supervisão deve ser verificar se as plataformas adotaram termos de uso que proíbam o compartilhamento dessas categorias de discurso e garantir que os sistemas de recomendação e moderação de conteúdo estejam treinados para moderar esse conteúdo.

IV O papel da sociedade

Apesar da importância da ação regulatória, a responsabilidade pela preservação da internet como uma esfera pública saudável reside, acima de tudo, nos cidadãos. A educação midiática e a conscientização dos usuários são etapas fundamentais para a criação de um ambiente livre, mas positivo e construtivo na rede mundial de computadores. Os cidadãos devem estar cientes de que as redes sociais podem ser injustas e perversas, violar direitos fundamentais e regras básicas da democracia. Eles devem estar atentos para não passar informações recebidas sem questionamento crítico. Nas palavras de Jonathan Haidt,[76] "[q]uando nossa esfera pública é governada pela dinâmica da multidão, sem mínima observância do devido processo legal, o resultado não é justiça e inclusão; mas, ao contrário, uma sociedade que ignora o contexto, a proporcionalidade, a misericórdia e a verdade". Os cidadãos são a força mais importante para lidar com essas ameaças.

V Novos desenvolvimentos sobre o tema

1 Estados Unidos: *Twitter v.* Taamneh e Gonzalez *v. Google*

Em maio de 2023, a Suprema Corte norte-americana decidiu dois casos relacionados à responsabilização de plataformas digitais. O primeiro, *Twitter v.* Taamneh,[77]

[76] HAIDT, Jonathan. Why the past 10 years of American life have been uniquely stupid. *The Atlantic*. Disponível em: https://www.theatlantic.com/magazine/archive/2022/05/social-media-democracy-trust-babel/629369/. Acesso em: 8 maio 2023. Tradução livre e ligeiramente editada.

[77] *Twitter, Inc. v. Taamneh*, 598 US ___ (2023).

discutiu a responsabilidade do *Facebook*, do *Twitter* e da *Google* por um ataque terrorista executado pelo Estado islâmico em Istambul no ano de 2017. A família de uma das vítimas fatais processou os réus alegando que eles teriam conhecimento do uso de suas plataformas pela organização terrorista e teriam falhado ao não impedir essas atividades, em violação a dispositivo da Lei de Antiterrorismo (18 U.S.C. §2333(d)(2)).[78] Os autores alegaram ainda que os algoritmos de recomendação das plataformas teriam facilitado as atividades de recrutamento, financiamento e propaganda do Estado islâmico, e que os réus teriam se beneficiado financeiramente de arrecadações publicitárias incluídas nesse material.

Em decisão unânime, a Suprema Corte rejeitou a alegação, concluindo que a mera disponibilização de plataforma digital com algoritmos que recomendam conteúdo a partir de *inputs* e histórico de usuários não caracteriza, por si só, conduta ilícita. A Corte entendeu que a relação dos réus com o Estado islâmico era igual à mantida com todos os demais usuários: impessoal, passiva e indiferente. Os algoritmos de recomendação são agnósticos quanto ao conteúdo recomendado, sendo influenciados exclusivamente por dados coletados dos usuários, de modo que a Corte entendeu que os requerentes não demonstraram ação deliberada ou vontade consciente de favorecer especificamente a organização terrorista. A capacidade do Estado islâmico de se beneficiar dessas plataformas foi considerada meramente incidental aos serviços prestados e ao modelo de negócio dos réus.

No segundo caso, Gonzalez *v. Google*,[79] discutia-se igualmente a responsabilidade da *Google* pela morte de uma cidadã americana em um atentado terrorista ocorrido em Paris. Os autores da ação, irmãos da vítima, alegaram que a Google seria responsável direta e subsidiariamente pelo ataque terrorista por ter permitido o uso de sua plataforma *YouTube* por integrantes do Estado islâmico. A Corte, novamente de forma unânime, considerou que a resolução desse caso deveria ser idêntica à conferida ao caso *Twitter v*. Taamneh. Na decisão de apenas três páginas, porém, ressalvou que nenhum desses dois casos foi proposto para discutir o dispositivo

[78] "In an action under subsection (a) for an injury arising from an act of international terrorism committed, planned, or authorized by an organization that had been designated as a foreign terrorist organization under section 219 of the Immigration and Nationality Act (8 U.S.C. 1189), as of the date on which such act of international terrorism was committed, planned, or authorized, liability may be asserted as to any person who aids and abets, by knowingly providing substantial assistance, or who conspires with the person who committed such an act of international terrorism". Tradução livre: "Em uma ação processada sob a subseção (a) por dano decorrente de um ato de terrorismo internacional cometido, planejado ou autorizado por uma organização designada como uma organização terrorista estrangeira nos termos da seção 219 da Lei de Imigração e Nacionalidade (8 U.S.C. 1189), a partir da data em que tal ato de terrorismo internacional foi cometido, planejado ou autorizado, a responsabilidade pode ser reconhecida em relação a qualquer pessoa que ajude e incite, conscientemente fornecendo assistência substancial, ou que conspire com a pessoa que cometeu tal ato de terrorismo internacional".

[79] *Gonzalez v. Google LLC*, 598 U.S. ___ (2023).

legal que confere às plataformas imunidade por conteúdo publicado por terceiros, deixando aberta a possibilidade de revisão judicial desse modelo de responsabilidade civil, atualmente previsto na Seção 230 do *Communications Decency Act*.[80]

2 Brasil: não votação do PL nº 2.630

Em maio de 2020, o Senado Federal iniciou discussões sobre o Projeto de Lei nº 2.630/2020, que institui a Lei Brasileira de Liberdade, Responsabilidade e Transparência na Internet. A versão final aprovada no Senado em 30.6.2020 e remetida à Câmara dos Deputados estabelece normas sobre transparência para provedores de redes sociais e serviços de mensageria privada com dois milhões ou mais de usuários registrados no Brasil. Em abril de 2023, depois de quase três anos aguardando votação na Câmara, o relator, Deputado Orlando Silva, apresentou novo texto e foi aprovado um regime de urgência com previsão de votação em 2.5.2023. Não obstante, no dia previsto para votação, o relator pediu a retirada de pauta, alegando falta de tempo hábil para examinar todas as sugestões recebidas quanto à nova versão do projeto. Entre os pontos mais controvertidos estão a definição da autoridade responsável pela fiscalização da lei e o compartilhamento de receitas de publicidade com entidades jornalísticas. Desde então, o PL novamente perdeu força.

3 Brasil: declaração de inelegibilidade do Ex-Presidente Jair Bolsonaro

Em 30.6.2023, o Tribunal Superior Eleitoral declarou a inelegibilidade do Ex-Presidente Jair Bolsonaro por 8 anos por abuso de poder político e uso indevido dos meios de comunicação.[81] A condenação teve como fundamento reunião realizada no Palácio da Alvorada com embaixadores no dia 18.7.2022, na qual o ex-presidente fez campanha eleitoral direcionada aos seus eleitores atacando o sistema de votação. Entre outros fundamentos, a Corte equiparou o caso ao julgado no RO nº 0603975-86 (caso do Deputado Francischini), que já havia reconhecido que a disseminação de fatos inverídicos acerca da lisura do pleito, em benefício do candidato, configura abuso de poder político ou de autoridade e/ou uso indevido

[80] §230: "No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider". Tradução livre: "Nenhum provedor ou usuário de um serviço de computador interativo deve ser tratado como o editor ou orador de qualquer informação fornecida por outro provedor de conteúdo de informação".

[81] POR maioria de votos, TSE declara Bolsonaro inelegível por 8 anos. *Tribunal Superior Eleitoral*, 30 jun. 2023. Disponível em: https://www.tse.jus.br/comunicacao/noticias/2023/Junho/por-maioria-de-votos-tse-declara-bolsonaro-inelegivel-por-8-anos.

dos meios de comunicação quando redes sociais são usadas para esse fim. Esse entendimento decorre da constatação de que a liberdade de expressão não protege a disseminação de desinformação eleitoral, sob pena de a democracia sucumbir ao charlatanismo político.

VI Conclusão

A rede mundial de computadores permitiu o acesso ao conhecimento, à informação e ao espaço público por bilhões de pessoas, mudando o curso da história. No entanto, o uso indevido da internet e das mídias sociais pode trazer sérias ameaças à democracia e aos direitos fundamentais. Algum grau de regulação, portanto, tornou-se necessário para enfrentar os comportamentos inautênticos e os conteúdos ilegítimos. É essencial, no entanto, agir com transparência, proporcionalidade e procedimentos adequados, para que o pluralismo, a diversidade e a liberdade de expressão sejam preservados. A educação midiática e a conscientização das pessoas de boa-fé – que felizmente constituem a maioria – são medidas decisivas para o uso construtivo das novas tecnologias.

Referências

A GLOBAL dialogue to guide regulation worldwide. *Unesco*, 2023. Disponível em: https://www.unesco.org/en/internet-conference. Acesso em: 8 maio 2023.

AUDITING Algorithms: the existing landscape, role of regulator and future outlook. *Digital Regulation Cooperation Forum*, 23 set. 2022. Disponível em: https://www.gov.uk/government/publications/auditing-algorithms-the-existing-landscape-role-of-regulators-and-future-outlook. Acesso em: 7 maio 2023.

BALKIN, Jack M. Free speech in the algorithmic society: big data, private governance, and new school speech regulation. *University of California*, Davis, v. 51, p. 1149-1210, 2018. Disponível em: https://lawreview.law.ucdavis.edu/issues/51/3/Essays/51-3_Balkin.pdf. Acesso em: 7 maio 2023.

BALKIN, Jack M. Free speech is a triangle. *Columbia Law Review*, v. 118, n. 7, p. 2011-2056, 2018. Disponível em: https://columbialawreview.org/wp-content/uploads/2018/11/Balkin-FREE_SPEECH_IS_A_TRIANGLE.pdf. Acesso em: 5 maio 2023.

BALKIN, Jack M. How to regulate (and not regulate) social media. *Journal of Free Speech Law*, v. 71, 2021; *Knight Institute Occasional Paper Series*, n. 1, March 2020; *Yale Law School, Public Law Research Paper Forthcoming*, 20 nov. 2019. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3484114. Acesso em: 7 maio 2023.

BALZ, Dan. A year after Jan. 6, are the guardrails that protect democracy real or illusory? *The Washington Post*, Washington, 6 jan. 2022. Disponível em: https://www.washingtonpost.com/politics/democracy-january-6/2022/01/06/2a1fc41e-6db4-11ec-a5d2-7712163262f0_story.html. Acesso em: 5 maio 2023.

BARROSO, Luís Roberto. O constitucionalismo democrático ou neoconstitucionalismo como ideologia vitoriosa do século XX. *Revista Publicum*, v. 4, 2018.

BARROSO, Luís Roberto. Technological revolution, democratic recession and climate change: the limits of law in a changing world. *International Journal of Constitutional Law*, v. 18, 2020.

BARROSO, Luna van Brussel. *Liberdade de expressão e democracia na era digital*: o impacto das mídias sociais no mundo contemporâneo. Belo Horizonte: Fórum, 2022.

BEYER, R. Can we fix what's wrong with social media? *Yale Law Report*, verão 2022.

BREXIT: Reaction from around the UK. *BBC*, Londres, 24 jun. 2016. Disponível em: https://www.bbc.com/news/uk-politics-eu-referendum-36619444. Acesso em: 5 maio 2023.

BUCKMAN, Ian. Hashing it out: how an automated crackdown on child pornography is shaping the Fourth Amendment. *Berkeley Journal of Criminal Law*, Berkeley, 13 abr. 2021. Disponível em: https://www.bjcl.org/blog/hashing-it-out-how-an-automated-crackdown-on-child-pornography-is-shaping-the-fourth-amendment/. Acesso em: 7 maio 2023.

DIAMOND, Larry. Facing up to the democratic recession. *Journal of Democracy*, v. 26, 2015.

DICHO, Michael; LOGVINENKO, Igor. Authoritarian populism, courts and democratic erosion. *Just Security*, 11 fev. 2021. Disponível em: https://www.justsecurity.org/74624/authoritarian-populism-courts-and-democratic-erosion/. Acesso em: 5 maio 2023.

DOUEK, Evelyn. Content moderation as systems thinking. *Harvard Law Review*, v. 2, p. 136, 2022. Disponível em: https://harvardlawreview.org/2022/12/content-moderation-as-systems-thinking/. Acesso em: 8 maio 2023.

DOUEK, Evelyn. Governing online speech. *Columbia Law Review*, v. 121, n. 3, 2021. Disponível em: https://columbialawreview.org/wp-content/uploads/2021/04/Douek-Governing_Online_Speech-from_Posts_As-Trumps_To_Proportionality_And_Probability.pdf. Acesso em: 7 maio 2023.

DWORKIN, Ronald. *Is democracy possible here?* Princeton: Princeton University Press, 2008.

DWORKIN, Ronald. *Taking rights seriously*. Cambridge: Harvard University Press, 1997.

ECPS – EUROPEAN CENTER FOR POPULISM STUDIES. *Digital Populism*. Disponível em: https://www.populismstudies.org/Vocabulary/digital-populism/. Acesso em: 5 maio 2023.

ELKIN-KOREN, Niva; PEREL, Maayan. Speech contestation by design: democratizing speech governance by AI. *Florida State University Law Review* [forthcoming]. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4129341https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4129341. Acesso em: 5 maio 2023.

ENGESSER, Sven *et al*. Populism and social media: how politicians spread a fragmented ideology. *Information, Communication & Society*, v. 20, 2017.

FACEBOOK statistics and trends. *Datareportal*, 19 fev. 2023. Disponível em: https://datareportal.com/essential-facebook-stats. Acesso em: 8 maio 2023.

FAUSTO, Sergio. O desafio democrático. *Revista Piauí*, v. 8, 2022.

FUSSEL, Sidney. Why the New Zealand shooting video keeps circulating. *The Atlantic*, 21 mar. 2019. Disponível em: https://www.theatlantic.com/technology/archive/2019/03/facebook-youtube-new-zealand-tragedy-video/585418/. Acesso em: 7 maio 2023.

GLOBAL Charter of Ethics for Journalists. *The International Federation of Journalists*, jun. 2019. Disponível em: https://perma.cc/7A2C-JD2S. Acesso em: 5 maio 2023.

GOLDSTEIN, Ariel. Brazil leads the third wave of the Latin American far right. *C-REX – Center for Research on Extremism*, 1º mar. 2021. Disponível em: https://www.sv.uio.no/c-rex/english/news-and-events/right-now/2021/brazil-leads-the-third-wave-of-the-latin-american-.html. Acesso em: 5 maio 2023.

HAIDT, Jonathan. Why the past 10 years of American life have been uniquely stupid. *The Atlantic*. Disponível em: https://www.theatlantic.com/magazine/archive/2022/05/social-media-democracy-trust-babel/629369/. Acesso em: 8 maio 2023.

HUMAN RIGHTS COMMITTEE. *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*. UN Doc A/HRC/32/38. 11 maio 2016. Disponível em: https://undocs.org/en/A/HRC/32/38. Acesso em: 8 maio 2023.

HUQ, Aziz; GINSBURG, Tom. How to lose a constitutional democracy. *UCLA Law Review*, v. 65, 2018.

ISSACHAROFF, Samuel. *Fragile democracies*: contested power in the era of constitutional courts. Cambridge: Cambridge University Press, 2015.

JACKSON, Vicki C. Knowledge institutions in constitutional democracies: reflections on the "press". *The Jorunal of Meida Law*, v. 14, 2022. Disponível em: https://doi.org/10.1080/17577632.2022.2142733. Acesso em: 5 maio 2023.

JONES, Seth G. The rise of far-right extremism in the United States. *Center for Strategic & International Studies*, nov. 2018. Disponível em: https://www.csis.org/analysis/rise-far-right-extremism-united-states. Acesso em: 5 maio 2023.

KADRI, Thomas E.; KLONICK, Kate. Facebook v. Sullivan: public figures and newsworthiness in online speech. *Southern California Law Review*, v. 93, p. 37-99, 2019. Disponível em: https://scholarship.law.stjohns.edu/faculty_publications/292/. Acesso em: 5 maio 2023.

KELLER, Daphne. Some humility about transparency. *The Center for Internet and Society Blog*, 19 mar. 2021. Disponível em: https://cyberlaw.stanford.edu/blog/2021/03/some-humility-about-transparency. Acesso em: 7 maio 2023.

KLONICK, Kate. The new governors: the people, rules, and processes governing online speech. *Harvard Law Review*, v. 131, p. 1598-1670, 2018. Disponível em: https://harvardlawreview.org/2018/04/the-new-governors-the-people-rules-and-processes-governing-online-speech/. Acesso em: 7 maio 2023.

KOSHIYAMA, Adriano; KAZIM, Emre; TRELEAVEN, Philip. Algorithm auditing: managing the legal, ethical, and technological risks of artificial intelligence, machine learning, and associated algorithms. *IEEE Transactions on Technology and Society*, v. 3, p. 128-142, abr. 2022. Disponível em: https://ieeexplore.ieee.org/document/9755237. Acesso em: 8 maio 2023.

KUO, Ming-Sung. Against instantaneous democracy. *International Journal of Constitutional Law*, v. 17, p. 554-575, 2019. Disponível em: https://doi.org/10.1093/icon/moz029. Acesso em: 5 maio 2023.

LANDAU, David. Abusive constitutionalism. *U.C. Davis Law Review*, v. 47, 2013.

LARSON, Erik J. *The myth of artificial intelligence*: why computers can't think the way we do. [s.l.]: Belknap Press, abr. 2021.

LEERSSEN, Paddy. The soap box as a black box: regulating transparency in social media recommender systems. *European Journal of Law and Technology*, v. 11, 2020. Disponível em: https://ssrn.com/abstract=3544009. Acesso em: 7 maio 2023.

LEIDIG, Eviane. Hindutva as a variant of right-wing extremism. *Patterns of Prejudice*, v. 54, n. 3, p. 215-237, 2020.

LESSIG, Lawrence. *They don't represent us*: reclaiming our democracy. Providence: Dey Street Books, 2019.

LEVITSKY, Steven; WAY, Lucan A. The rise of competitive authoritarianism. *Journal of Democracy*, v. 13, 2002.

MACCARTHY, Mark. Transparency requirements for digital social media platforms: recommendations for policy makers and industry. *Transatlantic Working Group*, 24 jun. 2020. Disponível em: https://ssrn.com/abstract=3615726. DOI: http://dx.doi.org/10.2139/ssrn.3615726. Acesso em: 8 maio 2023.

MAGARIAN, Gregory P. A internet e as mídias sociais. *In*: STONE, Adrienne; SCHAUER, Frederick. *Liberdade de expressão*. Oxford: Oxford University Press, 2021.

MESEROLE, Chris. How do recommender systems work on digital platforms? *Brookings*, 21 set. 2022. Disponível em: https://www.brookings.edu/techstream/how-do-recommender-systems-work-on-digital-platforms-social-media-recommendation-algorithms/. Acesso em: 5 maio 2023.

MINOW, Martha. *Saving the press*: why the Constitution calls for government action to preserve freedom of speech. Oxford: Oxford University Press, 2021.

MUDDE, Cas. *The populist zeitgeist*. Government and opposition. Cambridge: Cambridge University Press, 2004. v. 39.

NAHMIAS, Yifat; PEREL, Maayan. The oversight of content moderation by AI: impact assessment and their limitations. *Harvard Journal on Legislation*, v. 58, 2021. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3565025. Acesso em: 8 maio 2023.

ORWA, Robert; BINNS, Reuben; KATZENBACH, Christian. Moderação de conteúdo algorítmico: desafios técnicos e políticos na automação da governança de plataformas. *Big Data & Society*, v. 7, p. 1-15, 2020. Disponível em: https://journals.sagepub.com/doi/full/10.1177/2053951719897945. Acesso em: 7 maio 2023.

RUSSELL, Stuart. *Human compatible*: artificial intelligence and the problem of control. N. York: Penguin Books, 2019.

SCHEPPELE, Kim Lane. Autocratic legalism. *University of Chicago Law Review*, v. 85, 2018.

SCHWAB, Klaus. *A Quarta Revolução Industrial*. Tradução de Cássio Leite Vieira. São Paulo: Edipro, 2018. v. 1.

SHAFFER, Kris. *Data versus democracy*: how big data algorithms shape opinions and alter the course of history. Colorado: Apress, 2019.

SIYECH, Mohammed Sinan. An introduction to right-wing extremism in India. *New Eng. J. Pub. Pol.*, v. 1, 2021.

THAKUR, Dhanaraj; LLANSÓ, Emma. Do you see what I see? Capabilities and limits of automated multimedia content analysis. *Center for Democracy & Technology*, Washington, 20 maio 2021. Disponível em: https://cdt.org/insights/do-you-see-what-i-see-capabilities-and-limits-of-automated-multimedia-content-analysis/. Acesso em: 7 maio 2023.

WHATSAPP 2023 user statistics: how many people use WhatsApp? *Backlinko*, 5 jan. 2023. Disponível em: https://backlinko.com/whatsapp-users. Acesso em: 8 maio 2023.

WOOLDRIDGE, Michael. *A brief history of artificial intelligence*: what it is, where we are, and where we are going. New York: Flatiron Book, jan 2021.

WPFD 2021: attacks on press freedom growing bolder amid rising authoritarianism. *International Press Institute*, 30 abr. 2021. Disponível em: https://ipi.media/wpfd-2021-attacks-on-press-freedom-growing-bolder-amid-rising-authoritarianism/. Acesso em: 5 maio 2023.

WU, Tim. Is the first amendment obsolete? *In*: POZEN, David E. (Ed.). *The perilous public square*. N. York: Columbia University Press, 2020. *E-book Kindle*.

YOUTUBE User Statistic. *Global Media Insight*, 27 fev. 2023. Disponível em: https://www.globalmediainsight.com/blog/youtube-users-statistics/. Acesso em: 8 maio 2023.

ZAKARIA, Fareed. The rise of illiberal democracies. *Foreign Affairs*, v. 76, n. 22, 1997.

ZITTRAIN, Jonathan. Answering impossible questions: content governance in an age of disinformation. *Harvard Kennedy School – Misinformation Review*, 4 jan. 2020. Disponível em: https://misinforeview.hks.harvard.edu/article/content-governance-in-an-age-of-disinformation/. Acesso em: 8 maio 2023.

Informação bibliográfica deste texto, conforme a NBR 6023:2018 da Associação Brasileira de Normas Técnicas (ABNT):

BARROSO, Luís Roberto; BARROSO, Luna van Brussel. Democracia, mídias sociais e liberdade de expressão: ódio, mentiras e a busca da verdade possível. *Direitos Fundamentais & Justiça*, Belo Horizonte, ano 17, n. 49, p. 285-311, jul./dez. 2023.

Submissão: 22.09.2023
Pareceres: 25.09.2023, 11.10.2023
Aceite: 23.10.2023