

Discursos de ódio no contexto da desordem informacional: reflexões sobre estratégias de mitigação¹

Renê Alberto Moritz da Silva e Forster²
 Rodrigo Daniel Paiva Monteiro de Carvalho³

Resumo: A era do informacionalismo trouxe novas dinâmicas para os discursos de ódio. Se faz necessário, portanto, entender como fenômenos do mundo digital impactam esse tipo de manifestação. Este artigo apresenta uma revisão bibliográfica narrativa sobre estratégias de mitigação e combate aos discursos de ódio no contexto da chamada desordem informacional (entendida como a produção e circulação de informações maliciosas ou falas particularmente em ambiente digital). Por meio dessa pesquisa bibliográfica, temos o objetivo de oferecer um material introdutório elencando possíveis soluções para esses problemas. Nesse sentido, detalhamos quatro eixos de medidas de enfrentamento e resistência aos discursos de ódio e à desordem informacional: o educativo (no qual consideramos a importância do letramento), o classificatório (no qual consideramos iniciativas de checagem de fatos e também soluções algorítmicas para a detecção de conteúdos maliciosos ou falsos), o normativo (no qual tratamos da regulação interna e pública das mídias sociais) e o econômico (no qual consideramos a mobilização social pelo desfinanciamento de produtores ou divulgadores de conteúdo malicioso ou falso). Na análise de cada um deles, refletimos sobre suas potencialidades e também sobre suas limitações, defendendo a tese de que a complexidade dos problemas em questão exige uma combinação de diferentes ações.

Palavras-chave: Discursos de ódio; Desordem informacional; Desinformação; Fake News; Letramento.

Introdução

Não é difícil recapitular eventos motivados por discursos de ódio ao longo da história. As novas tecnologias digitais, contudo, estabeleceram um ponto de inflexão na circulação desses discursos. Para citar apenas um exemplo do impacto das novas tecnologias na propagação desses discursos, temos o caso Índia. O país assistiu ao surgimento dos chamados “linchamentos de *WhatsApp*”, episódios de violência coletiva desencadeados por rumores, no qual o mensageiro instantâneo parece ter sido responsável pela catalisação de discursos de ódio (ARUN, 2019; BANAJI *et al.*, 2019).

¹ O presente artigo integra o projeto FAPERJ E-26/010.000145/2016.

² Doutor em Estudos da Linguagem pela Pontifícia Universidade Católica do Rio de Janeiro. E-mail: rene.forster@gmail.com. ORCID iD: <https://orcid.org/0000-0002-3681-1505>.

³ Doutorado em Estudos Contemporâneos pelo Centro de Estudos Interdisciplinares do Século XX (CEIS20), da Universidade de Coimbra. Mestre em História Comparada pela Universidade Federal do Rio de Janeiro. E-mail: rodrigomonteirodecarvalho@gmail.com. ORCID iD: <https://orcid.org/0000-0002-0263-7692>.

O exemplo ilustra como a rapidez e a desintermediação da circulação de informação possibilitada pelas novas tecnologias, nas quais se incluem as redes sociais, permitiram que discursos discriminatórios ganhassem uma nova dimensão pela velocidade de sua propagação e pela difícil reversibilidade de seus efeitos. Além disso, o caso ilustra também como os discursos de ódio, na atualidade, são impactados pelo fenômeno da desordem informacional.

A “desordem informacional” é um termo que vem sendo adotado para englobar diferentes formas disfuncionais de circulação da informação, em particular, nos ambientes digitais (WARDLE; DERAKHSHAN, 2017). O termo abarca essencialmente três fenômenos rotulados, no inglês, de *misinformation* (que designa informação inverídica mas não mal intencionada), *malinformation* (que designa informação falsa, mas distorcida de maneira intencional) e *disinformation* (com a tradução já consolidada no termo “desinformação”, que designa informação falsa e necessariamente maliciosa⁴).

O desenvolvimento tecnológico associado à falta de regulação clara e eficaz dos ambientes digitais e à ausência de políticas educacionais com foco nas novas tecnologias engendraram esse circuito de propagação descontrolada de conteúdos inverídicos ou mal-intencionados, que é considerado hoje uma ameaça às democracias (OSCE, 2018; BAYER *et al.*, 2019; STARBIRD, 2019) e até mesmo à humanidade (WEF, 2018). Entre suas repercussões, estão danos à saúde pública (SOUSA JÚNIOR *et al.*, 2020; ISLAM *et al.*, 2020; PETERS *et al.*, 2018; JOLLEY; DOUGLAS, 2014), a perturbação de processos eleitorais (BENITES, 2018; POSETTI; MATTHEWS, 2018; HÄNSKA; BAUCHOWITZ, 2017) e, como dissemos, a potencialização de discursos de ódio (ARUN, 2019; MUTAHI; KIMARI, 2017; STANLEY, 2017).

Neste artigo, apresentamos uma revisão bibliográfica narrativa sobre possíveis estratégias de resistência e enfrentamento aos discursos de ódio no contexto da desordem informacional. Nesse sentido, buscamos, com base em um levantamento bibliográfico de diferentes campos do saber, inventariar ações que, abordando a propagação disfuncional de informação, poderiam responder ao problema dos discursos de ódio na contemporaneidade. O principal objetivo deste trabalho é oferecer um material introdutório ao tema, numa perspectiva transdisciplinar.

⁴ Mas veja-se Oliveira (2020) e Forster *et al.* (2021) para uma discussão acerca dessa terminologia.

Nessa revisão, partimos da premissa de que não existe uma solução única para os problemas em questão, uma vez que se trata de fenômenos multidimensionais. Desse modo, são necessárias estratégias diversificadas para mitigar os efeitos e reduzir a produção e circulação de conteúdos maliciosos ou falsos, incluindo aí os discursos de ódio.

Consideramos, assim, quatro eixos de estratégias para o combate aos discursos de ódio no âmbito da desordem informacional, tomando emprestadas algumas das categorias elencadas pelo trabalho de Oliveira (2020). Em primeiro lugar, apresentamos, na seção a seguir, o eixo de ações educativas, no qual tratamos das iniciativas voltadas para a promoção do Letramento. Em segundo lugar, são analisadas estratégias classificatórias, entendidas como aquelas que se baseiam na ideia de verificação da verdade. Nesse espectro, abordamos tanto o *fact-checking*, entendido como a apuração jornalística de informações, quanto as soluções tecnológicas de verificação de conteúdos. Em terceiro lugar, apresentam-se estratégias regulatórias das mídias digitais, no âmbito das quais consideramos a autorregulação e a regulação estatal. Por fim, consideramos o eixo de estratégias econômicas, no qual analisamos a repercussão da mobilização pelo desfinanciamento de produtores e divulgadores de informações falsas e/ou maliciosas. Na análise de todas essas abordagens, procuramos refletir tanto sobre potencialidades como limitações de cada uma dessas estratégias.

2 Estratégias Educativas

2.1 Letramento⁵

Medidas de caráter educativo enfocam o problema da desordem informacional no domínio de sua recepção, entendendo que usuários competentes de uma língua dispõem de ferramentas próprias para avaliar criticamente conteúdos maliciosos ou falsos. As iniciativas nessa direção têm buscado ampliar o grau de letramento dos indivíduos, em particular, no que tange a domínios mais específicos do letramento, como os letramentos informacional, midiático e digital (Cf. REGIS; AUDI; MAIA *et al.*, 2014; GOMES; PENNA; ARROIO *et al.*, 2020; Cf.

⁵ Os autores agradecem a Emanuelle Avila pela contribuição na redação e no levantamento bibliográfico realizado para esta seção.

FORSTER *et al.*, 2021). Entre estratégias educativas nesse âmbito, podemos citar como exemplo as que visam a ajudar os leitores a refletir sobre finalidades, pontos de vistas, vieses e orientações políticas de textos a partir do desenvolvimento de habilidades cognitivas como as de análise, síntese, dedução, indução e abstração (POTTER; BYRNE, 1998).

Pesquisas empíricas têm sugerido que iniciativas de letramento podem, de fato, mitigar o problema da desordem informacional. Segundo estudo realizado no Brasil, indivíduos com maior grau de letramento teriam menor propensão a acreditar em informações falsas (ESTABEL; LUCE; SANTINI, 2020). Na mesma direção, uma pesquisa norte-americana observou que a exposição de jovens ao letramento midiático os torna mais hábeis a julgar a acurácia de declarações (KAHNE; BOWYER, 2017). Há evidência também de que o grau de letramento impacta hábitos de consumo de notícias e a atenção direcionada a pistas de credibilidade (NEWMAN *et al.*, 2018).

Diante disso, alguns governos e organizações ao redor do mundo têm testado, na prática, iniciativas de letramento. Numa das ações mais sistemáticas e consistentes nessa seara, a Finlândia tem desenvolvido programas de letramento midiático que vão desde a educação infantil até o treinamento de funcionários públicos (KENNYBIRCH, 2019; MACKINTOSH, 2019). Governos, como os de Taiwan e da Itália, propuseram alterações no currículo com vistas a introduzir práticas de treinamento midiático (SMITH, 2017; HOROWITZ, 2017). Há também a atuação de ONGs, como, por exemplo, a First Draft News e a Media Monitoring Africa. Esta última, sediada na África do Sul, desenvolve diversas estratégias de letramento, como programas direcionados a crianças e adolescentes, além de oferecer recursos de checagem. No Brasil, as ações nesse sentido são ainda pontuais e assistemáticas. Numa delas, o TSE (Tribunal Superior Eleitoral), por ocasião das eleições municipais de 2020, desenvolveu uma série de vídeos, veiculados na TV e em mídias sociais, para capacitar os eleitores a identificar conteúdos falsos.

Parece, assim, haver uma conscientização crescente de que ações de natureza educativa são um caminho para o enfrentamento dos discursos de ódio e da desordem informacional. É preciso considerar, no entanto, que políticas educacionais têm impacto a longo prazo. Intervenções de letramento, em particular, parecem ser mais eficientes à medida que o tempo de instrução aumenta (JEONG *et al.*, 2012). Dessa forma, embora essas iniciativas venham sendo elencadas como uma solução para a propagação dos discursos de ódio e para a desordem

informacional, a urgência desses temas requer que ações educativas sejam parte de um conjunto de estratégias que incidam não apenas sobre os receptores, mas também sobre criadores e mídias, de modo, inclusive, a não deixar apenas sobre a audiência o ônus dos problemas em questão (BULGER; DAVIDSON, 2018).

3 Estratégias Classificadoras

3.1 *Fact-checking*

Como já ressaltado, na contemporaneidade, a proliferação dos discursos de ódio se articula ao circuito de conteúdos e notícias falsos. Um exemplo dessa articulação é o caso do chamado “kit gay”. O boato de que materiais educativos que direcionariam a escolha sexual de estudantes eram distribuídos nas escolas do país serviu de base para toda uma campanha de ódio contra pessoas LGBTQIA+ (ALBUQUERQUE MARANHÃO FILHO, 2018). O caso mostra que o enfrentamento à rede de conteúdos falsos pode ser um recurso importante no combate aos discursos de ódio.

O método mais difundido para o combate à circulação de informações falsas talvez seja a verificação por agências de checagem de fatos (ou *fact-checking*, em inglês). Os resultados dessas análises comumente são apresentados na forma de uma escala de veracidade, que classifica as informações em gradações que vão desde completamente verdadeiras a completamente falsas (WALTER *et al.*, 2020, p. 355). Originalmente, essas iniciativas examinavam a credibilidade do conteúdo de discursos de pessoas públicas, notadamente políticos e candidatos em campanha eleitoral, mas com a proliferação de conteúdos falsos em ambiente digital, também passaram a ser alvos de escrutínio fotos, vídeos, memes e quaisquer outras formas de comunicação. Em um levantamento realizado pela Duke Reporter’s Lab, divulgado em outubro de 2020, constavam 304 projetos ativos em 84 países, um crescimento de 45% em comparação ao ano anterior (STENCEL; LUTHER, 2020). No Brasil, a atividade de checagem de fatos independente chegou a obter chancela governamental. Em 2020, às vésperas das eleições municipais brasileiras, nove agências foram selecionadas para participar

do projeto “Coalizão para Checagem - Eleições 2020”, promovido pelo Tribunal Superior Eleitoral (TSE) e passaram a ter seu conteúdo divulgado pelo site oficial do órgão.

Existem, contudo, divergências quanto a efetividade dessa abordagem. Estudos apontam que a exposição a checagem de fatos até pode ter um efeito positivo em corrigir informações equivocadas, mas não o suficiente para alterar predisposições e comportamentos de segmentos do público (ECKER *et al.*, 2011; SWIRE-THOMPSON *et al.*, 2017; THORSON, 2016; WALTER *et al.*, 2020). Desmentidos também poderiam suscitar reinterpretação de informações de modo a reajustá-las às percepções pré-concebidas (GAINES *et al.*, 2007; SCHAFFNER; ROCHE, 2016). Outros estudos indicam ainda que as correções podem surtir efeito contrário do esperado ao ajudar a sedimentar a crença em informações falsas em determinadas audiências (HOLMAN; LAY, 2019; LEWANDOWSKY *et al.*, 2012; NYHAN; REIFLER, 2010), ou que a mera repetição da desinformação, ainda que para corrigi-la, pode resultar em maior disseminação e aceitação do conteúdo falso (BERINSKY, 2017; SCHWARZ; NEWMAN; LEACH, 2016; SCHWARZ *et al.*, 2007).

Em sentido inverso, há estudos que não encontraram evidências de que a verificação de fatos possa gerar efeitos contrários (ECKER; LEWANDOWSKY; CHADWICK, 2020; HAGLIN, 2017; WOOD; PORTER, 2019). Segundo outros estudos, as correções poderiam promover efeitos positivos longevos (CARNAHAN; BERGAN; LEE, 2020); reduziriam o compartilhamento de desinformação nas redes sociais (MENA, 2020; NEKMAT, 2020; OZTURK; LI; SAKAMOTO, 2015) e diminuiriam significativamente a crença em informações falsas, mesmo em públicos politicamente polarizados (GOTTFRIED *et al.*, 2013; HAMELEERS; VAN DER MEER, 2020; NYHAN; REIFLER, 2017; PORTER; WOOD; KIRBY, 2018; WEEKS, 2015).

Além do debate sobre a eficácia da checagem de fatos, há ainda a discussão sobre sua legitimidade. Quem teria o mandado para definir o que é ou não verdadeiro? Na ausência de um padrão-ouro para se averiguar a verdade, as agências de checagem teriam se outorgado esse papel, mesmo que existam dúvidas sobre a viabilidade desse esforço na aferição de determinadas asserções políticas (KUKLINSKI *et al.*, 1998) e sobre o possível enviesamento político das análises (ALBUQUERQUE, 2021). Há também questionamentos sobre a ausência de métodos objetivos para a escolha dos fatos a serem checados, o que geraria o risco de se produzir um desbalanço da percepção do público em relação a diferentes alvos das verificações,

sobretudo em campanhas eleitorais (USCINSKI; BUTLER, 2013), além da discrepância encontrada entre vereditos dados por diferentes iniciativas acerca do mesmo fato verificado (LIM, 2018; MARIETTA; BARKER; BOWSER, 2015).

3.2 Soluções tecnológicas

Uma segunda classe de iniciativas baseadas na ideia de verificação da verdade são as ferramentas automáticas de verificação. Essas soluções tecnológicas para a desordem informacional se baseiam, em geral, no desenvolvimento de mecanismos de checagem algorítmicos (de conteúdos ou de perfis de usuários), comumente baseados em tecnologias de aprendizado de máquina (Cf. BURKHARDT, 2017). Recursos dessa natureza podem ser destinados à implementação como mecanismos de filtragem pelas plataformas. O Youtube, por exemplo, sob acusações de que seu sistema de recomendação de vídeos poderia estar induzindo a proliferação de conteúdos maliciosos e falsos, supostamente implementou estratégias que reduziram a recomendação de conteúdos conspiratórios (FADDLOUL; CHASLOT; FARID, 2020). Mecanismos de checagem automática também são oferecidos a usuários como recurso adicional pelas próprias plataformas ou por terceiros. Entre as iniciativas nessa direção, estão as que permitem usuários do Twitter verifiquem se determinados perfis ou conteúdos seriam ilegítimos⁶.

Um dos métodos para realizar checagens dessa natureza se dá pela análise de dados técnicos de sites noticiosos. A organização Global Disinformation Index (GDI)⁷, por exemplo, com base na tecnologia de aprendizado de máquina, desenvolveu um algoritmo que, analisando metadados de sites de notícias, poderia classificá-los como propensos ao compartilhamento de desinformação (MELFORD; FAGAN, 2019). Outro método possível se baseia na análise automática de padrões linguísticos típicos desses sites (Cf. FORSTER *et al.*, 2021). O Núcleo Interinstitucional de Linguística Computacional (NILC) da USP, por exemplo, oferece uma plataforma online na qual se podem verificar se notícias são falsas ou não com base em uma classificação linguística⁸.

⁶ Por exemplo: <https://pegabot.com.br/>; <https://hoaxy.osome.iu.edu/>; <https://botometer.osome.iu.edu/>.

⁷ <https://disinformationindex.org/>.

⁸ <http://nilc-fakenews.herokuapp.com/>.

A vantagem dos métodos de checagem automática é a de possibilitar combater o problema da desinformação em larga escala. Contudo, como já apontado por Márcio Ribeiro e Pablo Ortellado, verificar a verdade de uma notícia compreende a análise de uma relação entre o mundo e a linguagem. Uma vez que a máquina não tem acesso ao mundo, resta então às máquinas a análise de padrões indiretos. Em relação à detecção de robôs, por exemplo, já se verificou a possibilidade de que a checagem automática possa levar a falsos positivos (RAUCHFLEISCH; KAISER; 2020). A censura a conteúdo ou usuários legítimos é, então, uma desvantagem dessas estratégias. Há ainda limitações quanto a base de aprendizado desses algoritmos, o que poderia enviesar seus resultados. A iniciativa do NILC, por exemplo, não inclui em sua base notícias em parte verdadeiras (Cf. MONTEIRO, 2018), o que pode implicar limitações a sua aplicabilidade, uma vez que conteúdos de desinformação nem sempre são totalmente falsos (Cf. seção 2.1.3).

4 Estratégias Regulatórias

4.1 Autorregulação

Enquanto estratégias educativas e classificatórias abordam a problemática dos discursos de ódio e da desordem informacional focando o domínio da recepção, medidas regulatórias visam a enfrentar esses problemas no domínio de sua produção e circulação. Nesse sentido, temos medidas de regulação externa, que desincentivam ou punem a produção e a difusão desse tipo de conteúdo na esfera jurídica (Cf. seção 4.2), mas também temos medidas de autorregulação, entendidas como aquelas por meio das quais as plataformas utilizadas para essa difusão buscam coibir esse tipo de conteúdo. Refletimos aqui sobre as medidas de autorregulação das mídias sociais, uma vez que são o *locus* da desordem informacional na contemporaneidade (Cf. FORSTER *et al.*, 2021).

Em uma das frentes de atuação nesse sentido, as plataformas de redes sociais implementaram medidas para limitar o compartilhamento de conteúdos (sejam eles verdadeiros ou falsos). O *WhatsApp*, por exemplo, passou a limitar o número de encaminhamentos possíveis de uma mensagem. Numa outra frente, as plataformas passaram a oferecer ferramentas de alerta

contra conteúdos potencialmente maliciosos ou falsos. O *Instagram*, por exemplo, diante da crise sanitária da Covid-19, passou a sinalizar postagens sobre a doença, inserindo nestas publicações banners com links para sites oficiais. A mesma plataforma implementou também alertas sinalizando para os usuários que se pode estar prestes a publicar algo ofensivo (INSTAGRAM, 2021). Além disso, recursos de checagem de informações suspeitas hoje vêm sendo integrados a grandes plataformas, como Google, *Facebook* e *WhatsApp*. Outras iniciativas, como a cobrança pela leitura de conteúdos estão no radar dessas companhias (MOSSERI, 2017).

Essas ações, contudo, parecem ter um impacto limitado no universo da desordem informacional. Por exemplo: embora a limitação do compartilhamento de mensagens no *WhatsApp* possa diminuir a velocidade de propagação de uma informação, essa barreira parece não afetar efetivamente o alcance final deste conteúdo (MELO *et al.*, 2019). Uma das razões para a ineficácia desta medida pode estar no fato de que a plataforma ainda disponibilize uma série de outros recursos permitindo o disparo em massa de mensagens (como a ferramenta de “lista de transmissão”). É preciso também considerar que limitar um dos mecanismos de compartilhamento de mensagens pode até dificultar a repercussão de um boato espontâneo, mas talvez tenha pouco impacto sobre conteúdos falsos produzidos de forma profissional (Cf. FORSTER *et al.*, 2021).

O fato é que num contexto de concentração extrema de mercado, a regulação privada das companhias oferece pouca ou nenhuma vantagem competitiva às plataformas de mídias sociais (Cf. OBSERVACOM, 2019). Diante dos vários mecanismos estruturais das mídias sociais que favorecem a desordem informacional (Cf. FORSTER *et al.*, 2021), parece pouco provável que medidas paliativas, como as adotadas até agora, tenham um impacto significativo contra a circulação de conteúdos maliciosos ou falsos.

4.2 Regulação externa

O combate aos discursos de ódio se faz, em muitos países, por meio de um aparato normativo específico, como é o caso do Brasil. Nesta seção, contudo, conforme a proposta deste artigo, procuramos entender como esse enfrentamento é feito no contexto específico do

universo digital e, mais particularmente, em sua articulação com o fenômeno da desordem informacional.

Nesse sentido, podemos observar a adoção de algumas categorias típicas de estratégias para o combate à desordem informacional. Encontramos normas que obrigam ou estimulam as mídias sociais (i) a facilitarem a denúncia de conteúdo, (ii) a removerem publicações e (iii) a prestarem contas das medidas tomadas contra informação falsa ou maliciosa (ROUDIK *et al.*, 2019). Iniciativas como essas, de regulação pública das mídias sociais, surgem em razão da percepção de que a autorregulação (Cf. seção 4.1) dessas plataformas não tem sido eficiente. Essas medidas têm a função limitar os potenciais ganhos que essas empresas possam vir a ter com condutas e mecanismos que favorecem a desordem informacional (Cf. FORSTER, 2021). Reconhecem, além disso, a responsabilidade das corporações de mídia na difusão de conteúdos maliciosos em circunstâncias em que nem sempre os produtores originais desses materiais podem ser rastreáveis ou imputáveis.

Um exemplo nessa direção é o Código de Condutas para Desinformação da União Europeia, ao qual as plataformas foram convidadas a aderir voluntariamente (EUROPEAN COMMISSION, 2018). Leis nesse sentido também vêm sendo adotadas por alguns países. A França, por exemplo, desenvolveu uma legislação pioneira, visando, em particular, à proteção do processo eleitoral. Os juízes, durante as eleições, podem solicitar a remoção imediata de conteúdo falso (FERREIRA, 2019; LE MONDE, 2018). Na Alemanha, a legislação contempla ainda a possibilidade de sanções a plataformas de mídias sociais com multas pesadas por desobediência a determinações judiciais (GESLEY, 2019).

No Brasil, ainda não existe uma legislação específica que aborde o papel das plataformas de mídia social na desordem informacional. Na verdade, o Marco Civil da Internet até mesmo desfavorece qualquer responsabilização dessas empresas por publicações de terceiros (FERREIRA, 2019; BRANT *et al.*, 2021). O combate ao problema tem se dado, então, por recurso a normas inespecíficas do Código Penal (como as que estabelecem punições para os crimes de calúnia, difamação, injúria e denunciação caluniosa) e, na esfera política, do Código Eleitoral (como as que regem a propaganda eleitoral ou afirmam a liberdade de voto) (PONTES; VALENTE; CAZARRÉ, 2018; ROUDIK *et al.*, 2019). Exemplo da aplicação do Código Penal, em particular, é o chamado “Inquérito das Fake News” no qual os referidos artigos são usados na acusação (BRASIL, 2020). Alguns projetos em tramitação na Câmara e no Senado visam à

regulação de plataformas de mídias sociais e incluem medidas como a incorporação obrigatória de mecanismos de checagem, a restrição ao compartilhamento de mensagens e a notificação aos produtores de conteúdo falso (Cf. OLIVEIRA; CRUZ; SILVA, 2018; ROUDIK *et al.*, 2019; Para uma discussão veja-se: ITS, 2020).

Apesar de já adotadas em outros países e aparentemente apoiadas pela população brasileira (AVAAZ, 2020a), medidas de regulação das plataformas ainda não são um consenso. Defensores da liberdade de expressão argumentam que normas de regulação podem ser concebidas e aplicadas de forma a limitar a diversidade de opiniões (Cf. DIGITAL FUTURE SOCIETY, 2020). A censura poderia ser motivada inclusive por leis que preveem sanções para a não retirada de conteúdos suspeitos, mas que não preveem punição para a retirada de conteúdos legítimos. No Canadá, por exemplo, leis de combate à desinformação já foram julgadas como uma violação inconstitucional à liberdade de expressão (ROUDIK *et al.*, 2019). Nos Estados Unidos, o “direito constitucional de mentir” já foi evocado na defesa produtores de conteúdos falsos.

Um instrumento indireto de regulação dessas plataformas poderia estar em normas de proteção de dados pessoais. Leis nesse sentido, na medida em que limitam a extensão em que as redes sociais podem compartilhar ou usar os dados de seus usuários, teriam potencial tanto para dificultar a utilização dessas informações para o marketing digital ilegítimo quanto para impedir que as próprias plataformas utilizem esses dados para estratégias de promoção de conteúdos que gerem as chamadas “bolhas informacionais”. No Brasil, ainda é incerto o quanto a aplicação da Lei Geral de Proteção de Dados (LGPD) poderá ser efetiva nesse sentido. Por exemplo, parece ser necessária a construção de uma certa arquitetura institucional que leve a Justiça Eleitoral a considerar elementos da LGPD para coibir a interferência da desinformação nos processos eleitorais (CRUZ *et al.*, 2019).

Assim, a medida justa da regulação das plataformas parece estar em criar normas que ao mesmo tempo desincentivem as empresas a se beneficiarem com conteúdo falso ou malicioso, mas que preservem a liberdade de expressão. Isso se garante, inclusive, limitando a concentração neste mercado e obrigando essas empresas a serem mais transparentes nos seus critérios de curadoria de conteúdo. Há ainda o desafio de desenvolver um conjunto de normas que se mantenham relevantes mesmo diante da evolução da tecnologia das redes e que consigam regular nacionalmente sistemas concebidos em escala global.

5 Estratégias Econômicas

5.1 Mobilização Social pelo Desfinanciamento

Movimentos da sociedade civil têm alcançado algum êxito combatendo o financiamento de produtores de conteúdos falsos e maliciosos. Esses produtores conseguem financiamento de grandes empresas, mesmo infringindo limites éticos, a partir de tecnologias que permitem a camuflagem de vínculos entre produtores de conteúdo e seus financiadores. (Cf. MELFORD; FAGAN, 2019; AVAAZ, 2020b; BRADSHAW; BAILEY; HOWARD, 2021; Cf. FORSTER *et al.*, 2021).

Um exemplo paradigmático nessa direção é o *Sleeping Giants Brasil*. Surgido nos Estados Unidos e replicado no país, o movimento tem atuado para expor empresas que, por mecanismos diretos ou indiretos, financiam conteúdo falso. Numa das iniciativas nessa direção, por exemplo, o *Sleeping Giants* expôs patrocinadores de um programa noticioso que divulgava informações falsas e conteúdo homofóbico (VIDON, 2021). A iniciativa, embora tenha alcance pontual, tem o mérito de denunciar o sistema de financiamento da desordem informacional (Cf. LEMOS, 2020; VEIGA, 2020; MAIA; GULLINO, 2020).

Considerações Finais

Conforme buscamos demonstrar com o levantamento bibliográfico apresentado, o enfrentamento aos discursos de ódio no contexto da desordem informacional requer a mobilização de diferentes estratégias, uma vez que nenhuma delas parece ser suficiente para abarcar o problema como um todo. Parece ser necessário também a atuação de diferentes setores da sociedade. Algumas das iniciativas requerem estrutura de estado. É o caso das iniciativas de ordem jurídica. É, em parte, também o caso das de iniciativas de ordem educacional. Outras propostas estão situadas predominantemente na esfera de atuação de organizações da sociedade

civil, como o *fact-checking*, ou a autorregulação das empresas. É necessário, além disso, considerar que a evolução desses fenômenos acompanha a evolução das tecnologias da informação, o que significa dizer que se trata de um cenário dinâmico. A inteligência artificial e o aprendizado de máquina, por exemplo, que são considerados por alguns como aliados na detecção de conteúdos falsos e maliciosos, são também, ao mesmo tempo, uma potencial ameaça. Se enquadram aí as ferramentas de geração automática de textos, que alimentadas por um banco de dados, podem recorrer à internet para produzir textos coerentes sobre qualquer assunto, alegadamente similares aos produzidos por humanos (Cf. ZELLERS *et al.*, 2019; FLORIDI; CHIRIATTI, 2020). Recursos como esses podem ter um grande impacto na indústria de desinformação. Outro exemplo é a possibilidade de que algoritmos recriem ou criem vieses discriminatórios. Num caso recente, o Twitter admitiu que seu algoritmo de apresentação de imagens possuía um viés racial que desprivilegiava rostos negros (HERN, 2020). Exemplos como esses sugerem que, na contemporaneidade, o enfrentamento dos discursos de ódio exige um olhar atento às transformações das tecnologias da informação.

Referências

ALBUQUERQUE MARANHÃO FILHO, E. M. de; COELHO, F. M. F.; DIAS, T. B. Fake news acima de tudo, fake news acima de todos: Bolsonaro e o “kit gay”, “ideologia de gênero” e fim da “família tradicional. *Correlatio*, São Paulo, v. 17, n. 2, p. 65-90, 2018.

ALBUQUERQUE, A. Quem vigia os vigias?: o combate às fake news pós-democracia. In: COSTA, M. *Qual o caminho do Brasil?: instituições, cultura e na política no século XXI*. Rio de Janeiro: Appris, 2021. n. p. Disponível em: https://www.researchgate.net/publication/354332306_Quem_vigia_os_vigias_O_combate_as_fake_news_na_pos-democracia_brasileira. Acesso em: 31 maio 2022.

AN UPDATE on our work to tackle abuse on Instagram. *Instagram*, [s. l.], 2021. Disponível em: <https://about.instagram.com/blog/announcements/an-update-on-our-work-to-tackle-abuse-on-instagram>. Acesso em: 18 jan. 2022.

ARUN, C. On WhatsApp, Rumours, and Lynchings. *Economic & Political Weekly*, Mumbai, v. 54, n. 6, p. 30-35, 2019.

AVAAZ. IBOPE: Ampla maioria dos brasileiros apoia medidas excluídas da lei das fake news, mostra nova pesquisa. AVAAZ, [s. l.], 2020a. Disponível em: <https://secure.avaaz.org/page/po/media/pressreleases/998.html>. Acesso em: 31 maio 2022.

AVAAZ. Why is Youtube Broadcastin Clime Misinformation to Milions? AVAAZ, [s. l.], 2020b. Disponível em: https://secure.avaaz.org/campaign/po/youtube_climate_misinformation/. Acesso em: fev. 2021.

BANAJI, S.*et al.* WhatsApp vigilantes: an exploration of citizen reception and circulation of WhatsApp misinformation linked to mob violence in India. *LSE*, [s. l.], 2019. Disponível em: <https://blogs.lse.ac.uk/mediase/2019/11/11/whatsapp-vigilantes-an-exploration-of-citizen-reception-and-circulation-of-whatsapp-misinformation-linked-to-mob-violence-in-india/>. Acesso em 31 maio 2022.

BAYER, J.*et al.* Disinformation and propaganda—impact on the functioning of the rule of law in the EU and its Member States. *HEC Paris Research Paper*, Rochester, p. 1-202, 2019.

BENITES, A. A máquina de ‘fake news’ nos grupos a favor de Bolsonaro no WhatsApp. *El País*, Brasília, DF, 2018. Disponível em: https://brasil.elpais.com/brasil/2018/09/26/politica/1537997311_859341.html. Acesso em: 5 jul. 2020.

BERINSKY, A. J. Rumors and health care reform: Experiments in political misinformation. *British Journal of Political Science*, Cambridge, v. 47, n. 2, p. 241-262, 2017.

BRADSHAW, S.; BAILEY, H.; HOWARD, P. *Industrialized Disinformation: 2020 Global Inventory of Organized Social Media Manipulation*. Oxford: Oxford Internet Institute, 2021.

BRANT, J.*et al.* *Regulação de combate à desinformação*. [S. l.]: Friedrich Eber Stigung, 2021. Disponível em: <https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKewjV5bamxYr4AhXvA7kGHhHpAh4QFnoECAQQAQ&url=http%3A%2F%2Flibrary.fes.de%2Fpdf-files%2Fbueros%2Fbrasilien%2F17529.pdf&usg=AOvVaw2MIC3ytjh9g2SM0Nt2GFqb>. Acesso em: 31 maio 2022.

BRASIL. SUPREMO TRIBUNAL FEDERAL. *Inquérito 4.781 Distrito Federal*. Trata-se de inquérito instaurado pela Portaria GP Nº 69, de 14 de março de 2019, do Excelentíssimo Senhor Ministro Presidente, nos termos do art. 43 do Regimento Interno desta CORTE. Relator: Min. Alexandre de Moraes, 26 de maio de 2020. Disponível em: <http://www.stf.jus.br/arquivo/cms/noticiaNoticiaStf/anexo/mandado27maio.pdf>. Acesso em: 10 jul. 2020.

BURKHARDT, J. M. Combating fake news in the digital age. *Library Technology Reports*, Chicago, v. 53, n. 8, 2017.

CARNAHAN, D.; BERGAN, D. E.; LEE, S. Do corrective effects last? Results from a longitudinal experiment on beliefs toward immigration in the US. *Political Behavior*, [s. l.], v. 43, n. 3, p. 1227-1246, 2020.

CRUZ, F.*et al.* Internet e Eleições no Brasil: Diagnósticos e Recomendações. *Internetlab*, [s. l.], 2019. Disponível em: <https://internetlab.org.br/pt/noticias/internet-e-eleicoes-no-brasil-diagnostics-e-recomendacoes/>. Acesso em: 31 maio 2022.

DIGITAL FUTURE SOCIETY. Dealing with disinformation. *Digital Future Society*, [s. l.], 2020. Disponível em: <https://digitalfuturesociety.com/report/dealing-with-disinformation/>. Acesso em: fev. 2021.

ECKER, U. K. H.; LEWANDOWSKY, S.; CHADWICK, M. Can corrections spread misinformation to new audiences? Testing for the elusive familiarity backfire effect. *Cognitive Research: Principles and Implications*, [s. l.], v. 5, n. 1, p. 1-25, 2020.

ECKER, U. K. H.; LEWANDOWSKY, S.; CHADWICK, M. *et al.* Correcting false information in memory: Manipulating the strength of misinformation encoding and its retraction. *Psychonomic Bulletin & Review*, [s. l.], v. 18, n. 3, p. 570-578, 2011.

ESTABEL, L. B.; LUCE, B. F.; SANTINI, L. A. Idosos, fake news e letramento informacional. *Revista Brasileira de Biblioteconomia e Documentação*, São Paulo, v. 16, p. 1-15, 2020.

EUROPEAN COMMISSION. Code of Practice on Disinformation. Digital Strategy, [s. l.], 2018. Disponível em: <https://digital-strategy.ec.europa.eu/en/policies/code-practice-disinformation>. Acesso em: 31 maio 2022.

FADDOUL, M.; CHASLOT, G.; FARID, H. A Longitudinal Analysis of YouTube's Promotion of Conspiracy Videos. *Cornell University*, New York, 2020. Disponível em: <https://arxiv.org/abs/2003.03318>. Acesso em: 31 maio 2022.

FERREIRA, R. R. Desinformação em processos eleitorais: um estudo de caso da eleição brasileira de 2018. 2019. Dissertação (Mestrado em Jornalismo e Comunicação) – Universidade de Coimbra, Coimbra, 2019.

FLORIDI, L.; CHIRIATTI, M. GPT-3: Its nature, scope, limits, and consequences. *Minds and Machines*, [s. l.], v. 30, n. 4, p. 681-694, 2020.

FORSTER, R. *et al.* Fake News: What Is It, How Is It Made and Why It Works?. *SciELO Preprints*, 2021. Disponível em: <https://doi.org/10.1590/SciELOPreprints.3294>. Acesso em: 26 maio 2022.

GAINES, B. J. *et al.* Same facts, different interpretations: Partisan motivation and opinion on Iraq. *The Journal of Politics*, Chicago, v. 69, n. 4, p. 957-974, 2007.

GESLEY. Initiatives to Counter Fake News: Germany. *Library of Congress*, Washington, DC, 2019. Disponível em: <https://www.loc.gov/law/help/fake-news/germany.php>. Acesso em: fev. 2021.

GOMES, S. F.; PENNA, J. C. B. de O.; ARROIO, A. Fake News Científicas: Percepção, Persuasão e Letramento. *Ciência & Educação (Bauru)*, Bauru, v. 26, e20018, 2020.

GOTTFRIED, J. A. *et al.* Did fact checking matter in the 2012 presidential campaign?. *American Behavioral Scientist*, [s. l.], v. 57, n. 11, p. 1558-1567, 2013.

HAGLIN, K. The limitations of the backfire effect. *Research & Politics*, [s. l.], v. 4, n. 3, p. 1-5, 2017.

HAMELEERS, M.; VAN DER MEER, T. G. L. A. Misinformation and polarization in a high-choice media environment: How effective are political fact-checkers? *Communication Research*, [s. l.], v. 47, n. 2, p. 227-250, 2020.

HÄNSKA, M.; BAUCHOWITZ, S. Tweeting for Brexit: how social media influenced the referendum. In: MAIR, J. *Brexit, Trump and the Media*. UK: Abramis Academic Publishing; Bury St Edmunds, 2017. p. 31-35.

HERN, A. Twitter apologises for 'racist' image-cropping algorithm. *The Guardian*, [s. l.], 2020. Disponível em: <https://www.theguardian.com/technology/2020/sep/21/twitter-apologises-for-racist-image-cropping-algorithm>. Acesso em: 19 jan. 2022.

HOLMAN, M. R.; LAY, J. C. They see dead people (voting): Correcting misperceptions about voter fraud in the 2016 US presidential election. *Journal of Political Marketing*, [s. l.], v. 18, n. 1-2, p. 31-68, 2019.

HOROWITZ, J. In Italian Schools, Reading, Writing and Recognizing Fake News. *The New York Times*, New York, 2017. Disponível em: <https://www.nytimes.com/2017/10/18/world/europe/italy-fake-news.html>. Acesso em: mar. 2021.

ISLAM, M. S. *et al.* COVID-19-related infodemic and its impact on public health: A global social media analysis. *The American Journal of Tropical Medicine and Hygiene*, Arlington, v. 103, n. 4, p. 1621, 2020.

JEONG, S.-H.; CHO, H.; HWANG, Y. Media literacy interventions: A meta-analytic review. *Journal of Communication*, Oxford, v. 62, n. 3, p. 454-472, 2012.

JOLLEY, D.; DOUGLAS, K. M. The effects of anti-vaccine conspiracy theories on vaccination intentions. *PloS one*, San Francisco, v. 9, n. 2, e89177, 2014.

KAHNE, J.; BOWYER, B. Educating for democracy in a partisan age: Confronting the challenges of motivated reasoning and misinformation. *American Educational Research Journal*, [s. l.], v. 54, n. 1, p. 3-34, 2017.

KENNYBIRCH, R. How Finland shuts down fake news. *Apolitical*, London, 2019. Disponível em: https://apolitical.co/en/solution_article/how-finland-shuts-down-fake-news. Acesso em: mar. 2021.

KUKLINSKI, J. H. *et al.* "Just the facts, ma'am": Political facts and public opinion. *The Annals of the American Academy of Political and Social Science*, [s. l.], v. 560, n. 1, p. 143-154, 1998.

LE MONDE. Le Parlement adopte les propositions de lois sur les infos. *Le Monde*, [s. l.], 2018. Disponível em: https://www.lemonde.fr/actualite-medias/article/2018/11/20/le-parlement-adopte-les-propositions-de-loi-sur-les-infos_5386208_3236.html. Acesso em: fev. 2021.

LE MOS, R. Contra fake news, siga o dinheiro. *ITS*, Rio de Janeiro, 2020. Disponível em: <https://itsrio.org/pt/artigos/contra-fake-news-siga-o-dinheiro/>. Acesso em: fev. 2021.

LEWANDOWSKY, S. *et al.* Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest*, [s. l.], v. 13, n. 3, p. 106-131, 2012.

LIM, C. Checking how fact-checkers check. *Research & Politics*, [s. l.], v. 5, n. 3, p. 1-7, 2018.

MACKINTOSH, E. Finland is winning the war on fake news. What it's learned may be crucial to Western democracy. *CNN*, Atlanta, 2019. Disponível em: https://edition.cnn.com/interactive/2019/05/europe/finland-fake-news-intl/?utm_medium=social&utm_source=fbCNNi&utm_content=2019-05-18T09%3A04%3A45&fbclid=IwAR0g0DqxY9iIDOTnkZzqKKZKcDcXAOo-MAq1Q7c3ERmmxgGoFLLbgJH6sRA. Acesso em: mar. 2021.

MAIA, G.; GULLINO, D. BB tira anúncios de site acusado de compartilhar fake news, Carlos Bolsonaro reclama, e Wajngarten diz estar 'contornando'. *EXTRA*, Rio de Janeiro, 2020. Disponível em: <https://extra.globo.com/noticias/brasil/bb-tira-anuncios-de-site-acusado-de-compartilhar-fake-news-carlos-bolsonaro-reclama-wajngarten-diz-estar-contornando-24437282.html?fbclid=IwAR0D06-L4jIgSICvsRh7jukysiouRW1wz8NngkZoQE4Gg5gW9dQ8rS2N3aIO>. Acesso em: fev. 2021.

MELFORD, C.; FAGAN, C. Curring the funding of disinformation. The Ad-tehc solution. *Global Disinformation Index*, [s. l.], 2019. Disponível em: https://disinformationindex.org/wp-content/uploads/2019/05/GDI_Report_Screen_AW2.pdf. Acesso em: fev. 2021.

MELO, P. *et al.* Can WhatsApp Counter Misinformation by Limiting Message Forwarding?. *Cornell University*, [s. l.], 2019. Disponível em: <https://arxiv.org/abs/1909.08740>. Acesso em: 31 maio 2022.

MENA, P. Cleaning up social media: The effect of warning labels on likelihood of sharing false news on Facebook. *Policy&Internet*, [s. l.], v. 12, n. 2, p. 165-183, 2020.

MONTEIRO, R. *Detecção Automática de Notícias Falsas para o Português*. 2018. Relatório Final de Projeto –Universidade de São Paulo, São Paulo, 2018.

MOSSERI, A. Testing Subscriptions Support in Instant Articles. *Facebook*, 2017. Disponível em: <https://www.facebook.com/formedia/blog/testing-subscriptions-support-in-instant-articles>. Acesso em: mar. 2021.

MUTAH, P.; KIMARI, B. The impact of social media and digital technology on electoral violence in Kenya. *IDS*, [s. l.], 2017. Disponível em: <https://opendocs.ids.ac.uk/opendocs/handle/20.500.12413/13159>. Acesso em: 31 maio 2022.

NEKMAT, E. Nudge effect of fact-check alerts: source influence and media skepticism on sharing of news misinformation in social media. *Social Media+ Society*, [s. l.], v. 6, n. 1, p. 1-36, 2020.

NEWMAN, N. *et al.* Reuters Institute Digital News Report 2018. *Reuters Institute for the Study of Journalism*, Oxford, 2018. Disponível em: <https://reutersinstitute.politics.ox.ac.uk/sites/default/files/digital-news-report-2018.pdf>. Acesso em: 31 maio 2022.

NYHAN, B.; REIFLER, J. When corrections fail: The persistence of political misperceptions. *Political Behavior*, [s. l.], v. 32, n. 2, p. 303-330, 2010.

OBSERVACOM. Contribuições para uma regulação democrática das grandes plataformas que garanta a liberdade de expressão na internet. *Observacom*, [s. l.], 2019. Disponível em: <https://www.observacom.org/contribuicoes-para-uma-regulacao-democratica-das-grandes-plataformas-que-garanta-a-liberdade-de-expressao-na-internet/>. Acesso em: maio 2022.

OLIVEIRA, R.; CRUZ, R.; SILVA, F. Caça Às Bruxas Às Fake News: Os Possíveis Desdobramentos da Criminalização das “Notícias Falsas”. In: POLIDO, F.; ANJOS, L.; BRANDÃO, L. *III Seminário Governança das Redes, 2018*, Belo Horizonte. III Seminário Governança das Redes: políticas, internet e sociedade. Belo Horizonte: Instituto de Referência em Internet e Sociedade, 2018. Disponível em: <http://bit.ly/38A7YbJ>. Acesso em: fev 2021.

OLIVEIRA, T. M. de *et al.* Como enfrentar a desinformação científica? Desafios sociais, políticos e jurídicos intensificados no contexto da pandemia. *Liinc em Revista*, Rio de Janeiro, v. 16, n. 2, p. e5374-e5374, 2020.

ORGANIZATION FOR SECURITY AND CO-OPERATION IN EUROPE. Joint Declaration on Freedom of Expression and “Fake News”, Disinformation, and Propaganda. OSCE, [s. l.], 2018. Disponível em: <https://www.osce.org/fom/302796>. Acesso em: 23 set. 2020.

OZTURK, P.; LI, H.; SAKAMOTO, Y. Combating rumor spread on social media: The effectiveness of refutation and warning. In: HAWAII INTERNATIONAL CONFERENCE ON SYSTEM SCIENCES. 48., 2015. *Proceedings* [...]. [S. l.]: IEEE, 2015. p. 2406-2414.

PETERS, A. *et al.* Fighting the good fight: the fallout of fake news in infection prevention and why context matters. *Journal of Hospital Infection*, London, v. 100, n. 4, p. 365-370, 2018.

PONTES, F.; VALENTE, J.; CAZARRÉ, M. Tudo sobre fake news: projetos de lei, políticas das redes sociais e mais. *Justificando*, [s. l.], 2018. Disponível em: <http://www.justificando.com/2018/07/17/tudo-sobre-fake-news-projetos-de-lei-politicas-das-redes-sociais-e-mais/>. Acesso em: fev. 2021.

PORTER, E.; WOOD, T. J.; KIRBY, D. Sex trafficking, Russian infiltration, birth certificates, and pedophilia: A survey experiment correcting fake news. *Journal of Experimental Political Science*, Cambridge, v. 5, n. 2, p. 159-164, 2018.

POSETTI, J.; MATTHEWS, A. A short guide to the history of ‘fake news’ and disinformation. *International Center for Journalists*, Washington, DC, v. 7, p. 1-19, 2018.

POTTER, W. J.; BYRNE, S. *Media literacy*. Thousand Oaks: Sage Publications, 1998.

RAUCHFLEISCH, A.; KAISER, J. The False positive problem of automatic bot detection in social science research. *PloS one*, San Francisco, v. 15, n. 10, p. e0241045, 2020.

REGIS, F.; AUDI, G.; MAIA, A. Do letramento ao letramento midiático: práticas e competências cognitivas na cibercultura. In: CONGRESSO BRASILEIRO DE CIÊNCIAS DA COMUNICAÇÃO, 37., Foz do Iguaçu, 2014. Foz do Iguaçu: Intercom, 2014. n. p. Disponível em: <http://www.intercom.org.br/papers/nacionais/2014/resumos/R9-1412-1.pdf>. Acesso em: maio 2022.

ROUDIK, P. *et al.* Initiatives to Counter Fake News in Selected Countries. *The Law Library of the Congress*, [s. l.], 2019. Disponível em: <https://digitalcommons.unl.edu/scholcom/179/>. Acesso em: maio 2022.

SCHAFFNER, B. F.; ROCHE, C. Misinformation and motivated reasoning: Responses to economic news in a politicized environment. *Public Opinion Quarterly*, [s. l.], v. 81, n. 1, p. 86-110, 2016.

SCHWARZ, N.; NEWMAN, E.; LEACH, W. Making the truth stick & the myths fade: Lessons from cognitive psychology. *Behavioral Science & Policy*, [s. l.], v. 2, n. 1, p. 85-95, 2016.

SCHWARZ, N.; NEWMAN, E.; LEACH, W. *et al.* Metacognitive experiences and the intricacies of setting people straight: Implications for debiasing and public information campaigns. *Advances in Experimental Social Psychology*, [s. l.], v. 39, p. 127-161, 2007.

SMITH, N. Schoolkids in Taiwan Will Now Be Taught How to Identify Fake News. *Time*, [s. l.], 2017. Disponível em: <https://time.com/4730440/taiwan-fake-news-education/>. Acesso em: mar. 2021.

SOUSA JÚNIOR, J. H. *et al.* Da Desinformação ao Caos: uma análise das Fake News frente à pandemia do Coronavírus (COVID-19) no Brasil. *Cadernos de Prospeção*, Salvador, v. 13, n. 2, p. 331, 2020.

STANLEY, S. Misinformation and hate speech in Myanmar. *First Draft*, [s. l.], 2017. Disponível em: <https://firstdraftnews.org/latest/misinformation-myanmar/>. Acesso em: 23 set. 2020.

STARBIRD, K. Disinformation's spread: bots, trolls and all of us. *Nature*, Bethesda, v. 571, n. 7766, p. 449-450, 2019.

STENCEL, M.; LUTHER, J. Fact-checking count tops 300 for the first time. *Duke Reporter's Lab*, [s. l.], 2020. Disponível em: <https://reporterslab.org/fact-checking-count-tops-300-for-the-first-time/>. Acesso em: mar. de 2021.

SWIRE-THOMPSON, B. *et al.* Processing political misinformation: comprehending the Trump phenomenon. *Royal Society open science*, London, v. 4, n. 3, p. 160802, 2017.

THORSON, E. Belief echoes: The persistent effects of corrected misinformation. *Political Communication*, [s. l.], v. 33, n. 3, p. 460-480, 2016.

USCINSKI, J. E.; BUTLER, R. W. The epistemology of fact checking. *Critical Review*, [s. l.], v. 25, n. 2, p. 162-180, 2013.

VEIGA, E. A luta anônima de três brasileiros contra sites de fake news. *DW*, [s. l.], 2020. Disponível em: <https://www.dw.com/pt-br/a-luta-an%C3%B4nima-de-tr%C3%AAs-brasileiros-contra-sites-de-fake-news/a-54153272>. Acesso em: fev. 2021.

VIDON, F. Depois de perder 38 patrocínios com campanha do Sleeping Giants, Sikêra Jr. mantém anúncios da Caixa e do MEC nos seus intervalos. *EXTRA*, Rio de Janeiro, 2021. Disponível em: <https://extra.globo.com/noticias/brasil/depois-de-perder-38-patroc%C3%ADnios-com-campanha-do-sleeping-giants-sikera-jr-mantem-an%C3%BAncios-da-caixa-do-mec-nos-seus-intervalos-25111984.html>. Acesso em: 31 maio 2022.

WALTER, N. *et al.* Fact-checking: A meta-analysis of what works and for whom. *Political Communication*, [s. l.], v. 37, n. 3, p. 350-375, 2020.

WARDLE, C.; DERAKHSHAN, H. Information disorder: Toward an interdisciplinary framework for research and policy making. *Council of Europe Report*, [s. l.], v. 27, 2017. Disponível em: <https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwjOisHDzor4AhUxA7kGHfzFANGQFnoECBAQAQ&url=https%3A%2F%2Frm.coe.int%2Finformation-disorder-toward-an-interdisciplinary-framework-for-research%2F168076277c&usg=AOvVaw0ryt6Dg6xHJ8jw6ZMSKOHM>. Acesso em: maio 2022.

WEEKS, B. E. Emotions, partisanship, and misperceptions: How anger and anxiety moderate the effect of partisan bias on susceptibility to political misinformation. *Journal of Communication*, [s. l.], v. 65, n. 4, p. 699-719, 2015.

WOOD, T.; PORTER, E. The elusive backfire effect: Mass attitudes' steadfast factual adherence. *Political Behavior*, [s. l.], v. 41, n. 1, p. 135-163, 2019.

WORLD ECONOMIC FORUM. Global Risks Report. *World Economic Forum*, [s. l.], 2018. Disponível em: <https://reports.weforum.org/global-risks-2018/preface-2/>. Acesso em: mar. 2021.

ZELLERS, R. *et al.* *Defending Against Neural Fake News*, [s. l.], p. 1-21, 2019. Disponível em: <https://arxiv.org/abs/1905.12616>. Acesso em: maio 2022.

Hate speeches in the context of information disorder: a reflection towards solutions

Abstract: The informationalism era has brought new dynamics to hate speech. It is therefore necessary to understand how the digital world impacts this kind of manifestation. This article presents a review on mitigation strategies for hate speech in the context of the so-called information disorder, a term that describes the production and circulation of malicious or false information, particularly in the digital realm. Through this bibliographic research, we aim to offer an introductory text listing possible solutions to these problems. We detail four categories of measures to combat hate speech and information disorder: educational (in which we analyze literacy strategies); classificatory (in which we consider fact-checking initiatives and also algorithmic solutions for the detection of malicious or false content); normative (in which we analyze internal and public regulation of social media); and economic (in which we consider social mobilization for defunding producers or disseminators of malicious or false content). In analyzing each of those strategies, we reflect on their potential and also on their limitations, defending the thesis that the complexity of the problems we face requires a combination of different actions.

Keywords: Hate speech; Information disorder; Disinformation; Fake news; Literacy.

Recebido em: 20 de janeiro de 2021.

Aceito em: 26 de maio de 2022.