

Desinformação nas plataformas: ações de combate adotadas pelo *Twitter* durante a pandemia da Covid-19

Luciana Miranda Costa

Doutora; Universidade Federal do Rio Grande do Norte, Natal, RN, Brasil;
lmirandaeua@hotmail.com; <https://orcid.org/0000-0003-3843-4499>

Lizete Barbosa da Nóbrega

Mestra; Universidade Federal do Rio Grande do Norte, Natal, RN, Brasil;
liz.nobrega4@gmail.com; <https://orcid.org/0000-0002-2915-5220>

Carolina Toscano Maia

Mestra; Universidade Federal do Rio Grande do Norte, Natal, RN, Brasil;
caroltmaia@gmail.com; <https://orcid.org/0000-0001-9825-1948>

Resumo: Neste artigo propomos uma reflexão crítica à luz do capitalismo de plataformas sobre as ações realizadas pelo *Twitter* para combater a desinformação durante a pandemia da Covid-19. Para isso, contextualizamos o fenômeno da desinformação nesse período e realizamos um estudo de caso a partir de pesquisa documental das atividades do *Twitter*. Destacamos que a estratégia de enfrentamento da plataforma se baseia em duas perspectivas principais: frear informações enganosas por meio de moderação de conteúdo e fornecer informações seguras. Por fim, tensionamos as propostas, apontando o perigo de um solucionismo algorítmico e a limitação de algumas ações da plataforma.

Palavras-chave: Desinformação; Twitter; Pandemia; Covid-19; Alfabetização midiática

1 Introdução

A Organização Mundial da Saúde (OMS) anunciou em 11 de março de 2020 a pandemia da Covid-19, doença respiratória causada por um novo tipo de vírus identificado como Sars-Cov-2 (EBC, 2020). Os primeiros casos de infecção foram localizados em Wuhan, na província de Hubei na China, que rapidamente se espalharam por todo o mundo. E os altos números de contágio e de mortes

devido ao vírus, além do agravamento da circulação de informações falsas ou equivocadas que ainda prejudicam o combate à doença, foram evidenciados.

Nesse sentido, a OMS e a Organização Pan-Americana da Saúde emitiram alertas sobre esse tipo de conteúdo chamando a atenção para terminologias que rapidamente ganharam o noticiário: a “infodemia”, que corresponde a “um excesso de informações, algumas precisas e outras não, que tornam difícil encontrar fontes idôneas e orientações confiáveis quando se precisa” (OPAS, 2020, p. 2), e a “desinformação”, definida como “uma informação falsa ou imprecisa cuja intenção deliberada é enganar” (OPAS, 2020, p. 2), causando distúrbios capazes de agravar a pandemia.

A circulação de conteúdos que indicam falsos tratamentos para a doença ou que sugerem a automedicação como estratégia de prevenção, minimizando os efeitos do vírus, podem provocar uma falsa sensação de segurança na população. Além disso, existem publicações capazes de despertar sentimentos como medo, tristeza, agressividade ou descrença, numa dimensão simbólica que se estende às ações (GOULART; MUÑOZ, 2020), comprometendo o cumprimento de orientações seguras para prevenção e tratamento da Covid-19.

A crença nesse tipo de conteúdo também é relacionada a uma cultura da “pós-verdade”, que se refere a circunstâncias nas quais os apelos à emoção e à crença pessoal são mais influentes na formação da opinião pública do que os fatos objetivos (OXFORD, 2016). Condição reforçada pela afirmação de que “é claro e nítido como o *Facebook* e o *Twitter*, além dos *sites* de buscas, a exemplo do *Google*, aceleraram e fortaleceram a pós-verdade” (BUCCI, 2019, p. 59). O autor baseia tal afirmação em dois motivos: o primeiro pelas características das plataformas que permitem maior velocidade e alcance das informações, já o segundo, pela questão econômica em que a desinformação se reverteria em ganhos financeiros para as plataformas, a partir de seu modelo de negócios (SRNICEK, 2018).

A desinformação não é um fenômeno necessariamente novo. Mentiras por disputa de poder e estratégias políticas pautadas em “fatos alternativos” sempre existiram na história da humanidade (KAKUTANI, 2018; KEYES, 2018). O que muda na construção do fenômeno atual é a sua infraestrutura material, impulsionada por plataformas digitais e seus modelos de negócios,

além do contexto sociopolítico. O grande *boom* das chamadas *fake news* se dá efetivamente em 2016, com eventos como a eleição do presidente americano Donald Trump nos Estados Unidos, em 8 de novembro, e as tratativas sobre o “Brexit” (contração de “*British Exit*”), consumado em 31 de dezembro de 2020, com a saída do Reino Unido da União Europeia.

O debate político e social em torno das *fake news* cresceu e o próprio termo começou a ser questionado. Tandoc Jr., Lim e Ling (2018) fazem uma categorização do termo a partir da revisão de trabalhos acadêmicos, em seis tipos: sátira, paródia, fabricação, manipulação, publicidade e propaganda. Por também compreender que é um fenômeno complexo e multifacetado, optamos por utilizar a nomenclatura de desinformação proposta por Wardle e Derakhshan (2017), que a definem como um tipo de desordem informacional que se apresenta a partir de “informações falsas conscientemente compartilhadas para causar danos a uma pessoa, grupo, organização ou país” (WARDLE; DERAKHSHAN, 2017, p. 20, tradução nossa¹).

A desinformação, no contexto atual, pode ser associada ao extrativismo de dados no qual informações falsas circulam e se adequam com muito mais intensidade ao meio digital, em total consonância a modelos de negócios baseados em cliques “que foram aperfeiçoados pelos gigantes extrativistas de dados” (MOROZOV, 2018, p. 169). As plataformas atuam, desta forma, como mediadores ativos, em uma base tecnológica digital conectada, possibilitando a comunicação e transações entre indivíduos e organizações. Elas oferecem serviços, desde a coleta ao processamento de dados, que se baseiam em conexões estabelecidas na *internet* e que são marcados por efeitos de rede (VALENTE, 2019).

Por essa razão, as plataformas digitais têm sido indicadas como importantes atores na construção de um cenário favorável à circulação de conteúdos enganosos pela quantidade expressiva de usuários que possuem, por sua facilidade e baixo custo para fabricação de informações, ampla visibilidade e, principalmente, por seu modelo de negócios, que lucra a partir do alto engajamento de publicações, independentemente de serem verdadeiras ou falsas. Essas informações costumam atrair a atenção dos usuários pelo apelo emocional que despertam e quando são monetizadas tornam-se lucrativas.

Ninguém discorda de que ao menos um pedaço da responsabilidade pela desvalorização da verdade factual cabe às plataformas sociais e à *internet*, onde se instalaram confortavelmente as forças dedicadas à produção das notícias fraudulentas. (BUCCI, 2019, p. 51).

Portanto, a desinformação pode se espalhar a partir de canais ou estratégias também utilizados para divulgar conteúdos jornalísticos. “A circulação de desinformação é facilitada nas mídias sociais pelas próprias *affordances* dessas plataformas” (RECUERO; SOARES; ZAGO, 2021, p. 6). No caso do *Twitter*, por exemplo, os autores afirmam que as mensagens curtas da plataforma facilitam a difusão e visibilidade da desinformação, por meio da utilização de “títulos caça-cliques” (*click baits*) (RECUERO; SOARES; ZAGO, 2021).

Neste artigo propomos um olhar crítico sobre as ações de combate à desinformação (WARDLE; DERAKHSHAN, 2017) adotadas pelo *Twitter* durante a pandemia da Covid-19. A escolha pelo *Twitter* se deu, principalmente, pelo seu expressivo e crescente número de usuários. Em 2020, a plataforma alcançou a marca de 186 milhões de usuários em todo o mundo, um aumento de 34% se comparado a 2019 (G1, 2020b). Ela também se tornou um meio muito utilizado para comunicação, frequentemente baseada em informações falsas, de políticos como o presidente brasileiro Jair Bolsonaro e o ex-presidente americano Donald Trump (OTT; DICKINSON, 2019; SANTOS; CIOCCARI; MORAES, 2020); além de possibilitar alto engajamento no debate sobre o coronavírus (DAPP FGV, 2020).

A partir dos resultados alcançados com a análise das informações, compreendemos que há um claro conflito de interesses nas ações realizadas pelo *Twitter*, que apesar de comportarem aspectos positivos para a sociedade, são ainda insuficientes para combater o fenômeno, pois ampliam o poder das *Big Techs*² em relação à moderação de conteúdos.

2 Metodologia

Para alcançar os objetivos propostos, realizamos um estudo de caso (YIN, 2005; GIL, 2018), com o objetivo de investigar com mais profundidade o fenômeno da desinformação no contexto cotidiano, a partir de elementos de caráter descritivo

e interpretativo. Neste processo buscamos problematizar a atuação das plataformas no enfrentamento à desinformação durante a pandemia da Covid-19, com um recorte específico no *Twitter*. Assim, foram estabelecidos os protocolos e definições que orientaram os próximos passos do estudo empreendido.

Partimos de uma coleta de dados baseada nos métodos que compõem a pesquisa bibliográfica e documental. Para tanto, realizamos uma observação sistemática do *blog* da plataforma (<https://blog.twitter.com>) desde o início da pandemia, decretada pela OMS em março de 2020, coletando informações no *site* e comunicados oficiais da empresa. Esses comunicados foram publicados por diretores e gestores da empresa, que buscavam detalhar as atualizações sobre as estratégias e abordagens adotadas especificamente sobre a doença (ROTH; PICKLES, 2020; GADDE; DERELLA, 2020), além de material institucional publicado pela plataforma em seu *blog* oficial (TWITTER, 2020a; 2020b).

Também consultamos as políticas de comunidade do *Twitter* (TWITTER, 2021d), que incluem as regras do que é ou não permitido na plataforma. Essa consulta se deu a fim de identificar quais ações já tinham sido tomadas e quais foram implementadas especificamente na pandemia da Covid-19 (TWITTER, 2021c). Por fim, ainda na fase de coleta, com o objetivo de contrapor as informações obtidas pelo *site* oficial da plataforma, testamos como usuários da mídia social algumas funcionalidades anunciadas a fim de caracterizar as alterações indicadas.

A partir do material coletado e de sua descrição, passamos à análise das atividades realizadas pelo *Twitter*, indicando de forma cronológica as ações tomadas e divulgadas pela plataforma em seus canais oficiais. Para tanto, levamos em conta proposições teóricas e conceitos sobre desinformação (WARDLE; DERA KHSHAN, 2017; TANDOC JR, LIM, LING, 2018), suas causas e potenciais saídas. De forma mais contundente, consideramos em nossa interpretação o significativo papel dos modelos de negócios (MOROZOV, 2018; BUCCI, 2019; SRNICEK, 2018) das plataformas que também lucraram com a desinformação.

3 A plataforma Twitter

A plataforma foi criada em 2006 e é caracterizada como uma rede social e serviço de *microblogging*, correspondendo a *blogs* com *posts* curtos, que inicialmente não poderiam exceder os 140 caracteres, sendo expandidos para 280 em 2017. Ela foi idealizada para que as pessoas pudessem se conectar em tempo real a partir do questionamento: o que você está fazendo? Para Berti *et al.* (2009), a grande aceitação da plataforma se deu principalmente pela obrigatoriedade de mensagens curtas, o que garantia a objetividade nos conteúdos publicados, assim como pela facilidade para alimentar o perfil do usuário a partir do computador ou *smartphone* em qualquer lugar ou momento do dia.

O crescimento do número de usuários do *Twitter* também está relacionado à popularidade que alcançou durante as eleições norte-americanas de 2008. Berti *et al.* (2009) destacam que o então candidato, Barack Obama, que possuía perfil na plataforma, foi por algum tempo o usuário com maior número de seguidores e postava frequentemente sobre suas atividades durante a campanha. Seguindo essa tendência, o dia da eleição também foi um marco, pois muitos usuários fizeram a cobertura em tempo real das votações e do processo de apuração através de postagens que podiam ser acessadas por qualquer pessoa.

O *Twitter* também pode ser caracterizado como uma plataforma publicitária (SRNICEK, 2018). Esse tipo de empresa lucra com a extração de dados dos usuários, gerados a partir da utilização do sistema, e da mineração e comercialização dos dados para anunciantes. Assim, a plataforma ganha com o tempo de permanência, atividades e interação dos usuários, estabelecendo mecanismos para que esse período de utilização seja cada vez maior.

Apesar de se colocar como uma empresa de tecnologia, assim como o *Google* e *Facebook*, o *Twitter* é também apontado como uma plataforma que possui cada vez mais decisões editoriais (BELL *et al.*, 2017). A partir de 2016, o *Twitter* começou a adotar mudanças no seu algoritmo de modo a personalizar o *feed* de notícias do usuário. Os *tweets*, que antes eram mostrados de forma cronológica inversa (os mais recentes primeiro), com a alteração, passaram a ser

mostrados com base nas interações, aparecendo em primeiro lugar “os *tweets* mais importantes de quem você segue” (BELL *et al.*, 2017).

Com essa mudança e a partir da “inteligência artificial”, o *Twitter* pauta o que aparece na *timeline* do usuário, a fim de prever, por meio de um modelo de pontuações, “o quão interessante e envolvente um *tweet* seria especificamente para você” (KOUMCHATZKY; ANDRYEYEV, 2017, tradução nossa³). A própria plataforma alerta para os impactos desses modelos, “escolher a abordagem certa para construir e executar modelos de previsão pode ter consequências profundas na experiência de cada indivíduo que usa o *Twitter*” (KOUMCHATZKY; ANDRYEYEV, 2017, tradução nossa⁴). Assim, com essa personalização, o *Twitter* pode aprimorar os dados que extrai e segmentar a publicidade ao seu público, agregando ainda mais valor. Como já mencionado, em 2020, a plataforma alcançou a marca de 186 milhões de usuários em todo o mundo. Essa popularidade também permite que a rede controle fluxos de comunicação em nível global.

Considerando o início da pandemia da Covid-19, um levantamento realizado pela Diretoria de Análise de Políticas Públicas da Fundação Getúlio Vargas (DAPP FGV, 2020), em março de 2020, já indicava o alto engajamento de conteúdos relacionados ao vírus Sars-Cov-2 no *Youtube*, *Facebook*, *Twitter* e *WhatsApp*.

Em relação ao *Twitter*, foram identificados 12,7 milhões de *tweets* sobre o novo coronavírus no Brasil. No período de 0h de 13 de março às 12h de 17 de março, houve 66,7% de interações, com debate sobre o vírus de modo mais abrangente (se distanciando da polarização política de direita ou esquerda) e com a participação de influenciadores digitais e personalidades públicas como “atores, com maior ênfase a publicações sobre cuidados com a saúde, mensagens de humor de tom informativo, comentários negativos sobre o presidente Bolsonaro e repercussão de casos internacionais de pessoas infectadas” (DAPP FGV, 2020). Outros 20,6% de interações correspondiam a articulações a partir da oposição ao governo e perfis partidários de esquerda.

[...] com forte presença de contas da imprensa, o que aumentou desde o fim de semana e permanece com foco crítico na postura do presidente em relação à pandemia e no debate político e econômico sobre o SUS, recursos públicos e proteção a populações vulneráveis (DAPP FGV, 2020, *online*).

Nesse contexto, o *Twitter* anunciou em 2020 atualizações para a remoção de conteúdos falsos ou potencialmente enganosos, capazes de causar danos às pessoas, assim como parâmetros para suspensão de contas (GADDE; DERELLA, 2020). Também incluiu mensagens de alerta com explicações sobre a doença e tratamentos, buscando se aproximar de iniciativas de *Media Literacy*.

No âmbito internacional, destaca-se a análise da Comissão Europeia relativa ao ano de 2020, com reconhecimento aos esforços das plataformas, incluindo o *Twitter*, para combate às informações falsas na pandemia da Covid-19 (UE, 2020). Esse esforço se deu a partir da divulgação de informações verdadeiras sobre a doença por meio de ferramentas como painéis ou *pop-ups* e do fornecimento de espaço publicitário gratuito para que as organizações governamentais e internacionais pudessem promover campanhas informativas sobre a pandemia.

4 As ações de combate à desinformação pelo *Twitter*

Desde 2017, como resposta ao fenômeno da desinformação dentro de sua rede, o *Twitter* já indicava algumas medidas consideradas protetivas, que iremos descrever a seguir.

“O *Twitter* tem intensificado o trabalho contra as tentativas de manipulação, incluindo *spam* e contas automatizadas mal-intencionadas, bem como quaisquer atividades que violem nossos Termos de Serviço” (TWITTER, 2017). A plataforma buscava detalhar como atuava para impedir a ação de contas inautênticas e robôs que tentavam pautar o debate público nesse espaço, influenciando, por exemplo, os assuntos mais comentados na rede. “Desde junho de 2017, conseguimos detectar uma média diária de 130.000 contas que tentaram manipular os *Trending Topics* - e tomamos medidas para evitar essa influência” (TWITTER, 2017). O foco era prevenir o envio de *spam* e padrões suspeitos.

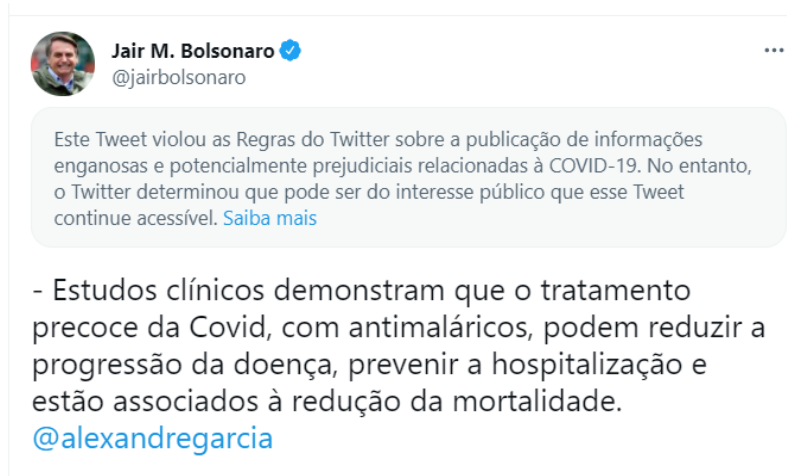
O *Twitter* intensificou suas ações de combate aos conteúdos falsos em fevereiro de 2020 (TWITTER, 2020a), em relação à chamada mídia manipulada e/ou sintética, como a plataforma a caracterizou em 2019. A rede entende como mídia sintética e manipulada “qualquer foto, áudio ou vídeo que tenha sido significativamente alterado ou produzido com a intenção de enganar as pessoas ou modificar seu significado original” (HARVEY, 2019). A ação também é conhecida como *deepfakes* ou *shallowfakes*. Ao categorizar a mídia como alterada ou fabricada segundo seus critérios, passou a aplicar um aviso na mensagem, o que, supostamente, diminuiria seu alcance.

A plataforma considera que a mídia é enganosa nas seguintes circunstâncias: (1) se o conteúdo foi substancialmente editado de uma maneira que altere fundamentalmente sua composição, sequência, tempo ou enquadramento; (2) qualquer informação visual ou auditiva (como novos trechos de vídeos, áudios com superposição de sons ou legendas modificadas) que foi adicionada ou removida; e (3) se a mídia representando uma pessoa real foi fabricada ou simulada (TWITTER, 2021a).

Em março de 2020 essa política foi expandida para “conteúdos que vão diretamente contra orientações para Covid-19” (GADDE; DERELLA, 2020). Para realizar ações de combate a conteúdos falsos, foi expandido o *prompt* (comando) de pesquisa da mídia a fim de que os usuários, ao pesquisar ‘coronavírus’ ou ‘covid-19’ no campo de busca da plataforma, pudessem encontrar com mais facilidade informações seguras sobre a doença. Para isso, estabeleceu-se uma parceria com entidades de saúde nacionais e internacionais, e foi reunido, em um só lugar, *tweets* com dicas e orientações sobre a doença. No mesmo anúncio dos *prompts* realizado no *site*, o *Twitter* informou o envio de créditos do *Ads for Good* (anúncios para o bem, em tradução livre) para instituições sem fins lucrativos que criassem campanhas de verificação, como o *International Fact-Checking Network*.

O *Twitter* passou a incluir aviso nos *tweets* dos usuários de que possivelmente aquela informação poderia ser enganosa, com uma marcação ao conteúdo e um aviso antes do compartilhamento e curtida. Incluiu também contextos adicionais ao assunto, além de reduzir a visibilidade do *tweet* e fornecer um *link* com mais esclarecimentos sobre o tema.

Figura 1 – Aviso do *Twitter* sobre informações enganosas sobre a pandemia



Fonte: Bolsonaro (2021).

Figura 2 – Aviso do *Twitter* sobre impossibilidade de engajamento com conteúdo enganoso



Fonte: Print da tela do *Twitter* em *tweet* considerado enganoso (BOLSONARO, 2021) capturado pelas autoras.

A plataforma adotou ainda a remoção de *tweets* que poderiam causar danos à população e, no caso da Covid-19, prejudicar o combate à doença, assim como expor as pessoas ao risco de contaminação a partir da desinformação. Com essa mudança, conteúdos que tenham capacidade de aumentar a exposição ao vírus, que possam causar complicações no desempenho do sistema público de saúde para lidar com a doença e/ou que possam levar à discriminação e isolamento de comunidades e/ou negócios com base na percepção de pertencimento a grupos marginalizados não poderão “ser compartilhados no *Twitter* e estão sujeitos a remoção. Contas que violam essa regra repetidamente podem ser suspensas permanentemente” (GADDE; DERELLA, 2020).

A fim de exemplificar quais conteúdos estão sujeitos à remoção, a plataforma especificou:

- a) negação das recomendações de autoridades de saúde locais ou globais;
- b) descrição de supostas curas alegadas para covid-19;
- c) descrição de tratamentos prejudiciais ou medidas de proteção ineficazes;
- d) negação de fatos científicos estabelecidos;
- e) afirmações específicas em torno das informações da Covid-19 que têm como objetivo manipular;
- f) afirmações específicas e não verificadas que incitam as pessoas a agir e causam pânico generalizado;
- g) afirmações específicas e não verificadas feitas por pessoas que se passam por um funcionário, organização ou governo de saúde;
- h) propagação de informações falsas ou enganosas sobre os critérios ou procedimentos de diagnóstico;
- i) declarações falsas ou enganosas sobre como diferenciar Covid-19 de outra doença;
- j) informações de que grupos específicos ou nacionalidades nunca são ou são mais suscetíveis à Covid-19;
- k) e afirmações de que grupos específicos ou nacionalidades são mais suscetíveis a Covid-19 (GADDE; DERELLA, 2020).

Para categorizar o trabalho, o *Twitter* informou que, por entender a ampla dimensão que a desinformação pode ter, dividiu esse conteúdo em três categorias: “informações enganosas”, “afirmações questionáveis” e “afirmações não confirmadas”. A primeira engloba afirmações comprovadamente falsas ou enganosas por autoridades de saúde, a segunda se configura como declarações em que a credibilidade é contestável ou desconhecida e, por fim, a terceira categoria aborda conteúdos ainda não verificados quando compartilhados.

Figura 3 – Tabela de propensão ao dano divulgada pelo *Twitter*

Informações enganosas	Aviso	Remoção
Afirmações questionáveis	Aviso	Alerta
Afirmações não confirmadas	Sem medida	Sem medida*
	Moderada	Grave
Propensão a dano		

Fonte: Roth; Pickles (2020).

Em maio de 2020, a plataforma informou que esses *tweets* seriam identificados a partir de “sistemas internos para monitorar proativamente o conteúdo relacionado à Covid-19. Esses sistemas ajudam a garantir que não estamos ampliando o alcance desses *tweets* com esses avisos e alertas” (ROTH; PICKLES, 2020). O *Twitter* também declarou que não realizará esse monitoramento sozinho: “continuaremos a contar com parceiros confiáveis para identificar conteúdos que provavelmente podem causar danos no mundo *offline*” (ROTH; PICKLES, 2020). Alguns desses parceiros da plataforma, como já mencionado, são autoridades de saúde pública e governos (GADDE; DERELLA, 2020). Até 14 de julho de 2020, foram removidos 14.900 *tweets* e foram notificadas mais de 4,5 milhões de contas com alertas e avisos sobre o conteúdo.

A partir dessa política, o *Twitter* já colocou alertas de mensagens falsas em *tweets* de figuras como o então presidente americano Donald Trump (COLOMÉ, 2020), em relação à Covid-19 e às eleições americanas, e já apagou *posts* do presidente Jair Bolsonaro e do ex-ministro do Meio Ambiente do Brasil, Ricardo Salles (G1, 2020a), que traziam conteúdos com desinformação e que incentivavam aglomerações, na contramão do que recomenda a OMS.

Figura 4 – Aviso do *Twitter* sobre exclusão de *posts* do presidente



Fonte: G1(2020a).

Nesses casos, o *Twitter* optou por não remover o *tweet*, mas, por entender que é de interesse público, mantê-lo oculto com “um aviso que contextualize a violação de regra e que permita que as pessoas cliquem para ver o conteúdo” (TWITTER, 2021b). Além disso, em janeiro de 2021, o *Twitter* decidiu suspender temporariamente a conta do filósofo Olavo de Carvalho, apoiador do presidente Jair Bolsonaro, por divulgar informações negacionistas sobre a pandemia do coronavírus.

Em março de 2021, a plataforma anunciou um sistema de aplicação de penalidades para contas que insistirem em infringir os termos de uso relacionados à Covid-19. A plataforma indica que quando houver uma violação não será realizada nenhuma ação na conta, apenas a marcação do *tweet*. Nos casos de duas ou três violações serão aplicadas 12 horas de bloqueio e com quatro violações será determinado sete dias de bloqueio. A partir de cinco violações será aplicada a suspensão permanente da conta (G1, 2021; TWITTER, 2021c). De acordo com o *Twitter*, as pessoas serão notificadas diretamente quando um aviso ou uma solicitação de remoção de um *tweet* implicar em medidas adicionais na conta.

Já em abril de 2021, a plataforma anunciou uma atualização para aprimorar os mecanismos de combate à desinformação sobre a vacina da Covid-19. A partir dela, são exibidos avisos sobre a vacinação na linha do tempo dos usuários, com *links* para uma página que reúne informações sobre a segurança, eficácia, disponibilidade e planos de imunização das pessoas de instituições como a OPAS, OMS, Anvisa e Instituto Butantã, assim como *links* de notícias sobre a vacinação e *posts* de perfis oficiais de governos estaduais, autoridades políticas e da saúde que tratam do calendário vacinal (LISBOA, 2021).

Figura 5 – Aviso do *Twitter* ao pesquisa sobre vacina na plataforma



Fonte: *Print* da tela do *Twitter* capturado pelas autoras ao pesquisar pelo termo 'vacina' na rede social.

Figura 6 – Informações sobre Covid-19 no explorar do *Twitter*



Fonte: Print da tela do *Twitter* capturado pelas autoras.

Ainda em 2021, em um compilado apresentado sobre sua política de combate à desinformação na pandemia, a plataforma reuniu atualizações e especificou o que é necessário em uma publicação para a etiqueta de conteúdo relacionado à Covid-19 ser inserida ou removida. Acrescentando que ela aborda quatro categorias de comportamentos e conteúdos enganosos: (1) informações falsas ou enganosas sobre a natureza do vírus (transmissão, suscetibilidade das pessoas ao vírus, sintomas e associação da pandemia ou das vacinas para Covid-19 a conspirações de forças do mal ou de poder); (2) informações falsas ou enganosas sobre eficácia e/ou segurança de medidas de prevenção, tratamentos ou outras precauções para mitigar ou tratar a doença (segurança ou eficácia de tratamentos ou medidas de prevenção, venda ou facilitação de medicamentos, efeitos ou impactos adversos do recebimento de vacinas, além de sua associação a tentativas deliberadas para causar danos ou controlar populações), (3) informações falsas ou enganosas sobre regulamentações oficiais, restrições ou isenções relacionadas a orientações de saúde (eficácia e a segurança de equipamentos de proteção individual, medidas de prevenção, orientações sobre

lockdown e quarentena, desenvolvimento, teste e aprovação de vacinas), (4) informações falsas ou enganosas sobre a prevalência do vírus, risco de infecção ou morte (informações relativas a resultados de teste, hospitalizações ou taxa de mortalidade, capacidade de o sistema de saúde pública lidar com a crise, pesquisas deturpadas ou infundadas que diminuem o significado da doença).

As consequências do compartilhamento do usuário vão variar dependendo “da gravidade, do tipo da violação e do histórico de violações anteriores cometidas pela conta” (TWITTER, 2021c). Além disso, também indica que um usuário não pode criar contas falsas que deturpem afiliação ou compartilhar conteúdo que represente falsa afiliação a um médico, funcionário ou agência de saúde pública, instituição de pesquisa ou que sugira conhecimento em questões da Covid-19 (TWITTER, 2021).

5 A política do Twitter contra a desinformação

Após a pesquisa documental, com viés exploratório e descritivo, das ações anunciadas pelo *Twitter* para combater a desinformação na plataforma, partimos para a análise dessas iniciativas à luz dos conceitos de capitalismo de plataformas (SRNICEK, 2018; MOROZOV, 2018) e do solucionismo algorítmico (MOROZOV, 2018). A estratégia do *Twitter* está ligada à oferta da “informação correta” ao usuário a partir de dois movimentos: frear informações enganosas por meio de moderação de conteúdo e fornecer informações seguras.

Iniciaremos nosso percurso analítico pela moderação de conteúdo e pela dimensão tecnológica/material da questão para, a seguir, entendermos como o estímulo a informações seguras sobre a doença pode contribuir para um processo de “alfabetização midiática”.

Como já mencionado, o modelo de negócios de plataformas publicitárias, como o *Twitter*, envolve a extração de dados dos usuários, a mineração/segmentação deles e venda para empresas anunciantes. Essa é uma das formas em que a publicidade *online* se apresenta e vem dominando o mercado das propagandas. Com esse modelo de negócio, cabe às plataformas usarem estratégias para prender a atenção dos usuários.

Quanto mais tempo os membros gastam em uma plataforma, mais anúncios uma empresa vende e mais lucro ela tem; e a forma de maximizar o engajamento é ‘coletando e analisando seus dados, usando isso para prever o que fará você reagir mais intensamente e, em seguida, dando-lhe mais disso’. Isso não apenas cria os filtros que isolam as pessoas em bolhas partidárias, mas também favorecem mensagens simplistas e provocativas. (KAKUTANI, 2018, p. 155).

Assim, os algoritmos trabalham de forma a criar um ambiente confortável para o usuário, com opiniões favoráveis ao seu pensamento, independentemente de sua veracidade. Como sintetiza Morozov:

O modelo de negócios da *Big Tech* funciona de tal maneira que deixa de ser relevante se as mensagens disseminadas são verdadeiras ou falsas. Tudo o que importa é se elas viralizam [...]. Sob a ótica das plataformas digitais, as *fake news* são apenas as notícias mais lucrativas. (MOROZOV, 2018, p. 11).

Dessa forma, com um modelo que incentiva a viralização (BELL *et al.*, 2017), a desinformação ganha espaço, o que torna difícil obter resultados efetivos no combate à desinformação, visto que o ambiente é propício para essa disseminação. Para Morozov (2018), que compreende as informações falsas como um subproduto do capitalismo, “a única solução que não se equivoca em seu diagnóstico, nem subjuga as elites, é repensar completamente os fundamentos do capitalismo digital” (MOROZOV, 2018, p. 186). O autor defende que a saída é repensar as redes de comunicação atuais e livrá-las do modelo alicerçado na coleta de dados.

Porém, uma vez que a questão do ‘extrativismo de dados’ faz parte do inconsciente político mais generalizado do mundo contemporâneo, dificilmente ela aparece nos debates políticos. Daí a maneira preferencial de lidarmos com as notícias falsas: reforçar ainda mais efetivamente a nossa confiança nos gigantes da tecnologia e dar-lhes ainda mais poder para identificar e distinguir o que é falso e o que é genuíno e verdadeiro. (MOROZOV, 2018, p. 169).

Portanto, mostra-se inviável, considerando o cenário atual, alcançar uma solução efetiva e definitiva para um problema que surge na infraestrutura digital e é multifacetado. O *Twitter*, como vimos, aposta em alguns fatores para tentar diminuir a circulação de informações enganosas, mas o realiza intensificando a mediação por meio de algoritmos, justamente como apontou Morozov (2018). Essa prática de solucionismo por meio dos algoritmos, que ajuda a amenizar,

mas não resolve o problema, é algo que se observa nas plataformas em geral, tanto no *Twitter*, como no *Facebook*, por exemplo (COSTA; NÓBREGA; MAIA, 2021).

Esse solucionismo, portanto, acaba por conceder às plataformas ainda mais força e poder na construção de um discurso proativo, mas apenas parcialmente efetivo para sociedade. Cabendo a moderação a uma empresa privada que é parte economicamente interessada no compartilhamento de desinformação, a remoção de conteúdos falsos por parte das próprias *Big Techs* torna-se, no mínimo, vulnerável (MARTINS, 2020), ainda que se reconheça a importância de frear a circulação de conteúdos falsos. Acrescente-se a isso a não menos grave “opacidade” das atividades realizadas por meio dos algoritmos (PASQUALE, 2015; CASTRO, 2020; HARTMANN; IUNES, 2020).

A ideia de que o problema será resolvido com mais informação e mais automação acaba por camuflar as causas ligadas ao modelo de negócio das plataformas e o *Twitter* não é exceção. Morozov (2018) defende que esse positivismo e fé nos algoritmos em determinar o que seria verdadeiro ou falso não aguça “a capacidade de as pessoas reconhecerem uma propaganda” (MOROZOV, 2018, p. 179).

No mínimo, isso pode apenas embotar a nossa capacidade de pensar criticamente sobre as informações que passaram pelo crivo da verificação algorítmica. Falso consenso, imposto por algoritmos baseados em dados falhos, é tão ruim quanto notícias falsas. (MOROZOV, 2018, p. 179).

Nesse sentido, destaca-se a importância da implementação de medidas por parte de governos e instituições sociais que superem uma ação reativa limitada à remoção ou ocultação de conteúdos, visando garantir transparência sobre os modelos de negócios e seu funcionamento, e incentivar/educar o usuário a lidar criticamente com os conteúdos em circulação, especialmente em plataformas digitais.

O panorama de “desordem de informação” que surge em situações críticas como a Covid-19 também coloca em evidência a necessidade de uma legislação que estabeleça diretrizes e penalidades para as empresas de tecnologia da informação e comunicação que potencializam comportamentos e conteúdos suspeitos nas redes sociais (embora não se assumam como empresas de conteúdo, o que tem sido questionado), além de ações voltadas aos cidadãos comuns, políticos e personalidades públicas de diversas áreas. (SANTOS; CAPRINO, 2020, p. 59).

Na pesquisa “*Fake news*: como as plataformas enfrentam a desinformação”, realizada pelo Interviços (2021), os autores concluíram, por meio da descrição e análise das medidas das plataformas, que “as informações são escassas e dispersas, de maneira que a aferição da concretude das ações fica restrita às alegações das empresas, aproximando-se mais de promessas não comprováveis” (BARBOSA; MARTINS; VALENTE, 2020). E, especificamente sobre o *Twitter*, ainda há uma ampla disseminação de desinformação sobre a pandemia na plataforma, como, por exemplo, a circulação de informações falsas sobre a substância hidroxicloroquina provocada por perfis apoiadores do governo Bolsonaro (ARAUJO; OLIVEIRA, 2020) e a disseminação de desinformação relacionada a uma possível cura para a pandemia, com discursos cada vez mais politizados e polarizados (RECUERO; SOARES, 2020).

6 Alfabetização midiática e divulgação científica no combate à desinformação

O reconhecimento da complexidade do fenômeno da desinformação, que se apresenta de modo dinâmico e persistente, demonstra os desafios para o seu enfrentamento. Dessa forma, é preciso desenvolver um conjunto de esforços contínuos para combater suas causas e efeitos. E, como já indicado, apesar do reconhecimento às ações adotadas pelas plataformas, como o *Twitter*, na luta contra a desinformação, elas ainda são limitadas e insuficientes.

Entre as diferentes competências necessárias para amenizar os impactos das informações falsas nas diversas esferas sociais, destaca-se a capacidade de lidar criticamente com os conteúdos recebidos, especialmente no âmbito digital. Assim, a *Media Literacy*, traduzida como Alfabetização Midiática para o português, desponta como uma possibilidade.

Guess *et al.* (2020) observam que a alfabetização em mídia digital captura de maneira útil as habilidades e competências necessárias para navegar com sucesso em um ambiente fragmentado e complexo pertencente ao ecossistema de informações. Para os autores, as pessoas obtêm cada vez mais notícias e informações em plataformas digitais que não possuem controles

editoriais tradicionais, permitindo que políticos e outros atores disseminem amplamente a desinformação, que ainda pode ser impulsionada pela ação dos algoritmos nos *feeds* dessas redes. Dessa forma, “sem as habilidades necessárias de alfabetização em mídia digital, as pessoas frequentemente são vítimas de alegações duvidosas que encontram neste contexto” (GUESS *et al.*, 2020, p. 15.536, tradução nossa⁵).

Os autores também ressaltam que embora os esforços para melhorar a alfabetização digital *online* não sejam a única solução para o problema, “eles podem ser uma maneira vital e econômica de reduzir a vulnerabilidade das pessoas aos conteúdos falsos” (GUESS *et al.*, 2020, p. 15.542, tradução nossa⁶), garantindo um acesso mais saudável às informações nos regimes democráticos. Em sentido complementar, destaca-se a importância dessa competência ser apoiada pelas instituições de ensino, fortalecendo a leitura crítica de informações das gerações futuras (CHAVES; MELO, 2019).

Como vimos, o *Twitter* tem contribuído para a alfabetização midiática através do fornecimento de informações mais seguras para os usuários, em suas ferramentas de busca e ao privilegiar veículos de comunicação em páginas como *Trending Topics*.

Figura 7 – Print de seção ‘Covid-19’ nos destaques do *Twitter*, em que a plataforma destaca as principais notícias do dia, além de conteúdos reforçando a confiança das vacinas



Fonte: *Print* da tela do *Twitter*, capturado pelas autoras em 20 de outubro de 2021, destacando os acontecimentos da CPI da Pandemia.

No contexto da pandemia da Covid-19, a plataforma também disponibiliza *links* de acesso a conteúdos criados por instituições consideradas confiáveis como a OPAS, OMS, Anvisa e Instituto Butantã, assim como inclui painéis ou *pop-ups* de informações verdadeiras sobre a doença. Já para o *retweet*, através do compartilhamento de conteúdos pertencentes a outros perfis, exibe o questionamento: quer ler o artigo primeiro?

Figura 8 – Print de tela para *retweet* no Twitter



Fonte: Print da tela do Twitter capturado pelas autoras.

Ainda que esse incentivo seja incapaz de reparar toda a ambiência propícia à circulação de desinformação permitida pela plataforma, contribui com o debate sobre a necessidade de inclusão de diferentes agentes para a construção de um processo de alfabetização midiática contínuo.

Para ampliação das boas práticas de combate aos conteúdos enganosos, no entanto, é preciso que se forme uma rede que envolva diversos atores sociais, como escolas, governo, sociedade e organizações de mídia (SPINELLI; SANTOS, 2020). Diante da intensa circulação de informações e atualizações frequentes, capacitar a sociedade para a diferenciação de informações verdadeiras e falsas, de fatos e opiniões, assim como de preconceitos e distorções, demanda domínios específicos.

Singer (2019) também defende que para enfrentar sucessivos ataques de desinformação é preciso uma ação articulada em nível governamental, empresarial e dos indivíduos. Ele ressalta a necessidade de se investir em “alfabetização digital” a partir das próprias plataformas, para atuar proativamente com a intenção de alertar os usuários contra os perigos da desinformação. O autor enfatiza que o que viraliza ou deixa de viralizar pode moldar resultados de eleições e até mesmo de guerras. Para ele, as plataformas

devem alterar seus algoritmos para elevar ou reduzir o alcance de um conteúdo, “mas isso só vai acontecer quando elas saírem do modo de negação ativa. Não dá para ser proativo se você insiste que isso não é um problema” (SINGER, 2019, p. 105), ou quando, acrescenta-se, se lucra com o problema, gerando resistência e pouca transparência em políticas de combate à desinformação.

A partir de um conjunto amplo e multifacetado de medidas, a Unesco recomendou quatro campos de atuação para o combate a desinformação na pandemia: (1) identificação da desinformação, com respostas de monitoramento, de verificação de fatos e investigativas; (2) produtores e distribuidores, com respostas legislativas, pré-legislativas e políticas, além de campanhas nacionais e internacionais contra a desinformação; (3) produção e distribuição, com respostas curatoriais, técnicas, algorítmicas e econômicas; (4) apoio aos públicos-alvo da desinformação, através de respostas éticas, normativas e educacionais, assim como esforços de empoderamento e credibilidade (POSETTI; BONTCHEVA, 2020).

Outra via de enfrentamento da desinformação na pandemia está baseada no fornecimento de informações especializadas e seguras à sociedade. Para Câmara e Braighi (2020), as pessoas que têm acesso a informações precisas e verdadeiras sobre medidas de prevenção e tratamento da Covid-19, possuem mais chances de preservar a própria saúde. No entanto, a própria comunicação da ciência enfrenta uma série de desafios, envolvendo disputas de narrativas tanto na mídia tradicional quanto nas mídias sociais, com a entrada de novos atores nesses espaços digitais, como os próprios cientistas ou as mediações algorítmicas (OLIVEIRA, 2020; CÂMARA; BRAIGHI, 2020).

Os autores Câmara e Braighi (2020), assim como Dantas e Deccache-Maia (2020), evidenciam a necessidade da adoção e intensificação de estratégias contra a desinformação baseadas na divulgação científica, capazes de viralizar informações balizadas pela verdade científica.

Podemos atingir esse objetivo com atividades de divulgação científica, cujo propósito é estender o acesso aos conhecimentos científicos produzidos, estimulando o senso crítico e a alfabetização científica do público em geral. (DANTAS; DECCACHE-MAIA, 2020, p. 3).

Enquanto a comunidade científica tem respondido de forma efetiva à crise sanitária provocada pelo surto em âmbito global, coube a população também repensar suas práticas. “Para isso, estar bem informado e agir adequadamente tornou-se essencial para a autoproteção, bem como da sociedade” (SOUZA; BRANDÃO, 2020, p. 8). As autoras alertam que a desinformação deixa as pessoas mais vulneráveis à doença, despertando o medo, o caos e a histeria.

A forma como muitas *fake news* são divulgadas, fazendo uso de fontes pseudocientíficas, dificulta a identificação, por grande parte da população, da falta de respaldo e da inverdade que encobre a má fé por trás dos seus conteúdos. Não há como garantir que a recepção dessas mensagens seja no sentido da crença cega, mas o aumento de atividades de divulgação científica pode criar um público mais atento e exigente com os conteúdos, sabendo minimamente identificar aquilo que vale a pena ser lido. (DANTAS; DECCACHE-MAIA, 2020, p. 12).

Desta forma, a divulgação científica se apresenta como um potencializador para um maior senso crítico da realidade pela população, tornando-a menos vulnerável às *fake news* (DANTAS; DECCACHE-MAIA, 2020). Nesse sentido, ela pode ser associada a um ato de resistência, favorável à construção da geração do conhecimento e à defesa da própria comunidade (CÂMARA; BRAIGHI, 2020), embora ainda esbarre em várias dificuldades, como a de despertar o interesse da população em ter contato com seus conteúdos (DANTAS; DECCACHE-MAIA, 2020).

7 Considerações finais

Neste artigo buscamos refletir e problematizar sobre algumas das razões que tornam o *Twitter* um cenário favorável à circulação de conteúdos enganosos. Entre elas, destaca-se a quantidade expressiva de usuários que possui, condição que promove uma ampla visibilidade aos conteúdos disponibilizados, além de sua alta popularidade, que também permite o controle de fluxos de comunicação em nível global. Apenas no mês de março de 2020 foram identificados 12,7 milhões de *tweets* sobre o novo coronavírus no Brasil, com a participação de personalidades públicas e instituições oficiais nesse contexto (DAPP FGV, 2020).

A facilidade e baixo custo para produção de informações e o formato curto das mensagens, facilmente manipulado pela lógica caça-cliques, também se somam ao cenário. Mas é principalmente por meio de seu modelo de negócios, com a extração de dados dos usuários, a mineração/segmentação deles e a venda para empresas anunciantes, que o *Twitter* e outras plataformas como *Facebook* lucram a partir do alto engajamento de publicações, independentemente de serem verdadeiras ou falsas.

O *Twitter* adota medidas de combate à desinformação a partir de duas perspectivas principais: a primeira em função da moderação de conteúdos, limitando o alcance e a visibilidade das informações falsas. A segunda está voltada para o fornecimento de informações seguras sobre a doença, através de espaços disponibilizados para divulgação de orientações de organizações como a OMS e de *links* direcionados a informações confiáveis sobre a pandemia, com estímulo à leitura crítica.

No entanto, esse tipo de solucionismo, por meio da moderação de conteúdos, implica em ratificar mais poder às próprias plataformas na construção das narrativas que irão circular na sociedade, minimizando-se que se trata de empresas privadas que possuem seus próprios interesses econômicos e que lucram com a desinformação por meio do alto engajamento de publicações.

Embora se reconheçam as ações descritas do *Twitter* como necessárias no combate ao fenômeno e como um incentivo importante à leitura crítica dos conteúdos, por meio de uma aproximação às práticas da alfabetização midiática, a iniciativa é bastante limitada diante do grande volume de informações falsas e de toda facilidade ainda existente para circulação de desinformação, que conta com a cumplicidade da própria plataforma, em virtude de seu modelo de negócios.

Financiamento

Este artigo foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES).

Referências

- ARAÚJO, R.; OLIVEIRA, T. Desinformação e mensagens sobre a hidroxicloroquina no Twitter: da pressão política à disputa científica. **Atoz: novas práticas em informação e conhecimento**, Curitiba, v. 9, n. 2, p. 196-205, 2020.
- BARBOSA, B.; MARTINS, H.; VALENTE, J. Fake news: como as plataformas enfrentam a desinformação. **Intervozes**, São Paulo, 2020.
- BELL, E; OWEN, T.; BROWN, P. D.; HAUDA, C.; RASHIDIAN, N. A imprensa nas plataformas: como o Vale do Silício reestruturou o jornalismo. **Revista de Jornalismo da ESPM**, São Paulo, n. 2, p. 48-83, 2017.
- BERTI, O.; CARDOSO, L.; MELO, L.; RODRIGUES, A. A história do *Twitter* no Piauí – de simples *microblogging* a instrumento de utilização jornalística. In: ENCONTRO NACIONAL DE HISTÓRIA DA MÍDIA, 7., 2009, Fortaleza. **Anais [...]**. Fortaleza: Alcar, 2009. p. 1-14.
- BOLSONARO, J. **Estudos clínicos demonstram que o tratamento precoce da Covid, com antimaláricos, podem reduzir a progressão da doença, prevenir a hospitalização e estão associados à redução da mortalidade.** @alexandregarcia. 15 jan. 2021. Twitter: @jairbolsonaro. Disponível em: <https://twitter.com/jairbolsonaro/status/1350149912009334784>. Acesso em: 13 jul. 2021.
- BUCCI, E. **Existe democracia sem verdade factual?** São Paulo: Estação das Letras e Cores, 2019.
- CÂMARA, M.; BRAIGHI, A. informação contra o coronavírus: uma proposta de midiativismo científico dos blogs de ciência da Unicamp (?). In: ENCONTRO VIRTUAL DA ABCIBER 2020, 2020. **Anais [...]**. [S. l.]: Associação Brasileira de Pesquisadores em Ciberultura, 2020.
- CASTRO, J. C. L. Controle via agência em plataformas algorítmicas. **Galáxia**, São Paulo, n. 44, p. 144-157, 2020.
- CHAVES, M.; MELO, L. Educação midiática para notícias: histórico e mapeamento de iniciativas para combater a desinformação por meio da educação. **Revista Mídia e Cotidiano**, Rio de Janeiro, v. 13, n. 3, p. 62-82, 2019.
- COLOMÉ, J. P. **Por que o conflito entre Trump e o Twitter esconde uma das batalhas de nosso tempo.** [S.l.] El País, 31 mai. 2020. Disponível em: <https://brasil.elpais.com/tecnologia/2020-05-31/por-que-o-conflito-entre-trump-e-o-twitter-esconde-uma-das-batalhas-de-nosso-tempo.html>. Acesso em: 13 jul 2021.

COSTA, L; NÓBREGA, L; MAIA, C. Combate à desinformação na pandemia da Covid-19: a reação das plataformas digitais. **Revista Eptic**, São Cristovão, v. 23, n.1, p. 162-177, jan./abr. 2021.

DANTAS, L.; DECCACHE-MAIA, E. Divulgação científica no combate às Fake News em tempos de Covid-19. **Research, Society and Development**, Vargem Grande Paulista, v. 9, n. 7, p. 1-18, 2020.

DAPP FGV. Análise Estratégia de Mídias Digitais. **Sala de Democracia Digital**. [S. l.], 2020. Disponível em: https://static.poder360.com.br/2020/03/COVID-19_REPORT_2020_03_18.pdf. Acesso em: 13 jul. 2021.

EBC. **Organização Mundial da Saúde declara pandemia de coronavírus**. [S. l.], 11 mar. 2020. Disponível em: <https://agenciabrasil.ebc.com.br/geral/noticia/2020-03/organizacao-mundial-da-saude-declara-pandemia-de-coronavirus>. Acesso em: 13 jul 2021.

G1. **Twitter apaga publicações de Jair Bolsonaro por violarem regras da rede**. [S. l.], 29 mar. 2020a. G1: Política.

G1. **Mesmo com aumento de usuários, Twitter tem prejuízo de US\$ 1,23 bilhão no 2º trimestre**. [S. l.], 23 ago. 2020b. G1: Economia.

G1. **Twitter vai desativar conta que desrespeitar 5 vezes as regras de desinformação sobre Covid-19**. [S. l.], 1 mar. 2021. G1: Economia.

GADDE, V.; DERELLA, M. **Uma atualização sobre nossa estratégia contínua durante a COVID-19**. Blog Twitter, [s. l.], 16 mar. 2020.

GIL, A. C. **Como elaborar projetos de pesquisa**. 6. ed. São Paulo: Atlas, 2018.

GOULART, A. H.; MUÑOZ, I. K. Desinformação e pós-verdade no contexto da pandemia da Covid-19: um estudo das práticas informacionais no Facebook. **Liinc em Revista**, Rio de Janeiro, v. 16, n. 2, p. e5397, 2020.

GUESS, A.; LERNER, M.; LYONS, B.; MONTGOMERY, J.; NYHAN, B.; REIFLER, J.; SIRCAR, N. A digital media literacy intervention increases discernment between mainstream and false news in the United States and India. **Pnas**, Washington, DC, v. 117, n. 27, p. 15536-15545, 2020.

HARTMANN, I. A.; IUNES, J. Fake News no contexto de pandemia e emergência social: os deveres e responsabilidades das plataformas de redes sociais na moderação de conteúdo online entre a teoria e as proposições legislativas. **Direito Público**, Brasília, DF, v. 17, n. 94, p. 388-414, nov. 2020. (Dossiê Especial Covid-19).

HARVEY, D. **Ajude-nos a criar nossa abordagem para mídia sintética e manipulada**. Blog Twitter, [s. l.], 11 nov. 2019.

KAKUTANI, M. **A morte da verdade**. Rio de Janeiro: Intrínseca, 2018.

KEYES, R. **A era da pós-verdade**: desonestidade e enganação na vida contemporânea. Petrópolis, RJ: Vozes, 2018.

KOUMCHATZKY, N; ANDRYEYEV, A. **Using deep learning at scale in Twitter's timelines**. Blog Twitter: Engineering, [s. l.], 9 may 2017.

Disponível em:

https://blog.twitter.com/engineering/en_us/topics/insights/2017/using-deep-learning-at-scale-in-tweets-timelines. Acesso em: 13 jul. 2021.

LISBOA, A. **Twitter anuncia reforço no combate à desinformação sobre vacinas da COVID-19**. Canal Tech, 23 abr. 2021. Disponível em:

<https://canaltech.com.br/redes-sociais/twitter-anuncia-reforco-no-combate-a-desinformacao-sobre-vacinas-da-covid-19-183703/>. Acesso em: 13 jul 2021.

MARTINS, H. Muito além das fake news: A desinformação em meio à crise social. **Le Monde diplomatique**, São Paulo, v. 14, n. 155, jun. 2020.

MOZOROV, E. **Big Tech**: a ascensão dos dados e a morte da política. São Paulo: Ubu Editora, 2018.

OLIVEIRA, T. Desinformação científica em tempos de crise epistêmica: circulação de teorias da conspiração nas plataformas de mídias sociais. **Fronteiras**, São Leopoldo, v. 22, n. 1, p. 21-35, jan./abr. 2020.

OXFORD LANGUAGES. **Pós-verdade**: a palavra do ano. Oxford, 2016.

OPAS. **Folha informativa sobre Covid-19**. [S. l.], 2020. Disponível em:

https://www.paho.org/bra/index.php?option=com_content&view=article&id=6101:covid19&Itemid=875. Acesso em: 13 jul. 2021.

OTT, B.; DICKINSON, G. **The Twitter Presidency**: Donald J. Trump and the politics of white rage. New York: Routledge, 2019. Disponível em:

<https://www.taylorfrancis.com/books/mono/10.4324/9780429054259/twitter-presidency-brian-ott-greg-dickinson>. Acesso em: 14 jul. 2021.

PASQUALE, F. **The black box society**: the secrets algorithms that control money and information. Cambridge: Harvard University Press, 2015.

POSETTI, J.; BONTCHEVA, K. **Desinfodemia**: decifrar a desinformação sobre a Covid-19. [S. l.]: UNESCO, 2020.

RECUERO, R.; SOARES, F. O discurso desinformativo sobre a cura do COVID-19 no Twitter: estudo de caso. **E-Compós**, Brasília, DF, v. 24, p. 1-29, jan./dez. 2021.

RECUERO, R.; SOARES, F.; ZAGO, G. Polarização, hiperpartidarismo e câmaras de eco: como circula a desinformação sobre Covid-19 no Twitter. **Contracampo**, Niterói, v. 40, n. 1, jan./abr. 2021.

ROTH, Y.; PICKLES, N. **Atualizando nossa abordagem a informações enganosas**. Blog Twitter, [s. l.], 11 maio 2020.

SANTOS, M.; CAPRINO, M. Covid-19 e desinformação: ações de fact checking e educação midiática. **Comunicação & Inovação**, São Caetano do Sul, v. 21, n. 47, p. 39-62, 2020.

SANTOS, R.; CIOCCARI, D.; MORAES, T. O clã Bolsonaro e o Twitter: comunicação política e influência na rede social. **Mediapolis – Revista de Comunicação, Jornalismo e Espaço Público**, Coimbra, n. 10, p. 65-81, 2020.

SINGER, P. “Guerra de likes”: precisamos dominar as ferramentas e fazer a verdade viralizar. In: BARBOSA, M. (org.). **Pós-verdade e fake news: reflexões sobre a guerra de narrativas**. Rio de Janeiro: Cobogó, 2019. p. 97-108.

SOUZA, R.; BRANDÃO, R. Divulgação científica para combater notícias falsas sobre a Covid-19: um projeto de extensão do Instituto Federal da Bahia. **Science and Knowledge in Focus**, Macapá, v. 3, n. 2, p. 7-24, dez. 2020.

SPINELLI, E.; SANTOS, J. Alfabetização midiática na era da desinformação. **ECCOM**, Lorena, v. 11, n. 21, p. 147-164, jan./jun. 2020.

SRNICEK, N. **Capitalismo de plataformas**. Buenos Aires: Caja Negra Editora, 2018.

TANDOC JR, E.; LIM, Z.; LING, R. Defining “fake news”. **Digital Journalism**, Oxon, v. 6, n. 2, p. 137-153, 2018.

TWITTER. **Twitter detalha medidas contra uso indevido de robôs e desinformação**. Blog Twitter, [s. l.], 28 set. 2017.

TWITTER. **Construindo regras publicamente**: nossa nova abordagem para mídia sintética e manipulada. Blog Twitter, [s. l.], 4 fev. 2020a.

TWITTER. **Intensificando nosso trabalho para proteger a conversa pública em torno do #coronavirus**. Blog Twitter, [s. l.], 6 mar. 2020b.

TWITTER. **Política de mídia sintética e manipulada**. Twitter: Central de Ajuda, [s. l.], 2021a.

TWITTER. **Sobre as exceções devido ao interesse público no Twitter**. Twitter: Central de Ajuda, [s. l.], 2021b.

TWITTER. **Política de informações enganosas sobre a COVID-19**. Twitter: Central de Ajuda, [s. l.], 2021c.

TWITTER. **Regras e políticas**. Twitter: Central de Ajuda, [s. l.], 2021d.

UE. **A UE em 2020**: Relatório Geral sobre a Atividade da União Europeia. 2020. Disponível em: <https://op.europa.eu/webpub/com/general-report-2020/pt/>. Acesso em: 13 jul 2021.

VALENTE, J. **Tecnologia, informação e poder**: das plataformas online aos monopólios digitais. 2019. Tese (Doutorado em Sociologia) – Instituto de Ciências Sociais, Universidade de Brasília, Brasília, DF, 2019.

WARDLE, C.; DERAKHSHAN, H. **Information Disorder**: toward an interdisciplinary framework for research and policy making. Strasbourg: Council of Europe, 2017.

YIN, R. K. **Estudo de caso**: planejamento e métodos. 3. ed. Porto Alegre: Bookman, 2005.

Disinformation on platforms: actions taken by Twitter during a Covid-19 pandemic

Abstract: In this article, we propose a critical reflection in the light of platform capitalism on the actions taken by Twitter to combat misinformation during the Covid-19 pandemic. For this, we contextualized the phenomenon of disinformation and the pandemic, and carried out a case study based on documentary research of the platform's activities. We emphasize that the platform's coping strategy is based on two main perspectives: stopping misleading information through content moderation and providing secure information. Finally, we tensioned the proposals, pointing out the danger of an algorithmic solutionism and the limitation of some platform actions.

Keywords: Disinformation; Twitter; Pandemic; Covid-19; Media literacy

Recebido: 19/07/2021

Aceito: 10/11/2021

Declaração de autoria

Concepção e elaboração do estudo: Luciana Miranda Costa, Lizete Barbosa da Nóbrega e Carolina Toscano Maia.

Coleta de dados: Luciana Miranda Costa, Lizete Barbosa da Nóbrega e Carolina Toscano Maia.

Análise e interpretação de dados: Luciana Miranda Costa, Lizete Barbosa da Nóbrega e Carolina Toscano Maia.

Redação: Luciana Miranda Costa, Lizete Barbosa da Nóbrega e Carolina Toscano Maia.

Revisão crítica do manuscrito: Luciana Miranda Costa.

Como citar:

COSTA, Luciana Miranda; NÓBREGA, Lizete Barbosa da; MAIA, Carolina Toscano. Desinformação nas plataformas: ações de combate adotadas pelo Twitter durante a pandemia da Covid-19. **Em Questão**, Porto Alegre, v. 28, n. 3, e-116919, jul./set. 2022. <https://doi.org/10.19132/1808-5245283.116919>



¹ Original: Dis-information. Information that is false and deliberately created to harm a person, social group, organization or country (WARDLE; DERAKHSHAN, 2017, p. 20).

² Empresas de tecnologia que dominam o mercado atualmente. Morozov (2018) conceitua como: “[...] as grandes empresas associadas a plataformas de uso intensivo de dados, quase todas situadas na América do Norte, e também cada vez mais na China [...]” (MOROZOV, 2018, p. 144).

³ Original: “The model’s score predicts how interesting and engaging a Tweet would be specifically to you” (KOUMCHATZKY; ANDRYEYEV, 2017).

⁴ Original: “choosing the right approach to building and running prediction models can have profound consequences on the experience of every individual using Twitter” (KOUMCHATZKY; ANDRYEYEV, 2017).

⁵ Original: “Without the necessary digital media literacy skills, people frequently fall victim to dubious claims they encounter in this context” (GUESS *et al.*, 2020, p. 15536).

⁶ Original: “They may prove to be a vital and cost-effective way to reduce people’s vulnerability to false news” (Guess *et al.*, 2020, p. 15542).