

*Essa é uma tradução para o português da nossa contribuição à consulta sobre desinformação aberta pela Relatora Especial da ONU sobre a promoção e proteção do direito à liberdade de opinião e expressão que foi originalmente escrita em inglês e pode ser acessada no site do ITS.

Desinformação e liberdade de opinião e expressão

Submissão do Instituto de Tecnologia e Sociedade à Relatora Especial da ONU sobre a promoção e proteção do direito à liberdade de opinião e expressão

A missão do Instituto de Tecnologia e Sociedade (ITS) é assegurar que o Brasil e o Sul Global respondam de maneira criativa e apropriada às oportunidades fornecidas pela tecnologia na era digital, e que seus potenciais benefícios sejam amplamente compartilhados pela sociedade.

Por meio de pesquisa e de parcerias com outras instituições, o ITS Rio analisa as dimensões legais, sociais, econômicas e culturais da tecnologia e promove melhores práticas de regulação que protejam a privacidade, a liberdade de expressão e o acesso ao conhecimento. Agradecemos esta oportunidade de contribuir para a consulta da Relatora Especial sobre desinformação.

Desafios globais da desinformação sob a perspectiva brasileira

[A desinformação não é um fenômeno novo](#), mas na era digital assumiu proporções sem precedentes. A internet diminui as barreiras para a circulação de informações e não conhece fronteiras. Conteúdo falso e enganoso pode circular de maneira fácil e rápida em todo o mundo. O Sul Global pode ser particularmente afetado e as instituições democráticas estão especialmente em risco.

O Brasil é um desses casos. Ressalte-se que é o [segundo país](#) em termos de tempo gasto nas redes sociais (média de 3,45 horas em comparação a menos de 2 horas na maioria dos países desenvolvidos) e estas são uma fonte [importante de notícias para a maioria dos brasileiros](#). As eleições de 2018 são um exemplo significativo do impacto que a desinformação pode ter no sistema democrático. Os candidatos se apoiaram muito em serviços digitais, com destaque para as mídias sociais. Pesquisadores divulgaram em um [artigo do NY Times](#) que no mês anterior à eleição apenas 4 das 50 mensagens mais compartilhadas no WhatsApp puderam ser consideradas totalmente verdadeiras. Em geral, a eleição foi considerada permeada de notícias falsas e/ou fora de contexto.

Além disso, a ampla e rápida disseminação de desinformação geralmente é resultado de comportamento coordenado e do uso de ferramentas de automação para divulgar digitalmente notícias falsas e políticas no país, como mostra [este estudo](#) do ITS.

A pandemia é outro fator que impactou o ambiente online. Foram relatados vários casos de informações falsas e enganosas sobre a atual crise de saúde que circularam no país. Em

um [estudo](#) publicado em novembro de 2020, o Brasil se destaca como o único em que drogas como cloroquina e ivermectina ainda perduram no debate público, mesmo depois de nenhuma pesquisa científica sustentar possuírem qualquer efeito positivo significativo contra a doença causada pelo vírus da COVID-19.

O mesmo estudo indica que a desinformação em torno da pandemia vem sendo usada em um contexto de disputas de poder nos diferentes níveis de governo. Foram relatados episódios em que apoiadores desacreditaram a segurança das vacinas produzidas no país sob a égide de governadores opositores.¹

A contribuição está estruturada em três seções que apresentam um panorama dos esforços empreendidos (i) pelo governo, (ii) pelas empresas de tecnologia e (iii) pela sociedade civil para conter o problema no Brasil. Serão destacadas oportunidades e riscos para implementar e garantir os direitos humanos - particularmente liberdade de expressão, acesso à informação e privacidade - no Sul Global.

1. Esforços do governo brasileiro no combate à desinformação

A desinformação foi percebida como um grande desafio por todos os níveis de governo no Brasil. Isso propiciou um cenário em que tanto o Legislativo quanto o Judiciário atuaram no enfrentamento das causas e efeitos do fenômeno.

Entre as iniciativas, destaca-se uma série de projetos de lei propostos no Congresso para discutir como lidar com o assunto por meio de várias lentes, desde a criminalização de condutas até a proposição de diversas obrigações aos provedores de serviços de internet (principalmente empresas intermediárias). O Congresso também estabeleceu uma Comissão Parlamentar de Inquérito bicameral para investigar uma alegada "rede de desinformação" financiada com fundos públicos.

O Judiciário também foi conclamado a lidar com inúmeras reclamações de "notícias falsas" em contexto de difamação, ataques à honra pessoal, intimidação, *bullying*, entre outros. As reclamações ocorreram tanto em processos civis como criminais. O Supremo Tribunal Federal também abriu um inquérito polêmico sob a alegação de que tanto a própria instituição (o Tribunal) foi alvo de uma campanha de desinformação quanto seus membros. Como o procedimento foi iniciado *proprio motu* - pelos próprios Ministros, houve denúncias de violação do devido processo legal e parcialidade, que [teriam sido encaminhadas à Comissão Interamericana de Direitos Humanos](#).

2. Iniciativas Legislativas

2.1. Cenário Legal Brasileiro

¹ Foi o caso, por exemplo, em que o presidente da Fundação Palmares (fundação federal brasileira) escreveu no [Twitter](#): 'Sou paulista e apelo aos meus familiares, a quem desejo tudo de bom no mundo: não tomem a vacina chinesa de Doria!'. O comunicado até 10 de fevereiro (data da redação deste documento) teve quase 20.000 curtidas e foi retuitado mais de 3,6 mil vezes e ainda estava disponível na plataforma de mídia social. Segundo o site de notícias do [Yahoo](#), a seguinte mensagem de Roberto Jefferson, político brasileiro próximo de Bolsonaro, foi replicada mais de 4.000 vezes no Twitter: "Laboratório chinês criando vacina contra o vírus chinês e com pesquisa patrocinada por um governador que é grande parceiro da China ? Eu não quero essa vacina, e você? ".

O Brasil não tem uma definição legal de desinformação, notícias falsas ou desordem informacional. Muitas vezes, o Direito Penal - calúnia, difamação, perturbação do processo eleitoral, falsas comunicações de crimes, denúncias caluniosas e outras infrações - é utilizado como forma de lidar com o fenômeno. Isso, entretanto, levanta uma série de dificuldades, uma vez que tais instrumentos não são adequados para acompanhar o ritmo, o volume, a velocidade e o impacto da desinformação online e podem ir contra os direitos humanos protegidos pela Constituição e pelas obrigações internacionais.

Vários projetos de lei foram apresentados no Congresso para mudar esse cenário. Suas estratégias referem-se a: (i) revisar a responsabilidade dos intermediários de internet - deve-se observar que, de acordo com a legislação brasileira em vigor (Marco Civil da Internet), os provedores de aplicações de internet não são responsáveis pelo conteúdo gerado por terceiros, a menos que não cumpram uma decisão judicial de remoção de conteúdo; (ii) aumentar as obrigações dos provedores de aplicações de internet para coletar e armazenar dados (apesar das obrigações de minimização da proteção de dados); (iii) limitação da criptografia; (iv) adicionar ao código penal e outras legislações criminais tipos penais ligados à desordem de informação; e (v) aumentar os poderes de investigação.

Nas seções seguintes serão destacados os aspectos positivos e negativos referentes a tais propostas. A relevância desse exercício é mostrar a infinidade de opções que podem impactar o gozo dos direitos humanos, especialmente a liberdade de expressão no Brasil e no mundo. Muitas das propostas promovidas no país ecoam outras iniciativas internacionais.

2.2 Marco Civil da Internet - o *juste milieu* pode ceder sob pressão

Uma das formas de lidar com a desinformação online é por meio da regulamentação da moderação de conteúdo e de mecanismos de responsabilização das plataformas pelos conteúdos disponibilizados em seus serviços. Nesse sentido, o marco regulatório brasileiro é considerado um dos mais equilibrados e uma grande conquista. O [Marco Civil da Internet](#) (Lei nº 12.965 / 2014) é o resultado de um longo processo colaborativo com amplo suporte multissetorial. A Lei estabelece os princípios orientadores para a governança da Internet no país, que incluem liberdade de expressão, acesso à informação e privacidade.

O Art. 19 da referida Lei criou um regime de responsabilidade para os provedores de aplicações de internet - mídias sociais, sites de vídeo, enciclopédias, aplicativos de mensagens e qualquer plataforma (incluindo notícias) que contenham comentários e contribuições de seus usuários - conhecido como **notificação judicial e retirada**.

De acordo com o Art. 19, as empresas de tecnologia não são responsáveis pelos atos de terceiros (seus usuários) até que uma ordem judicial declare que o conteúdo é ilegal. A partir desse momento, se não retirarem o conteúdo (texto, fotos ou vídeo) tornam-se responsáveis por ele. A exceção se deve a conteúdo que justifique mecanismos de remoção rápida, como nudez ou pornografia não autorizada (“pornografia de vingança”), que são regulamentados pelo Art. 21.

Por outro lado, as plataformas podem ser responsáveis por suas próprias ações - remoção ou redução da disponibilidade de conteúdo. Vários usuários, sentindo-se prejudicados com

a remoção de páginas, vídeos ou fotos, já processaram fornecedores na justiça brasileira. Observe-se que no Brasil não há previsão de imunidade para ações de moderação de conteúdo, como na Seção 230 nos Estados Unidos.

Assim, evidencia-se um compromisso alcançado no âmbito do Marco Civil da Internet, um *juste milieu*, onde por princípio os indivíduos têm liberdade de expressão online e as plataformas são incentivadas a moderar o conteúdo de forma responsável, sabendo que suas ações devem respeitar os direitos individuais. Cabe aos Tribunais - e não às empresas de tecnologia - decidir definitivamente se determinado conteúdo é legal ou ilegal, responsabilizando as prestadoras de serviço apenas se não cumprirem uma ordem judicial.

No entanto, há um sentimento de que esse regime pode não ser eficaz, especialmente contra as campanhas de desinformação. Assim, diversas iniciativas propõem emendar o Marco Civil da Internet com diversos riscos para o gozo dos direitos humanos no país.

2.3 Projetos de Lei sobre ‘Desinformação’

Hoje, estão em análise mais de 50 propostas sobre desinformação que, em síntese, iriam:

- abordar a criminalização da disseminação ou compartilhamento de informações falsas ou incompletas na Internet,
- responsabilizar as mídias sociais quando materiais falsos são publicados na internet e o conteúdo não é removido dentro de vinte e quatro horas,
- forçar as plataformas de mídia social a fornecer filtros e ferramentas para evitar a disseminação de informações prejudiciais, e
- responsabilizar os fornecedores de conteúdos e de serviços de Internet pelos danos causados pela divulgação de notícias falsas na Internet.

- **Projeto de lei nº 2630/2020 - Lei das ‘Fake News’**

Um dos projetos de lei que mais reverberou nos últimos tempos e foi aprovado em uma das casas legislativas foi o Projeto de Lei nº 2630/2020, também conhecido como Lei das ‘Fake News’. O processamento do projeto envolveu uma série de problemas, em grande parte devido ao seu ritmo acelerado, o que não permitiu uma participação social ideal. Este é um grande contraste com a história colaborativa do Marco Civil da Internet.

O projeto de lei passou por mudanças notáveis, com diferentes versões sendo publicadas a cada poucos dias. Em suas várias versões, o projeto de lei gerou diversas polêmicas sobre o controle de plataformas e usuários. Na versão aprovada pelo Senado Federal, algumas dessas críticas foram absorvidas e muitos dos dispositivos em questão foram retirados do texto. Muitas cláusulas polêmicas, no entanto, foram mantidas, como a obrigação de coletar e armazenar em massa informações sobre trocas de mensagens para possibilitar rastrear a origem e a disseminação de mensagens - desconsiderando questões de privacidade e contornando protocolos de criptografia.

Nesse cenário, o projeto de Lei das ‘Fake News’ acumula críticas nacionais e internacionais: desde a [Human Rights Watch](#), passando pela [Global Network Initiative](#), até o [Conselho Nacional de Direitos Humanos](#) e esta Relatora da ONU para a liberdade de expressão e

opinião. A [Coalizão Direitos na Rede](#), composta por 39 acadêmicos e ativistas de direitos civis e digitais no Brasil, incluindo o ITS, tem repetidamente apontado os riscos e fracassos do projeto, além da falta de apoio social e participação em sua construção.

O ITS destacou nesta [nota técnica](#) os problemas em apressar seu processamento e as imprecisões conceituais do texto. Entre as principais lições estão:

- **O texto pode encorajar a censura e prejudicar a liberdade de expressão na Internet.** Isso reverte um dos maiores avanços trazidos pelo Marco Civil da Internet do Brasil, pois o projeto pode levar as plataformas a monitorar todo o conteúdo publicado pelos usuários e suas intenções, visto que podem ser responsabilizadas pelas publicações de terceiros em suas plataformas. Assim, os provedores podem remover agressivamente conteúdo legal que tenha um leve indício desacordo com a Lei; inverter a presunção em favor da liberdade de expressão para contra; As plataformas serão direcionadas ao policiar o conteúdo e se tornarão, de fato, suas guardiãs. O "efeito inibidor" que esta obrigação pode ter sobre a liberdade de expressão não pode ser ignorado.
- **O texto traz definições imprecisas:** definições imprecisas ou mesmo conflitantes com outros textos legais podem gerar confusão e dar ensejo a abusos de poder e à atribuição de funções de políticas às plataformas, a fim de definir "desinformação" nos casos de violação da liberdade de expressão. Por exemplo, não há reservas quanto à desinformação, que às vezes pode ser resultado da ação de um usuário mal informado ou de um erro jornalístico. Esses casos de má informação não têm a intencionalidade esperada de desinformação e merecem tratamento diferenciado.
- Usuários suspeitos de criar bots ou contas anônimas podem ser obrigados a **apresentar IDs a empresas de tecnologia (identificação em massa), sem uma ordem judicial:** as contas podem ser relatadas como suspeitas por uma ampla variedade de razões - incluindo, em teoria, disputas políticas. Além disso, condicionar o registro nas redes sociais à apresentação de documentos em um país onde milhões de pessoas ainda não possuem prova legal de identidade pode exacerbar o já existente fosso digital.

- **Projeto de Lei nº 246/2021**

O [Projeto de Lei nº 246/2021](#) estabelece um regime de responsabilidade civil para os provedores de aplicações de internet nos casos de moderação de conteúdo, não apenas para remoção, mas também para rotulagem de conteúdo. O Projeto de Lei nº 246 parece transplantar uma discussão muito viva nos Estados Unidos a respeito da imunidade para moderação de conteúdo sob a Seção 230 do Communications Decency Act (conhecido como "[Bom Samaritano](#)"). Essa medida legal é, na melhor das hipóteses, supérflua, pois, como já foi mencionado, essa imunidade nunca existiu no Brasil. As ações da plataforma diretamente no conteúdo são "atos da plataforma" e, como tal, não estão cobertas pela imunidade do Art. 19 do Marco Civil da Internet no Brasil.

Se as empresas de internet agirem de forma abusiva e removerem ou mesmo rotularem o conteúdo de forma incorreta, elas podem ser responsabilizadas. O Brasil já possui diversos processos em que o autor de postagens ou o titular de contas removidas indevidamente

venceram seus processos no Judiciário. Os tribunais brasileiros já [condenaram o Google](#), por exemplo, por remover por engano paródias de músicas sob alegação de violação de direitos autorais. Um congressista já ganhou uma ação judicial contra o Facebook porque a empresa removeu sua conta alegando disseminação de notícias falsas e [a empresa foi obrigada](#) a republicar as postagens. Similarmente, o [Facebook já foi condenado](#) por remover por engano a fanpage de um deputado. Ou seja, a responsabilidade das empresas de tecnologia pela atividade de moderação já ocorre sem a necessidade de uma nova lei.

O perigo para os direitos humanos pode não estar nas obrigações presentes no projeto de lei, mas no ambiente que ele cria. Ele parece encorajar um ambiente contra a liberdade de expressão online, concentrando-se em remodelar o equilíbrio estabelecido no Marco Civil da Internet e forçando as empresas a protegerem o discurso que pode ser discriminatório, ofensivo ou até mesmo incitador da violência.

- **Projeto de Lei nº 291/2021**

O [Projeto de Lei nº 291](#) foi apresentado no início de fevereiro deste ano. O Projeto visa proibir a remoção de mensagens de usuários por um provedor de aplicação em desacordo com as garantias constitucionais de liberdade de expressão, comunicação e expressão de pensamento. No entanto, deve-se observar que, como está, o projeto de lei pode ter um alcance limitado.

O Projeto de Lei acrescenta ao Marco Civil da Internet prevendo a nulidade de quaisquer cláusulas contratuais e de termos de uso relativos à prestação de serviços de aplicações de Internet que prevejam a retirada de conteúdo do usuário por orientação política ou expressão de opinião. Hoje, em tese, não existe um termo de uso que permita a supressão de conteúdo por "orientação política ou expressão de opinião".

Este Projeto de Lei, se aprovado, também pode ter um "efeito inibidor" nas empresas de tecnologia que usam seus termos de uso como justificativa para remover conteúdo, pois podem estar sujeitos a penalidades elevadas. A natureza vaga dos termos usados no Projeto de Lei proposto aumenta esse risco.

2.4 Comissão Parlamentar de Inquérito das 'Fake News' ('Fake News CMPI')

A Comissão Parlamentar de Inquérito sobre 'notícias falsas' ('Fake News CPMI') está intimamente associada ao aumento de projetos de lei sobre desinformação. A Comissão foi motivada por denúncias de ["ataques cibernéticos que atentam contra a democracia e o debate público; a utilização de perfis falsos para influenciar os resultados das eleições 2018; a prática de cyberbullying sobre os usuários mais vulneráveis da rede de computadores, bem como sobre agentes públicos; e o aliciamento e orientação de crianças para o cometimento de crimes de ódio e suicídio"](#).

Em dezembro de 2019, a deputada Joice Hasselmann testemunhou expondo um grupo chamado ["gabinete do ódio"](#) dentro do governo do presidente Jair Bolsonaro. [Segundo Hasselmann](#), já se comprovou que quase R\$ 500 mil foram gastos na perseguição de adversários políticos, por meio de reportagens diárias globais e nacionais, além de análises de sentimento, que servem para orientar a tomada de decisões, além do amplo uso de bots

para espalhar desinformação online. 'Sugiro seguir a trilha do dinheiro, pois estamos falando de milhões', argumentou Hasselmann.

Como restou claro, uma das principais contribuições da CPMI foi a identificação de práticas de financiamento para campanhas de desinformação, o que motiva os projetos a primeiro identificar a estratégia "siga o dinheiro" no combate à indústria de notícias falsas. A CPMI das Fake News foi prorrogada por mais 180 dias e incluiria a disseminação de desinformação sobre a pandemia COVID-19, uma vez que, à medida que o coronavírus se espalhava pelo Brasil, [histórias sinistras começaram a surgir nas mídias sociais](#), muitas vezes via parlamentares com milhares de seguidores. Os riscos de disseminação da desinformação ganharam nova urgência devido à gravidade do problema de saúde.

Um dos pontos positivos do trabalho desenvolvido pela Comissão foi o fato de ter sido ouvida uma [vasta gama de atores](#). Jornalistas, especialistas no combate à desinformação e empresas de tecnologia prestaram depoimento. [Segundo a Relatora da Comissão](#), essas contribuições servirão de base ao legislativo para "estabelecer uma legislação que possa punir culpados sem ferir a liberdade de expressão".

3. Iniciativas judiciais - "ordens de bloqueio"

Em março de 2019, o Ministro Dias Toffoli, então Presidente do Supremo Tribunal Federal, ordenou a [abertura de inquérito](#) que passou a ser conhecido como 'Inquérito das Fake News' ([Inquérito nº 4781](#)) para apurar a suposta existência de calúnia, ameaças e denúncias fraudulentas contra os Tribunal, seus membros e parentes.²

O Ministro Alexandre de Moraes, relator do inquérito, proferiu diversas decisões polêmicas durante o processo. O magistrado ordenou que as redes sociais Facebook e Twitter bloqueassem o acesso às contas de 16 indivíduos investigados por supostamente espalharem desinformação e discurso de ódio online. No entanto, após o bloqueio nacional, os usuários fora do Brasil, ou que usam VPN, continuaram tendo acesso às referidas contas. O Ministro teria, então, pedido a [suspensão global dos perfis](#) depois de perceber que o bloqueio que ordenou anteriormente teve impacto reduzido. O evento teve uma repercussão ainda maior, pois um dos 16 indivíduos que teve suas contas bloqueadas globalmente pôde usar outra conta para fazer uma declaração bastante polêmica.

Especialistas e acadêmicos criticaram a polêmica decisão, que a princípio [foi contestada pelo Facebook e Twitter](#). A empresa ressaltou que qualquer pedido com efeito global deve ser tratado com cautela para não impactar os direitos das pessoas de outros países. Posteriormente, [multas mais altas e o risco de responsabilidade criminal de um de seus funcionários](#) levaram as empresas a cumprir a determinação, enquanto pendia recurso. [Informou-se](#) que o procedimento foi levado à consideração da Corte Interamericana de Direitos Humanos sob a alegação de violações dos direitos dos investigados, incluindo liberdade de expressão, informação e imprensa.

Solicita-se à Relatora que se debruce sobre a extensão adequada da ordem judicial. Levando em consideração que os países diferem sobre como implementar certos direitos -

² This measure raised controversy regarding the legality of the processing, as it was initiated ex officio, that is, without request of another body, such as the Public Ministry, the Attorney General's Office or a police authority.

como a liberdade de expressão - e como equilibrá-los entre si, é de se esperar que as **ordens globais que podem impactar os direitos humanos sejam uma exceção e apenas aplicáveis em situações de consenso global ou para proteger *jus cogens***.

4. Iniciativas de empresas de tecnologia

4.1 Apresentando o atrito aos ecossistemas de informação

Uma maneira promissora de conter a desinformação nas plataformas de mídia social é introduzir o atrito nos ecossistemas de informação que ajudaram a criar e agora habitam. Esta prática é relativamente nova - ganhou força durante a eleição presidencial de 2020 nos Estados Unidos - mas seus impactos têm sido principalmente positivos por duas razões: (1) ao diminuir a velocidade com que as informações são compartilhadas nas redes sociais, medidas que induzem a fricção propiciam mais tempo para que usuários, verificadores de fatos e plataformas atuem e enfrentem as campanhas de desinformação e (2) ao contar com uma abordagem de direcionamento comportamental (do inglês "nudge"), essas medidas têm um impacto menor (e às vezes inexistente) na liberdade de expressão.

As iniciativas que induzem a fricção são basicamente pequenos ajustes de design introduzidos na plataforma com o objetivo de tornar as informações sinalizadas como potencialmente falsas ou enganosas menos proeminentes e mais difíceis de encontrar ou compartilhar. Como observou o [New York Times](#) em um artigo que descreve como o Facebook e o Twitter se comportaram no dia das eleições nos Estados Unidos, 'é um sinal revelador de autoconsciência, como se a Ferrari tivesse percebido que só poderia impedir que seus carros batessem substituindo os motores com motores de kart'. Desde o seu surgimento nos anos 2000 e no início dos anos 2010, as plataformas digitais fomentaram espaços onde as informações podiam ser compartilhadas sem problemas e as investigações jornalísticas sérias do Times apareciam lado a lado com artigos falsos do jornal Denver Guardian, semanas mais antigos.

Outra maneira de ver isso é analisando a evolução do controverso lema do Facebook: "mova-se rápido e quebre as coisas". O problema é que, ao se mover muito rápido e não perceber as desvantagens associadas à velocidade com que as informações estavam sendo compartilhadas em sua plataforma sem atrito, o Facebook chegou muito perto de quebrar uma coisa muito importante: a verdade. Posteriormente, a empresa [atualizou seu lema](#) para "mover-se rapidamente com infraestrutura estável". Este é outro sinal revelador de autoconsciência. A empresa agora entende que, para se mover com rapidez e segurança, sua infraestrutura deve ser construída sobre uma base sólida, capaz de lidar com as desvantagens das informações de alta velocidade de antemão. Chega de 'se mover rápido, quebrar coisas e consertar mais tarde'.

Nesta contribuição, gostaríamos de destacar algumas dessas medidas indutoras de fricção. Observe-se que não temos como objetivo uma lista abrangente. O primeiro exemplo digno de nota vem do Twitter, quando a empresa apresentou um prompt 'leia antes de retuitar' para seus usuários em setembro de 2020. A ideia é provocar a mudança de comportamento inconsciente, e incitar as pessoas a lerem os artigos que pretendem compartilhar antes de clicarem em 'retuitar.' Como a empresa [observou](#) em um tweet próprio, "compartilhar um artigo pode desencadear uma conversa, então você pode querer lê-lo antes de twittar". Se um usuário tentar compartilhar um artigo que não leu, o Twitter mostrará uma tela dizendo

que 'as manchetes não contam a história completa' e que o usuário 'pode ler o artigo no Twitter antes de retuitar'.

Esta é uma inovação importante porque manchetes sensacionalistas são frequentemente usadas por usuários mal intencionados para espalhar informações falsas ou enganosas nas mídias sociais. De acordo com o Washington Post, um estudo de 2016 realizado por cientistas da Universidade de Columbia e do Instituto Nacional Francês descobriu que 59% dos links compartilhados nas redes sociais nunca foram realmente clicados: em outras palavras, a maioria das pessoas parece retuitar notícias sem nunca lê-las'. Ao introduzir o atrito na função de retuíte, o Twitter deu um grande passo para corrigir o problema. De acordo com as [próprias métricas da empresa](#), 'as pessoas abrem artigos com 40% mais frequência depois de ver o prompt' e 'algumas pessoas não terminam fazendo RTing após abrir o artigo', o que aponta para uma conversa mais informada sobre os artigos e diminui as chances que as pessoas acabarão divulgando manchetes falsas ou enganosas.

Outra inovação interessante também foi introduzida pelo Twitter durante a eleição presidencial nos Estados Unidos. A empresa decidiu mudar temporariamente a função de retuíte para incitar as pessoas a adicionarem suas próprias "citações" antes de compartilhar um pedaço de conteúdo. Como observou o [The Verge](#), "o Twitter espera que, ao introduzir algum atrito no processo, as pessoas possam considerar melhor exatamente o que estão retuitando ou aproveitar a oportunidade para adicionar sua própria perspectiva". Embora levar os usuários a adicionar suas perspectivas únicas ao debate promova um debate mais informado e refreie a desinformação, o Twitter [reverteu as mudanças](#) depois de notar que 45% dos Tweets com citações eram "afirmações de uma única palavra e 70% tinham menos de 25 caracteres". Resta saber se um 'Quote Tweet Prompt 2.0' pode aprender com essa experiência e induzir as pessoas de forma eficiente a entrar na conversa de uma forma mais significativa.

O Facebook também experimentou uma série de medidas temporárias de indução de atrito durante as eleições de 2020 nos Estados Unidos. Em um resumo das mudanças mais importantes, o [New York Times](#) observou que a empresa "implementou um novo e complicado processo de aprovação para anunciantes políticos e bloqueou novos anúncios políticos no período após o dia das eleições. Ele conteve alegações falsas e colocou em prática um "disjuntor de viralidade" para dar aos verificadores de fatos tempo para avaliar histórias suspeitas. E desligou temporariamente seu algoritmo de recomendação para certos tipos de grupos privados, para diminuir a possibilidade de agitação violenta ". Em outras palavras, o Facebook substituiu temporariamente o motor de sua Ferrari para permitir que verificadores de fatos e engenheiros monitorassem o veículo e, se necessário, mudassem de curso.

Além disso, o Facebook também introduziu o atrito em sua função de compartilhamento para ajudar a combater a desinformação COVID-19. A empresa chama a abordagem de 'Compartilhamento Informativo' e consiste em oferecer às pessoas mais informações básicas sobre um determinado conteúdo antes de clicarem em 'Compartilhar'. De acordo com o OneZero, "o Facebook adicionou telas de notificação fazendo as pessoas pausarem, apenas por um momento, e considerar algum contexto antes de compartilharem postagens relacionadas ao COVID-19 e também artigos desatualizados". Isso é consistente com a pesquisa do Professor David Rand do MIT sobre "[intervenção de deslocamento de precisão escalonável](#)". De acordo com Rand, incitar as pessoas a pensar sobre a exatidão de uma

informação sobre o COVID-19 antes de compartilhá-la reduz significativamente a probabilidade de repassarem informações falsas ou enganosas.

Do ponto de vista da legislação de direitos humanos, **ajustes de design que introduzam atrito aos ecossistemas de informação podem ser uma solução menos invasiva à liberdade de expressão**. Afinal, a maioria das iniciativas descritas acima são estímulos ou, em outras palavras, tentativas de empurrar o comportamento humano em uma determinada direção sem restringir diretamente as opções das pessoas ou tratá-las abertamente. Por exemplo, quando o Twitter mostra a seus usuários uma tela perguntando se eles gostariam de ler o artigo antes de compartilhar, a empresa não está restringindo os direitos do usuário, apenas o incitando a contribuir com uma conversa online mais informada, tornando-os mais conscientes do que estão compartilhando. Se o usuário decidir que não é do seu interesse realmente ler o conteúdo antes de retuitá-lo, pode simplesmente ignorar o prompt e prosseguir com a postagem.

Essa estratégia tem seus riscos. A decisão sobre o que "estimular", e em qual direção, pode impactar o conteúdo e afetar as escolhas das pessoas. **A transparência nas mudanças arquitetônicas é fundamental para ser compatível com os direitos humanos**.

4.2 Focar em perfis grandes e proeminentes

Concentrar-se em grandes contas tem se mostrado eficaz em conter a desinformação nas redes sociais. Essas empresas costumam enfatizar como é difícil monitorar o conteúdo na plataforma, identificar instâncias de desinformação que violem os padrões da comunidade e, finalmente, agir de acordo. No entanto, desenvolvimentos recentes no campo apontam para uma direção um pouco diferente. A experiência com a 'desplataformização' do ex-presidente dos EUA Donald J. Trump mostra que as plataformas podem percorrer um longo caminho na luta contra informações falsas ou enganosas, monitorando grandes contas que seriam responsáveis por uma parcela desproporcional de conteúdo falso. No entanto, ao contrário das medidas indutoras de fricção, **esta prática levanta sérias preocupações sobre as violações dos direitos humanos**, especialmente o direito à liberdade de expressão e o direito à igualdade.

Em uma pesquisa inovadora sobre fraude eleitoral pelo correio nos Estados Unidos, Yochai Benkler, professor de direito de Harvard, e seus co-autores, [descobriram](#) que 'ao contrário do foco da maioria dos trabalhos contemporâneos sobre desinformação, [a campanha de desinformação por fraude eleitoral] era fruto de um processo de mídia de massa conduzido por uma elite. A mídia social desempenhou apenas um papel secundário e de apoio.' De acordo com o estudo, Donald J. Trump manipulou o ecossistema de informações e transformou os meios de comunicação de massa em propagadores de sua elaborada mentira de fraude eleitoral. Trump conseguiu isso 'usando três práticas padrão básicas de jornalismo profissional [...]: foco institucional na elite (se o presidente diz isso, é notícia); busca de manchetes (se sangra, lidera); e equilíbrio, neutralidade ou evitar a aparência de tomar um lado.'

Os resultados deste estudo apontam para uma situação difícil, eles sugerem que o problema da desinformação da mídia de massa dirigida pela elite não pode ser resolvido por verificação de fatos ou ajustes de algoritmos por plataformas de mídia social. Embora os autores do estudo não se aventurem tão longe, uma conclusão plausível é que, para conter

efetivamente as campanhas de desinformação generalizadas, como as alegações de fraude eleitoral de Trump, as plataformas de mídia social devem encontrar maneiras de neutralizar a fonte. Embora por razões diferentes, Facebook, Twitter e outras plataformas suspenderam Trump por um período indefinido de tempo depois que o ex-presidente dos EUA incitou uma multidão violenta contra o Congresso em 6 de janeiro de 2021. Após 'a grande *deplataformização*', o [Washington Post](#) observou que vários institutos de pesquisa chegaram a uma conclusão semelhante: a desinformação online sobre a eleição caiu drasticamente.

Um estudo do Signal Labs, por exemplo, descobriu que a desinformação eleitoral caiu 75% depois que Trump perdeu o acesso às suas contas nas redes sociais. Como afirma o artigo do [Washington Post](#), 'a pesquisa do Signal e de outros grupos sugerem que um ecossistema de desinformação integrado e poderoso - composto de influenciadores de alto perfil, seguidores comuns e o próprio Trump - foi fundamental para levar milhões de americanos a rejeitar os resultados das eleições e pode ter problemas para sobreviver sem suas contas nas redes sociais'. Para ter certeza, isso levanta uma questão importante: até onde as plataformas de mídia social devem ir para conter as campanhas de desinformação? Parece claro que banir Trump para sempre é uma medida eficaz contra informações falsas ou enganosas sobre fraude eleitoral, mas é proporcional em relação ao direito à liberdade de expressão? Embora ainda não tenhamos uma resposta, acreditamos que o relatório temático anual da Relatora Especial deve reconhecer essa tensão.

Além disso, o foco em grandes contas também **levanta uma questão sobre igualdade**. Por um lado, o ponto mais óbvio é que as plataformas de mídia social precisariam aplicar seus padrões de comunidade de uma forma não coerente, restringindo a fala de figuras proeminentes na plataforma enquanto contas menores podem ser admitidas. Mas, por outro lado, o banimento de Donald J. Trump também mostra como as plataformas podem estar dispostas a agir nos EUA - onde, se falharem, sua reputação é mais vulnerável - mas não necessariamente em outros países onde podem ter menos a perder com a manutenção de contas. Embora outros líderes mundiais tenham sido acusados de usar contas de mídias sociais para semear discórdia e até mesmo incitar a violência, nem todos tiveram seus 'megafones digitais' restringidos. Por exemplo, como Tom Phillips observou em seu artigo para o [The Guardian](#), 'os apelos à tomada de providências têm sido particularmente abundantes no Brasil, que é liderado desde 2019 por Jair Bolsonaro, um tweeter-líder de extrema direita que se deleita em ser retratado como o "Trump tropical". No entanto, apesar das acusações de usar repetidamente as redes sociais para minar a democracia, incitar a violência e [espalhar desinformação](#), as contas do presidente brasileiro nas redes sociais continuam ativas.'

Há uma tensão importante na defesa da liberdade de expressão e em encontrar maneiras de restringir contas de alcance significativo acusadas de estarem ou que estejam de fato espalhando desinformação.

5. Iniciativas da sociedade civil

As organizações da sociedade civil têm um papel importante na redução da desinformação e na promoção dos direitos humanos. As iniciativas de educação midiática são de importância significativa nos esforços para lidar com a desinformação, especialmente quando reúnem diferentes partes interessadas.

5.1 Programa de Enfrentamento à Desinformação do Tribunal Superior Eleitoral do Brasil

Uma iniciativa promissora foi lançada pelo Tribunal Superior Eleitoral do Brasil (TSE). O [Programa de Enfrentamento à Desinformação](#) do TSE com vistas às eleições brasileiras de 2020 adotou uma abordagem multissetorial. O Programa contou com o apoio de [57 instituições](#), entre partidos políticos e entidades públicas e privadas.

O programa foi organizado sob diferentes estratégias, sendo de notável relevância: (i) Alfabetização midiática e informacional; (ii) Contenção de desinformação; (iii) Verificação de Identificação e Desinformação; (iv) Aperfeiçoamento da Ordem Legal; e (v) Melhoria dos Recursos Tecnológicos. A descrição completa do programa e seu plano estratégico podem ser encontrados em português [aqui](#).

Uma coalizão de verificadores de fatos foi formada para as eleições. A existência do grupo de nove verificadores foi publicada na página '[Fato ou Boato](#)', disponível no Portal da Justiça Eleitoral.

Somado a isso, o programa fez parceria com quatro das maiores plataformas de mídia social e mensageria do mundo - Google, Facebook, Twitter e WhatsApp. Por exemplo, no caso do [Google](#), os eleitores que buscavam informações sobre as eleições encontraram no topo dos resultados da pesquisa um conteúdo confiável e de utilidade pública elaborado pela Justiça Eleitoral.

Durante o período eleitoral, de 27 de setembro a 29 de novembro, foi criada uma plataforma para receber denúncias de contas suspeitas de envio de grandes quantidades de mensagens de desinformação. Recebeu-se 5.180 denúncias, 199 das quais foram indeferidas por não terem relação com as eleições. Após esse primeiro filtro, a Justiça Eleitoral encaminhou 4.981 denúncias ao WhatsApp, para verificar possíveis violações dos Termos de Serviço do App.

Após uma análise preliminar para remover números duplicados ou inválidos, o WhatsApp identificou 3.527 contas válidas e banuiu 1.042 números (29,5%) por violarem seus Termos de Serviço. Do total de contas banidas, mais de 64% foram bloqueadas de forma proativa e automática pelo sistema de integridade do WhatsApp, antes mesmo de serem denunciadas.

A criação do “Despacho Eleitoral no WhatsApp”, chatbot desenvolvido gratuitamente pela Infobip para facilitar o acesso a dados confiáveis sobre as eleições, também foi importante para a Justiça Eleitoral divulgar informações relevantes aos eleitores brasileiros. Durante o período eleitoral, este canal teve mais de um milhão de usuários únicos e mais de 18 milhões de mensagens foram trocadas com o robô dentro do aplicativo.

Essas parcerias multissetoriais podem ter um impacto na proteção dos direitos, pois aceleram as ações e oferecem uma oportunidade para que diferentes atores façam sua parte e sejam ouvidos.

5.2. Iniciativas de educação midiática da sociedade civil

Conforme destacado pela Comissão [Interamericana de Direitos Humanos](#), há uma 'necessidade urgente de promover programas de alfabetização digital voltados para o desenvolvimento de habilidades cívicas em uma perspectiva de coexistência democrática e com um enfoque de direitos humanos.' As iniciativas da sociedade civil podem endereçar essas lacunas. Várias delas focam em diferentes aspectos do fenômeno da desinformação.

'[Expondo a desinformação de robôs no Brasil](#)', por exemplo, é um esforço para garantir que as pessoas estejam cientes do uso de robôs nas redes sociais, especialmente aqueles que espalham desinformação. Por meio de um algoritmo que promove a transparência sobre o comportamento de bot nas mídias sociais, o projeto fortalece a capacidade de usuários da Internet, organizações da sociedade civil, organizações de mídia e formuladores de políticas para identificar e contextualizar as campanhas de desinformação. Até agora, o algoritmo do projeto ('Pegabot', bot catcher em inglês) foi usado mais de 100 mil vezes para verificar a probabilidade de perfis do Twitter serem automatizados. Além disso, 750 membros do sistema de justiça eleitoral participaram de treinamentos sobre desinformação e automação para que pudessem incorporar esse conhecimento no trato dos processos eleitorais afetados pela desinformação.

Assim, iniciativas de terceiros baseadas em educação midiática e transparência podem revelar-se maneiras importantes de enfrentar os desafios colocados pela desinformação. Elas não apenas podem aumentar a conscientização sobre o problema, mas também criar resiliência entre as pessoas afetadas.

5.3. O papel desempenhado por 'influenciadores digitais'

'Influenciadores digitais'³ são fundamentais para o ecossistema de informações de hoje. Nas mídias sociais, eles são importantes produtores e distribuidores de informações. Espontaneamente, os influenciadores movem tendências e temas, estruturam debates. No Sul Global, onde a maioria da população consome suas notícias por meio das redes sociais, os influenciadores são uma das fontes mais importantes.

As iniciativas destinadas a aumentar a conscientização sobre o papel dos influenciadores digitais na divulgação de informações podem melhorar o ambiente geral de informações.

Projetos como o '[Redes Cordiais](#)', que treina comunicadores de diferentes nichos para combater a desinformação e o discurso de ódio nas mídias sociais, ajudam a abrir um caminho para o futuro. Não restringem direitos, pelo contrário, empoderam não só os influenciadores, mas também o público em um diálogo sobre como combater a desinformação. Como exemplo, o Redes Cordiais reuniu em 16 meses 109 influenciadores - que no total têm mais de 68,3 milhões de seguidores - para treinar e discutir como identificar e traduzir para seus públicos o que é desinformação e discurso de ódio online. Assim, os influenciadores tornam-se propagadores responsáveis com o objetivo de encontrar e disseminar informações confiáveis sempre que discutem questões com seus públicos.

Portanto, a Relatora Especial pode reconhecer que as iniciativas destinadas a aumentar a conscientização sobre o papel dos influenciadores digitais na divulgação

³ Em geral, "influenciador digital" é entendido como um novo tipo de celebridade na Internet, uma pessoa que acumula muitos seguidores e é capaz de modificar a opinião ou comportamento de seu público.

de informações ajudam na luta contra a desinformação, apoiando a divulgação de informações confiáveis e aumentando o nível de alfabetização midiática da população.

Conclusão

Uma crise global de desinformação persiste. É particularmente desafiador navegar efetivamente pelo problema e ao mesmo tempo proteger a liberdade de expressão e o acesso à informação. A regulamentação e as salvaguardas podem desempenhar um papel, mas, como as experiências até agora demonstram, elas trazem seus próprios desafios e podem não ser suficientes.

A solução deve ir além do governo e além da imposição de duras obrigações aos intermediários e provedores de serviço na Internet - alterando os regimes de responsabilidade. Os direitos em jogo são de extrema importância, constituem a base da democracia e não cabe ao setor privado regulá-los. O que é necessário é reconhecer que se trata de um problema multifacetado, que requer uma abordagem holística. Diferentes atores, sob a orientação do governo, devem se engajar em iniciativas multissetoriais, em que cada ator tem um papel importante a desempenhar.

Recomendamos que a Relatora Especial considere profundamente a complexidade do cenário e busque uma abordagem holística, reconhecendo a variedade de contribuições que podem ser feitas a fim de criar um ambiente melhor e mais propício à proteção dos direitos das pessoas.