ANÁLISE

DEMOCRACIA E DIREITOS HUMANOS

REGULAÇÃO DE COMBATE À DESINFORMAÇÃO

Estudo de oito casos internacionais e recomendações para uma abordagem democrática

João Brant, João Guilherme Bastos dos Santos
Tatiana Dourado e Marina Pita
Março de 2021

O uso estratégico de informações falsas - *fake news* - e enganosas, visando alterar o comportamento de nichos específicos da sociedade, é prática notada muito antes da popularização na Internet.

A análise dos casos revela os desafios intrínsecos ao processo de construir soluções regulatórias que sejam, ao mesmo tempo, protetoras de direitos, mas eficazes na promoção do acesso à informação confiável.

O fenômeno da desinformação está profundamente imbricado com processos políticos e sociais, e não é possível pensar soluções legais e regulatórias de forma isolada da compreensão desses processos e dos contextos nacionais e locais.

FRIEDRICH EBERT STIFTUNG

DEMOCRACIA E DIREITOS HUMANOS

REGULAÇÃO DE COMBATE À DESINFORMAÇÃO?

Estudo de oito casos internacionais e recomendações para uma abordagem democrática

Índice

1

INTRODUÇÃO

O uso estratégico de informações falsas e enganosas visando alterar o comportamento de nichos específicos da sociedade é prática notada muito antes da popularização na Internet. Do uso da propaganda política em guerras (Lasswell, 1938) à adoção de atalhos informativos, por onde eleitores preenchem lacunas de informação sobre política (Popkin, 1994), a difusão deliberada de rumores e falsificações se adequa aos meios de comunicação e às infraestruturas tecnológicas vigentes em cada momento histórico. Seja visando soldados dispostos a deixar o campo de batalha, seja um grupo demográfico mais disposto a se abster, rejeitar ou votar em um candidato se em contato com informações específicas, o efeito dessas campanhas depende da eficácia da identificação de nichos e do direcionamento/entrega de informações personalizadas a eles.

Na última década, o aumento da fragmentação e da opacidade dos ambientes informacionais organizados em plataformas digitais como Facebook e WhatsApp tem gerado o 'enterramento' de parte do debate público[1]. Essa característica dada pela profusão de grupos fechados impede que haja escrutínio público sobre parte das ideias em discussão e dificulta a visibilidade de perspectivas contraditórias.

Sem transparência, a confiabilidade da informação é diretamente fragilizada. Em arenas de discussão invisíveis ao público e sem responsabilização moral e legal dos interlocutores, informações enganosas e fraudulentas têm mais chance de prosperar. Com todos os limites do modelo de comunicação tradicional e sua dificuldade para prover um debate plural e diverso, a transparência do debate público e a busca da credibilidade pelos meios de comunicação ajudaram a fazer da confiabilidade da informação um problema menos relevante até a década de 2010.

A prática da desinformação foi definida pela UNESCO (Organização das Nações Unidas para a Educação, a Ciência e à Cultura) e pelo PNUD (Programa das Nações Unidas para o Desenvolvimento) como "conteúdo falso, manipulado ou enganoso, criado e disseminado intencionalmente ou não, e que pode causar danos potenciais à paz, aos direitos humanos e ao desenvolvimento sustentável". De fato, seus efeitos são sentidos em muitos campos: em campanhas eleitorais, nos temas de saúde pública (como ficou evidente na pandemia de Covid-19), na disseminação de discurso de ódio contra grupos sociais ou de ataque a reputação de ativistas, e em todas as disputas relevantes no campo socioambiental, apenas para citar os exemplos mais evidentes.

A tendência de segmentação do perfil da audiência ajuda a entender o investimento de grupos de interesse, governos e campanhas políticas no cruzamento de dados que garanta maior precisão na entrega e persuasão ao efeito da mensagem. Antes da popularização de plataformas e aplicativos, no entanto, os dados envolvidos no tratamento de perfis eram de difícil acesso, como no cruzamento de dados como cartões de crédito, revistas preferidas e local de moradia (Howard, 2006). A entrega de informações envolvia um outro conjunto de esforços, em serviços diferentes dos envolvidos no tratamento de dados de perfis.

Plataformas e aplicativos baseados em perfis pessoais e entrega de propaganda direcionada tornam muito menos custoso a obtenção e cruzamento de dados pessoais (inclusive relacionados ao posicionamento político) e, também, unificam a fonte dos dados e a entrega de informações direcionadas. Campanhas de desinformação se encaixam neste tipo de cenário porque dependem da identificação de nichos e entrega personalizada de mensagens impulsionadas e anúncios políticos. Ferramentas como *dark posts* (em que só um tipo de público pode visualizar informações postadas, sem que elas fiquem visíveis a todos na timeline das páginas de campanha), e a própria segmentação do público em caixas de ressonância de reafirmação de crenças e valores reduzem possíveis efeitos colaterais que estas publicações poderiam ter em outros públicos.

Em 2016, dois eventos fortemente polarizados – eleições nos Estados Unidos e o referendo do Brexit –, revelaram nichos propensos a mudar seu comportamento com base em campanhas de desinformação. A reação política colocou no centro da discussão a viabilidade e a necessidade de regulamentação das plataformas online visando impedir sua utilização em campanhas baseadas em informações falsas e enganosas. No Brasil, as eleições presidenciais de

1 Ver BRANT, João. Modelo de aplicativos de mensagens enterra o debate público. Folha de S.Paulo, 1º de novembro de 2020. Disponível em: <https://www1.folha.uol.com.br/ilustrissima/2020/10/modelo-de-aplicativos-de-mensagens-enterra-o-debate-publico.shtml>

2018 também foram marcadas por práticas de desinformação com impacto relevante na percepção dos eleitores[2].

Em poucos anos, este tema tornou-se uma pauta política central na manutenção de processos democráticos contemporâneos, tendo ensejado convocação de membros e donos de empresas de tecnologia para prestar esclarecimento em diferentes parlamentos, desenvolvimento de leis nacionais para lidar com o problema e experiências de regulamentações pelas próprias empresas.

As audiências com integrantes da Cambridge Analytica no Parlamento Britânico e no Congresso Americano com donos de empresas como Facebook, além de grupos de trabalho da União Europeia, marcam o início da popularização desse debate, que ganhou ainda mais visibilidade com filmes como Privacidade Hackeada (*The Great Hack*/2019) e O Dilema das Redes (*The Social Dilemma*/2020). No Brasil, a CPMI das Fake News reuniu denúncias de dissidentes do bolsonarismo sobre o funcionamento do chamado Gabinete do Ódio, além de especialistas sobre o tema.

A produção descentralizada de publicações que circulam nessas plataformas digitais, nesse sentido, gera desafios transnacionais de difícil solução. Por um lado, ferramentas de denúncia podem ser apropriadas por usuários e campanhas políticas para atacar adversários. Por outro lado, frente à impossibilidade de analisar milhares de *posts* e horas de vídeos em um único dia, empresas investem em filtros algorítmicos que atuam para identificar conteúdos e perfis potencialmente falsos. Ao fazer triagens e escolhas que envolvem questões polêmicas, este tipo de ferramenta transfere a programadores e códigos de aprendizado de máquina decisões sobre publicações consideradas possivelmente nocivas, muitas vezes sem revisão humana. Vieses em mecanismos de busca e análise de imagens, que reproduzem padrões racistas e sexistas a partir de aprendizado de máquina, talvez sejam o exemplo mais conhecido de problemas neste sentido.

Um dos desafios do fenômeno da desinformação é que, por se espalhar significativamente em espaços 'subterrâneos', ele não tem como ser adequadamente mensurado. Pesquisadores de todo o mundo estão dedicados a fortalecer os métodos e condições de pesquisa, mas têm trabalhado a partir de recortes limitados, que não servem de amostra de um universo complexo e não delimitável. Além disso, há pouca colaboração das empresas no sentido de imprimir transparência a suas práticas e aos processos de intercâmbio de conteúdo. Isso faz com que seja difícil compreender também as tendências de crescimento ou diminuição de ocorrência do fenômeno.

Ao mesmo tempo, o fenômeno da desinformação está profundamente imbricado com processos políticos e sociais e não é possível pensar soluções legais e regulatórias de forma isolada da compreensão desses processos e dos contextos nacionais e locais.

Além disso, legislações que enfrentam o problema da manipulação informativa se deparam com desafios multiplataforma e transnacionais. Isto porque as informações transitam por diferentes plataformas interconectadas por meio de ferramentas de compartilhamento, com lógicas de filtragem e termos de uso diversos, e os países por onde essas informações circulam não são os mesmos em que as empresas e seus bancos de dados estão localizadas. Os esforços coordenados das empresas sem este tipo de incentivo mostram-se ainda limitados em diferentes países.

Este paper analisa casos em que diferentes países (Alemanha, França, Canadá, Índia e Singapura) buscaram criar regras e conjuntos de práticas para mitigar os efeitos nocivos das campanhas online centradas em informações falsas ou enganosas. É analisada também uma iniciativa de âmbito regional, o Código de Práticas da União Europeia. São também apresentadas iniciativas formatadas, mas ainda não aprovadas, no Reino Unido e no Brasil.

Na primeira parte, os casos são apresentados individualmente. Na seção seguinte, buscamos destacar aspectos transversais que ajudem a identificar as tensões de cada proposta em relação a direitos fundamentais como liberdade de expressão e privacidade. Na terceira parte, buscamos abordar as questões ausentes nos textos regulatórios, mas que deveriam ser considerados no contexto de combate à desinformação. Ao fim, apresentamos considerações para os setores interessados em construir soluções regulatórias de perfil democrático.

---

2 Cf. SANTOS, João Guilherme Bastos e FREITAS, Miguel. WhatsApp, política móvil y desinformación: ¿cómo se dio la viralización de las noticias falsas en las elecciones brasileñas? Centro de Estudios en Libertad de Expresión y Acceso a la Información. Universidad de Palermo: Buenos Aires, 2019. Disponível em: <https://www.palermo.edu/Archivos_content/2020/cele/febrero/WhatsApp-politica-movil-y-desinformacion.pdf> e SOLANO, Esther; BRANT, João; BRITO CRUZ, Francisco et alli. Secretos y mentiras: WhatsApp y las redes sociales en las elecciones presidenciales de Brasil en 2018. Centro de Estudios en Libertad de Expresión y Acceso a la Información. Universidad de Palermo: Buenos Aires, 2019. Disponível em: <https://www.palermo.edu/Archivos_content/2020/cele/febrero/Secretos-y-mentiras-WhatsApp-y-las-redes-sociales%20.pdf>

2

ESTUDO DE CASOS

2.1 ALEMANHA NetzDG

A principal referência de regulação de discurso em plataformas online é a Network Enforcement Act (Netzwerkdurchsetzungsgesetz - NetzDG) da Alemanha, cujo objetivo é conter o discurso de ódio e demais conteúdos e expressões enquadrados como crime pelo Código Penal do país. Para alcançar o objetivo, a NetzDG, aprovada em setembro de 2017, criou uma série de obrigações a empresas de redes sociais que tenham objetivo de lucro e mais de 2 milhões de usuários registrados na Alemanha.

Por conta do histórico de crimes contra a humanidade perpetrados no país, a regulação de restrição a discursos de ódio e ou a qualquer manifestação de apoio ao nazismo ou outras organizações consideradas inconstitucionais é robusta, o que não impede que a Alemanha esteja entre os dez países que mais respeitam a liberdade de expressão, em levantamento da organização Artigo 19 relativo aos anos de 2019/2020 (ARTIGO 19, 2020a) e seja o 11º país que mais respeita a liberdade de imprensa na lista da Repórteres Sem Fronteiras.

A legislação pode ser comparada à proposta do Reino Unido no sentido de criação de um dever de cuidado, com parâmetros, mas sem a responsabilização das plataformas por conteúdos individuais postados por terceiros. A lei prevê que as empresas enquadradas derrubem conteúdo ilegal ou comportamento criminoso tal qual estabelecido no Código Penal - e encaminhem para órgãos competentes as denúncias -; produzam relatórios de transparência quanto ao gerenciamento de conteúdo e reclamações de usuários e garantam mecanismos de apelação e acompanhamento das medidas tomadas aos denunciantes. Além disso, devem treinar as equipes de moderação com relação à legislação alemã e indicar responsável legal no país para responder às requisições regulatórias. A lei cria ainda órgão administrativo para aplicar sanções e a possibilidade de entidade de autorregulação certificada pelo órgão administrativo.

A previsão mais polêmica aprovada é a criação de um prazo exíguo, 24 horas, para a retirada de conteúdo "obviamente ilegal", uma expressão considerada ampla pela maioria dos críticos, que poderia levar ao excesso de intervenção das plataformas que priorizariam a segurança financeira e jurídica. Há a possibilidade de reação em prazo maior, caso seja necessária análise pormenorizada para classificar o conteúdo. O período de sete dias pode ser estendido se a empresa requerer apoio à entidade de autorregulação certificada pela entidade administrativa governamental.

As redes sociais não podem ser sancionadas por erros individuais no gerenciamento de um conteúdo denunciado. Multas só podem ser impostas por violações "sistemáticas" da lei. Ou seja, se a rede social cometer um erro de julgamento, ou for negligente ou gerenciar um item por engano, não enfrentará responsabilização. E, no entanto, o valor de multas por falhas sistêmicas de aplicação da lei pode chegar a 50 milhões de euros. Em junho de 2019, o Facebook foi multado em 2 milhões de euros por subnotificar o número de reclamações que recebeu sobre conteúdo ilegal em sua plataforma. Diferentemente da Alphabet, que controla o YouTube e Google, e do Twitter, o Facebook não fez adaptações na plataforma para permitir que os usuários fizessem, sem dificuldades, denúncias de conteúdo ilegal de acordo com o disposto na lei. A opção da empresa fundada por Mark Zuckerberg foi criar um formulário para denúncias em páginas não facilmente acessíveis, o que implicou em menor número de questionamento de conteúdos[1]. Além disso, há alegações de que o Facebook estaria optando por classificar o conteúdo como infringente das regras da própria empresa, em vez de infrações legais.

Vale destaque que boa parte da polêmica envolvendo a NetzDG foi apaziguada com a exclusão da previsão de controle de novos uploads de conteúdo banido, uma vez que o contexto ou os comentários em torno de determinada imagem ou vídeo podem ser determinantes para configurar discurso ilegal ou não. Porém, organizações dedicadas a coibir o terrorismo apontam que a ausência do dispositivo tornou a lei pouco efetiva e eficiente.

---

1 Após o prazo para a publicação do primeiro relatório de transparência sob o Network Enforcement Act no verão de 2018, os meios de comunicação alemães relataram que o Facebook apresentou um número muito menor de reclamações do que outras redes sociais. O relatório do Facebook listou 886 reclamações sobre conteúdo ilegal em 1.704 postagens, das quais 362 foram excluídas. Esse número foi significativamente menor do que os números relatados ªºpelo Google e Twitter. O Google recebeu 215.000 reclamações por postagens em sua plataforma de vídeo, YouTube, e excluiu 58.000 postagens, enquanto o Twitter relatou 265.000 reclamações e a remoção de 29.000 postagens (BUNDESAMT FUR JUSTIZ, 2019).

À época da discussão do projeto de lei, a proposta foi contestada por organizações de peso[2]. O próprio relator para liberdade de expressão da Organização das Nações Unidas, David Kaye, manifestou preocupação com possível impacto da elevação do nível de gerenciamento de conteúdo de terceiros por plataformas digitais, temerosas de serem responsabilizadas com pesadas sanções, que pudesse significar censura privada. As primeiras análises do impacto da lei dão conta que as plataformas não optaram pelo caminho de derrubar conteúdo tão logo fosse denunciado, com menos de 20% dos itens destacados por usuários como ilegais se tornando objeto de remoção[3]. À mesma conclusão chegou o estudo realizado pelo professor da Humboldt Universität de Berlin, Martin Eifert, que concluiu que não houve efeito de bloqueio excessivo e que a lei foi eficaz[4]. Ainda assim, críticos apontam que não há evidências de que tenha havido redução na circulação de discurso de ódio e outras expressões proibidas[5].

Apesar das críticas de entidades especializadas em liberdade de expressão, a NetzDG conta com forte apoio popular (JACOB, 2018). Recentemente, inclusive, foram aprovadas novas disposições, a saber: empresas submetidas à NetzDG terão de relatar, de forma proativa, casos graves de discurso de ódio às autoridades, o que atraiu a atenção de defensores da privacidade, que apontam que a regra violaria direitos previstos em lei.

2.2 FRANÇA

A França tem dado seguimento à sua tradição regulatória e é um expoente em termos de esforços para a normatizar o discurso ilegal online. Já em dezembro de 2018, o parlamento aprovou a Lei Contra Manipulação da Informação (Lei 1202/2018), que visa impedir a interferência estrangeira nas eleições e aumentar a transparência em anúncios em plataformas digitais durante o período eleitoral.

A lei estabelece que provedores de plataformas digitais têm o dever de cooperar no combate à desinformação; obrigação de designação de um representante legal para ser o respectivo ponto de contato em território francês; criar um meio visível e facilmente acessível para os usuários sinalizarem informações falsas e fornecer uma declaração anual ao *Conseil Supérieur de l'Audiovisuel* (CSA), órgão regulador das comunicações, detalhando as medidas adotadas contra a disseminação de informações falsas. A França já dispunha de regulação enquadrando a disseminação de desinformação como crime. Plataformas também têm o dever de, durante o período eleitoral (definido como três meses antes do primeiro dia da eleição geral até a votação), garantir a transparência em relação ao conteúdo informativo patrocinado ligado a debates de interesse público, indicando identidade, valor e como os dados pessoais são utilizados.

A lei também criou um novo procedimento judicial para estabelecer a interrupção da difusão de uma alegação imprecisa ou enganosa ou imputação de um fato que pode deliberadamente alterar a lisura da próxima votação ou que esteja sendo disseminado de forma artificial ou massiva através de um serviço de comunicação online. Os casos devem ser analisados por juiz que deve emitir decisão em até 48 horas. A lei ainda reforça o poder do CSA para conter qualquer tentativa de campanha de desestabilização ou desinformação por um serviço de televisão controlado ou influenciado por um estado estrangeiro.

Apesar de polêmica, a lei, que não prevê sanções às redes sociais, foi considerada constitucional com mínimos ajustes. As principais críticas apontam para as definições abrangentes de conteúdos passíveis de serem suprimidos, o que permite análise bastante subjetiva de plataformas e juízes. Além disso, há enorme preocupação de a lei ser explorada principalmente por políticos e grandes empresas, com maior capacidade de litigância. A lei atribui maior responsabilidade às plataformas da internet, cuja interpretação do âmbito do conteúdo a ser derrubado pode ser muito mais ampla do que os acórdãos anteriores do *Conseil Constitutionnel* (corte constitucional francesa) e do Tribunal de Justiça das Comunidades Europeias.

A aprovação da lei em 2018, porém, não foi suficiente para responder a todas as demandas de regulação do discurso online. Em maio de 2020, o parlamento francês aprovou a Lei de Combate ao Ódio Online, buscando responder ao crescimento da ocorrência de discurso de ódio e incitação à violência online. A chamada "Lei Avia" foi criticada por grupos de diferentes espectros políticos, organizações da sociedade civil como a Artigo 19 e a Electronic Frontier Foundation, além da Comissão Europeia (EUROPEAN COMISSION, 2019), entre outras.

A lei, que teve posteriormente boa parte de seus artigos considerados inconstitucionais, previa obrigação de todos os sites removerem conteúdos envolvendo abuso sexual infantil e terrorismo, a partir de notificação da polícia e demais órgãos da administração pública, dentro de uma hora após notificação.

Além disso, a lei obrigava que redes sociais e buscadores avaliassem em 24 horas se conteúdos denunciados por usuários como ilegais o são de fato. O escopo do projeto de lei foi estendido de 'discurso de ódio' ilegal para uma ampla gama de outros conteúdos, incluindo: apologia a atos que consti-

---

2 Dentre as quais Open Knowledge Foundation, Chaos Computer Club, Repórteres Sem Fronteiras Alemanha e Wikimedia Foundation Alemanha. A Artigo 19 também manifestou-se contrária à proposta.

3 "This study shows that the reality is in between these extremes. NetzDG has not provoked mass requests for takedowns. Nor has it forced internet platforms to adopt a 'take down, ask later' approach. Removal rates among the big three platforms ranged from 21.2% for Facebook to only 10.8% for Twitter" (ETHIKSON, 2018).

4 Cf. CAMPOS, 2020.

5 Before delving into the transparency reports, it is important to note that these data only cover removal decisions arising from NetzDG complaints, and do not account for other removals based on other types of complaints, referrals, or injunctions. Furthermore, the metric of takedowns does not reveal whether NetzDG has achieved its purpose of combating hate speech and other online excesses. The differences between complaint mechanisms and the reports themselves make certain types of comparison difficult. It is also hard to know how the volume of content removal compares to the overall volume of illegal speech online" (TWOREK, 2019).

tuem um crime contra a dignidade humana, crimes de guerra, crimes contra a humanidade, escravidão, crimes de colaboração com um inimigo, interferência voluntária na vida ou integridade física, agressão sexual, roubo agravado, extorsão ou destruição, degradação voluntária ou deterioração que é perigosa para uma pessoa, assédio sexual, tráfico humano, proxenetismo (cafetinagem), incitamento ou desculpas de atos de terrorismo e conteúdo de abuso infantil. Diferentemente da Lei Alemã que serviu de inspiração, a versão francesa não contava com dispositivos de análise de conteúdo denunciado como ilegal em prazo dilatado diante da necessidade de avaliação apurada.

Em 18 de junho de 2020, a Corte Constitucional francesa julgou boa parte da lei inconstitucional (CONSEIL CONSTITUTIONEL, 2020). Em sua decisão, afirmou que considerou que certas disposições infringem "a liberdade de expressão e comunicação e não são necessárias, adequadas e proporcionais ao fim prosseguido". Foram mantidas algumas, menos da metade, das normas legais previstas, dentre as quais a permissão para criação de um tribunal judicial especial dedicado ao ódio online e de um observatório do ódio online destinado a monitorizar e analisar o conteúdo online e o seu desenvolvimento. A CSA atuará como secretaria deste observatório.

2.3 CANADÁ - ELECTIONS MODERNIZATION ACT

No Canadá, a Elections Modernization Act (BILL C-76), aprovada em dezembro de 2018, tem partes que buscam responder aos desafios que o ambiente digital impõe ao processo eleitoral. A lei, no entanto, não trata de forma detalhada nenhuma das questões de ponta (algoritmos de visibilidade, microtargeting em plataformas, contas inautênticas e comportamento coordenado).

As plataformas, que se enquadram nos critérios da lei, precisam registrar as propagandas, peças de campanha ou partidárias em seu espaço publicitário (seja direto ou indireto), bem como seus contratantes e agentes econômicos (tanto da plataforma quanto do contratante). Os critérios para que plataformas online sejam contempladas variam de acordo com a língua, como o número de acessos por mês (3 milhões em inglês, 1 milhão em francês e 100 mil se a língua não for nenhuma das anteriores).

Há similaridades com medidas adotadas em outros países, uma vez que a lei busca colocar limites aos gastos dos partidos e trazer mais transparência à participação de terceiros no processo eleitoral (o que inclui atividades partidárias, campanha/propaganda e pesquisas de opinião). Ou ainda na proteção aos dados pessoais, uma vez que a proposta exige que partidos publiquem sua política de proteção de dados em seus sites (aplicada aos dados de cidadãos aos quais os partidos têm acesso), sob pena de perder seu registro.

A regulação toma cuidado para proteger indivíduos que falam de política na internet sem finalidades comerciais, separando-os expressamente dos anunciantes em diferentes momentos. No entanto, mesmo a distribuição de links é considerada como propaganda ("providing a link to an Internet page that does anything referred to in subparagraphs (i) and (ii)"), abrindo flancos para indefinição entre compartilhamento orgânico e propaganda.

2.4 ÍNDIA

A Índia tem no Information Technology Act 2000 sua principal referência normativa para lidar com o problema da distribuição de conteúdos falsos e enganosos por meio de plataformas de mídias sociais e de mensagens instantâneas. O IT Act 2000 foi formulado, no entanto, para regular o comércio e crimes eletrônicos e, nesse âmbito, tratou como delito a publicação de informações deturpadas e falsas (misrepresentation), de conteúdos obscenos, de mensagens fraudulentas, bem como a adulteração de documentos e a violação de privacidade, entre outras situações. Este estatuto dá poderes ao governo de assumir a função de julgar o que é crime cibernético e tem tido artigos contestados por restrição da liberdade de expressão e vigilância massiva da população. A lei tem sido atualizada com emendas ao longo do tempo.

Em 2015, o artigo 66A, que previa punições, entre multas e prisões, para envio de mensagens ofensivas por meio de serviços de comunicação, foi considerado inconstitucional. Isto depois de cartunistas, políticos, empresários e jovens terem sido presos e/ou terem tido contas bloqueadas por publicações com críticas ao governo, especialmente em torno do tema corrupção, e a símbolos nacionais. Apesar de ter sido invalidada pela Suprema Corte, a Internet Freedom Foundation tem demonstrado que governos e polícias estaduais continuam recorrendo ao artigo 66A para bloquear contas e sites considerados por eles ofensivos[6].

O artigo 79, que na lei atual isenta intermediários de responsabilidade por conteúdos publicados por terceiros, por sua vez, é objeto de emendas. Na mais recente, a Information Technology [Intermediary Guidelines (Amendment) Rules] 2018, passa a constar a obrigação para que intermediários emitam termos de uso que proíbam usuários de publicar conteúdos "obsceno e prejudicial" e que ameace a "saúde e segurança pública"; removam acesso a conteúdo considerado ilegal, em até 24 horas, quando notificados pelo governo, assim como façam uso de métodos automatizados para remover, por si só, conteúdos ilegais; prestem assistência técnica ao governo em 72 horas após solicitação e que permitam que o autor da informação em apreciação seja rastreado; constituam sede no país se possuem mais de 5 mil usuários indianos ou residentes na Índia. Especialistas, organismos e empresas de tecnologia consideram que emendas como essas atacam severamente o direito à livre expressão (direito fundamental previsto no artigo 19 da Constituição) e o direito à proteção de dados pessoais.

6 Disponível em: <https://internetfreedom.in/how-a-bill-becomes-a-zombie-the-journey-of-section-66a-of-the-information-technology-act-2000/>. Acessado em 12 de novembro de 2020.

O rascunho dessa emenda foi apresentado em dezembro de 2018 para consultas e manifestação de partes interessadas. Em 2020, carta assinada por pesquisadores, ativistas digitais, jornalistas, dentre outros, dirigida ao Ministério de Tecnologia Eletrônica e Informação, conhecido pela sigla MEITY, destacou como principais preocupações "o uso desproporcional da regulamentação governamental", o caráter vago de termos que afrontam a liberdade de expressão, como "conteúdos extremamente prejudicial, hostil e odioso"; e a questão da quebra da criptografia que pode afetar a segurança das mensagens que circulam em aplicativos como WhatsApp e a privacidade dos indivíduos[7]. A versão definitiva ainda precisa ser apresentada pelo governo e aprovada pelo parlamento para ser transformada em lei.

2.5 SINGAPURA

Singapura instituiu, no ano de 2019, o Ato de Proteção contra Falsidades e Manipulação Online (traduzido do inglês, Protection from Online Falsehoods and Manipulation Act), também chamado pela sigla PROFMA. Conhecida como a "Lei das Fake News" do país, o estatuto aprovado pelo Parlamento em maio (que começou a vigorar em setembro daquele ano) proíbe a propagação online de fatos considerados falsos, ao mesmo tempo que cria medidas para neutralizar os efeitos dessas mensagens. A lei prevê punições principalmente a indivíduos, mas também a provedores, em caso de descumprimento de decisões que partem do governo central e de órgãos da administração direta, o que é considerado abusivo por organismos como a Organização Internacional de Juristas, Repórteres Sem Fronteiras e Human Rights Watch.

No intuito de impedir a circulação de falsidades e a manipulação informativa online, a lei nomeia como declaração falsa de fatos o ato de comunicar para um ou mais usuários finais, pela internet e/ou SMS, afirmações e materiais (dentre mensagens, artigos, discursos, posts, imagens, áudios, vídeos etc.) integral ou parcialmente falsos ou enganosos. Com base na lei, um indivíduo não pode publicar declaração falsa de fato e mensagens propensas a serem prejudiciais à imagem, segurança, saúde, finanças e tranquilidade públicas, bem como às relações amistosas e ao clima eleitoral. Adicionalmente, no âmbito do discurso de ódio, não pode incitar sentimentos de inimizade, ódio e má vontade entre diferentes grupos de pessoas, assim como não pode diminuir a confiança pública em relação ao governo.

O que é considerado falso e enganoso, no caso de temas como políticas públicas, eleições, serviços públicos e órgãos institucionais, depende da avaliação da administração direta. O ministro, nesse sentido, pode nomear Autoridade Competente (conselho/diretoria estatutária, incluindo titular de qualquer órgão a serviço do governo ou de uma autoridade estatutária) para desempenhar função pública e cumprir as instruções do governo. Os indivíduos podem ser penalizados com multas, que chegam a 100 mil dólares singapurianos (cerca de US$ 75 mil), e prisão de até 10 anos.

A lei visa ainda extinguir qualquer financiamento e promoção de declarações falsas de fato com apoio de locais online (site, página de internet, chat e fóruns etc. hospedados em computador e que podem ser vista, ouvida ou percebida pela internet). Outro foco desta normativa é a elaboração de medidas para coibir qualquer comportamento coordenado inautêntico e usos indevidos de contas online e de bots. Pela lei, atividades realizadas com duas ou mais contas para promover engano aos indivíduos são comportamentos coordenados inautênticos. Por fim, a PROFMA prevê a criação de medidas para aperfeiçoar a divulgação de informações sobre conteúdos pagos (na lei, apresentados como qualquer declaração comunicada em qualquer lugar com remuneração) com finalidades políticas.

Foram criadas três medidas para garantir a execução da lei. A primeira se chama Direção de Correções (Correction Direction) e consiste no envio de um aviso de correção a um indivíduo, que precisa seguir as orientações para informar publicamente de que a declaração por ele divulgada é falsa ou contém alguma falsidade. "Uma pessoa que comunicou uma declaração falsa de fato em Singapura pode receber uma instrução de correção mesmo se a pessoa não souber ou não tiver motivos para acreditar que a declaração seja falsa", afirma a letra da lei[8]. A segunda foi chamada de Direção para parar a Comunicação (Stop Communication Direction), que determina que o indivíduo precisa parar de publicar declarações sobre determinado assunto, ou algo substancialmente semelhante, nos termos especificados. De igual maneira, a instrução para interrupção de comunicação pode valer mesmo em casos em que o indivíduo não saiba que aquela declaração é falsa à luz do entendimento do governo.

A terceira medida se chama Ordem de Bloqueio de Acesso (Access blocking order), que é efetivada se o cidadão não cumprir os avisos de correção e de interrupção da comunicação. Nesse caso, a lei habilita o ministro a ordenar que o local online onde a declaração falsa de fato seja comunicada a usuários finais tenha o acesso desativado. Enquanto nos outros casos, punições como multa e prisão são direcionadas ao indivíduo, nesta situação, o provedor que não cumprir a ordem de bloqueio de acesso será considerado culpado e condenado a pagar multa de 20 mil dólares singapurianos (cerca de U$S 15 mil) por dia, até o total de 500.000 dólares singapurianos (cerca de U$S 375 mil).

Os afetados pelas medidas podem apelar ao Tribunal Superior, mas, antes, precisam solicitar ao Ministro para cancelar

7 Disponível em: <https://www.businesstoday.in/top-story/draft-information-technology-rules-experts-write-to-meity-highlight-key-concerns/story/317640.html>. Acessado em 12 de novembro de 2020.

8 Tradução de: "A person who communicated a false statement of fact in Singapore may be issued a Correction Direction even if the person does not know or has no reason to believe that the statement is false". Disponível em: <https://docs.google.com/document/d/1W4aUzkQw_NFbr3TEwI2HAWM6q-u8RmyA/edit>. Acessado em 12 de novembro de 2020.

ou alterar as determinações e ter tido o pedido negado. A lei prevê que o Tribunal só pode anular o ato quando a pessoa não for, de fato, responsável por comunicar a declaração; não se configurar uma declaração de fato ou se a declaração de fato for verdadeira; se não for possível, tecnicamente, cumprir a decisão. A ordem fica em vigor, mesmo após interposto recurso, até ser anulada pelo Tribunal Superior ou o Tribunal de Apelação (Court of Appeal), ou se expirar.

A primeira ordem de correção ocorreu em novembro de 2019 e foi direcionada ao líder de oposição Brad Bowyer (Progress Singapore Party) em função de uma publicação no Facebook, na qual questionava a independência de empresas de investimento[9]. A notificação que informava que ele propagou informações falsas e enganosas foi publicada na página de fact-checking governamental de Singapura[10]. Desde então, ativistas e jornalistas são advertidos e penalizados por publicarem conteúdo e/ou comentários que contenham crítica a alguma ação e órgão de governo em suas próprias páginas em plataformas de mídias sociais. A ONG Human Right Watch afirma, no relatório mundial de 2020, que a liberdade de expressão pode ser considerada restrita em Singapura[11]. O contexto do PROFMA empurrou o país da 151 para a 158 posição no Índice Mundial de Liberdade de Imprensa do Repórteres sem Fronteiras. Para a ONG, "2019 viu uma deterioração significativa com a adoção de uma lei de fake news com disposições orwellianas que permitem ao governo atuar como combinação de Ministério da Verdade e gabinete de censura na era da mídia social"[12].

2.6 UNIÃO EUROPEIA - EU CODE OF PRACTICE ON DISINFORMATION

Um dos melhores exemplos de esforços coordenados, unindo empresas como Facebook, Twitter, Google, Microsoft e Mozilla, é o chamado Código de Prática da União Europeia sobre Desinformação (EU Code of Practice on Disinformation), de setembro de 2018. Trata-se de um código de boas práticas das plataformas e associações para impedir ou mitigar a divulgação de informações falsas, feito em resposta aos problemas expostos pela comunicação da Comissão Europeia no documento Tackling online disinformation: a European Approach (COMISSÃO EUROPEIA, 2018). O documento trata principalmente da exposição dos cidadãos europeus a informações falsas ou enganosas em escala massiva, em que a amplificação por redes sociais online (através de algoritmos, anúncios direcionados ou robôs) figura como peça-chave nas causas do problema. As soluções propostas pela comissão passam por mais transparência, diversidade de informações, indicação de credibilidade (tags de conteúdo verificado, ou de que está em disputa) e letramento midiático.

O código de boas práticas assinado pelas empresas também busca compatibilidade com outras propostas e normas, como a Carta dos Direitos Fundamentais da União Europeia, Convenção Europeia de Direitos Humanos, as regulações EU 2016/679 sobre as proteções a pessoas no que concerne o processamento e movimentação de seus dados pessoais, e as diretivas 2000/31/EC (que trata de comércio eletrônico, particularmente artigos 12 ao 15, que tratam das garantias a serem dadas pelos Estados-membros, assegurando que a prestação de serviços preserve a confidencialidade de comunicações e evite armazenamento ou uso indevido), 2005/29/EC (sobre práticas comerciais desleais entre empresas e consumidores) e 2006/114/EC (versando, entre outras coisas, sobre propaganda enganosa).

O código prevê condutas e boas práticas para lidar com a disseminação de informação falsa online, comprometendo-se a (a) ter mais cuidado com a distribuição de anúncios direcionados envolvendo informações falsas, (b) tornar mais transparente os contratantes e valores pagos em anúncios envolvendo temas políticos ou controvérsias específicas (issue-based advertising), (c) integridade dos serviços evitando contas falsas ou usos impróprios de robôs, e, por fim, empoderando (d) consumidores e (e) a comunidade de acadêmica/de pesquisa. Os signatários forneceriam relatórios de avaliação anual e ações que os signatários podem colocar em prática para lidar com a desinformação. O documento assinado pelas empresas é um código essencialmente propositivo e principiológico, sem imposições ou sanções claras, e os signatários podem deixar os compromissos sem maiores consequências.

No primeiro relatório anual (outubro de 2019), foram abordadas as políticas sobre anúncios e quantidade de anúncios que levaram a respostas das plataformas, exposição dos compradores dos anúncios mostrados a usuários, remoção e bloqueio de contas com comportamento perverso, notificação a usuários e parcerias com fact-checkers para valorizar conteúdos checados e iniciativas envolvendo facilitação de acesso a dados para pesquisadores.

2.7 REINO UNIDO

O governo do Reino Unido tem se debruçado e tateado soluções para o problema que envolve conteúdo nocivo e ilegal online. Em abril de 2019, a Secretaria de Estado de Digital, Cultura, Mídia e Esporte e a Secretaria de Estado do Ministério do Interior apresentaram análise do contexto, impacto e os primeiros passos para forjar medidas regulatórias. Após apresentado publicamente, o Online Harm White Paper foi colocado em consulta pública. Em fevereiro de 2020, o relatório com a sistematização das contribuições foi divulgado.

---

9 Disponível em: <https://docs.google.com/document/d/1FbUvV-FWW2mE1p6vR8ATVqNGRT2G_Cdfoq-JBW7Whoas/edit#>. Acessado em 20 de novembro de 2020.

10 Disponível em: <https://www.scmp.com/week-asia/politics/article/3039260/singapore-invokes-fake-news-law-first-time-over-politicians>. Acessado em 20 de novembro de 2020.

11 Disponível em: <https://www.hrw.org/world-report/2020/country-chapters/singapore>. Acessado em 20 de novembro de 2020.

12 Tradução de: "2019 saw a significant deterioration with the adoption of an anti-fake news law with Orwellian provisions that allows the government to act as a combination of Ministry of Truth and censorship office for the social media era". Disponível em: <https://rsf.org/en/singapore>. Acessado em 20 de novembro de 2020.

Consta deste relatório a intenção do executivo enviar um projeto de lei que crie um "dever de cuidado" para empresas que atuam em serviços que permitam a interação com conteúdo gerado por terceiros ou interação entre usuários.

Este dever de cuidado será detalhado em códigos a serem desenvolvidos, observando cada um dos temas a serem aprovados pelo Ministro do Interior, e aplicados e fiscalizados pelo *Office of Communications* (Ofcom), o regulador para os serviços de comunicações, que na proposta ganhará novas atribuições. Empresas poderão apresentar formas alternativas às definidas nos códigos e terão de justificar por que seriam mais eficientes e adequadas. As prioridades para a produção de código são a proteção de crianças e adolescentes quanto à exploração sexual e combate a conteúdo que ofereça ameaça à segurança nacional e terrorismo. Quanto ao enfrentamento de fenômeno da desinformação, o documento prevê que as empresas tomem medidas proporcionais e proativas para ajudar os usuários a compreenderem a natureza e a confiabilidade das informações que estão recebendo, mas sem maior detalhamento.

Vale destacar que o governo sinalizou, na consulta pública, e de forma genérica, a intenção de regular os serviços de comunicação privada, à qual os respondentes se opuseram. No entanto, afirma o governo, houve reconhecimento em algumas respostas - tanto de indivíduos quanto de organizações - de que o abuso, o assédio e algumas das atividades ilegais mais graves ocorrem em espaços privados, como fóruns da comunidade fechados e salas de bate-papo. A proposta apresentada indica que o Ofcom terá um conjunto de poderes para tomar medidas de fiscalização eficazes contra empresas que violarem o dever legal de cuidado. Isso pode incluir o poder de emitir multas substanciais e impor responsabilidades a pessoas físicas que ocupem cargo da alta administração das empresas.

O regulador terá o poder de exigir relatórios anuais de transparência das empresas enquadradas no escopo, delineando a prevalência de conteúdo em suas plataformas e as medidas tomadas. Esses relatórios serão publicados online, para que os usuários e pais possam tomar decisões informadas sobre o uso da Internet. O regulador também terá poderes para exigir informações adicionais, incluindo sobre o impacto dos algoritmos na seleção de conteúdo para os usuários, e para garantir que as empresas relatem de forma proativa os danos emergentes e conhecidos.

O regulador irá encorajar e supervisionar o cumprimento de compromissos das empresas com pesquisadores independentes, bem como as salvaguardas adequadas para proteção dos usuários. Supervisionará ainda a disponibilidade de canais para reclamação dos usuários e as respostas às reclamações. Está em avaliação a permissão de alguns órgãos especiais fazerem "super reclamações" ao regulador.

Grupos de direitos humanos expressaram preocupações de que a abordagem de aplicação proposta possa ser desproporcionalmente punitiva, e o regulador precisaria demonstrar que cumpriu o teste de proporcionalidade para adotar medidas de cerceamento da liberdade de expressão de acordo com as leis de direitos humanos. A preocupação é maior por conta da previsão de bloqueio de serviços online direto pelos provedores de conexão. Além disso, há receio geral de que a pressão em torno das empresas por soluções possa levar ao cerceamento privado da liberdade de expressão, por precaução contra eventuais punições. No caso de medidas para conter fenômenos cujo conceito é bastante aberto, como desinformação, há temor de cerceamento do exercício legítimo da liberdade de expressão.

O white paper foi publicado em 2019 e o relatório de consulta no início de 2020[13].

2.8 BRASIL - PROJETO DE LEI 2630/2020

No Brasil, um projeto de lei aprovado no Senado, o PL 2630/2020, de autoria do senador Alessandro Vieira, pretende regular redes sociais e serviços de mensageria instantânea para conter o fenômeno da desinformação e assegurar maior transparência de como as plataformas vêm gerenciando conteúdo. O texto foi aprovado no Senado Federal em junho de 2020, mas no momento de fechamento deste texto [dezembro de 2020] seguia sem apreciação da Câmara dos Deputados.

A proposição prevê a criação de uma entidade para acompanhar a aplicação da regulação, responsável também por seu detalhamento em código de conduta, com participação multissetorial, e instaurada no âmbito do legislativo. O PL também traz uma série de obrigações quanto a impulsionamento e publicidade, especialmente eleitoral, e proibição de robôs não identificados, além de relatório de transparência com uma série de requisitos. Há previsão de guarda da identificação de anunciantes e guarda de metadados de mensagens privadas que forem compartilhadas em grupos e que alcançarem determinado patamar de compartilhamento. Além disso, prevê alteração da lei que rege o serviço móvel privado (celular) para obrigar que as empresas de telecomunicações validem os dados cadastrados dos usuários.

Diante da realidade nacional, em que integrantes dos Poderes Executivo e Legislativo são considerados importantes disseminadores de desinformação e de ataques, a proposta incide sobre o uso de redes sociais pelo poder público e servidores públicos.

A redação votada pelo Senado incorporou pontos sobre o devido processo na moderação de conteúdos por parte das plataformas, como mecanismos de notificação e direito de defesa dos usuários, importantes para o exercício da liberdade de expressão. Caso a aplicação de moderação seja considerada inadequada, caberá à plataforma reparar dano, fruto da interferência.

13 Nota do editor: em dezembro de 2020, após o fechamento deste paper, foi publicada nova versão, ainda mais detalhada, da resposta do governo à consulta.

O projeto ainda prevê ações céleres das plataformas e reações às denúncias, oferecendo direito de resposta, num formato vago e pouco claro. O texto estabelece que a decisão do procedimento de moderação deverá assegurar "ao ofendido o direito de resposta na mesma medida e alcance do conteúdo considerado inadequado".

Entre os pontos mais debatidos e que divide opiniões está a guarda, a priori, de metadados de mensagens em serviços de mensageria. A proposta é que sejam guardados os metadados daquelas mensagens que alcançarem patamar de compartilhamento e em grupos, a fim de viabilizar, por ordem judicial, a identificação de responsáveis por mensagens consideradas ilícitas. Grande parte das entidades da sociedade civil defende a supressão total do artigo 10, e o WhatsApp, aplicação do Facebook, apresentou como alternativa a guarda de metadados de interações apenas a partir de instauração de investigação. Outras organizações da sociedade civil defendem o artigo com base na leitura de que a atual impossibilidade, na prática, de responsabilização legal dos responsáveis por conteúdo ilícito, funciona como incentivo à prática de desinformação.

3

ANÁLISE TRANSVERSAL DOS CASOS

A avaliação dos textos legais, apresentados no capítulo anterior, permite a identificação de questões transversais que oferecem uma chave de leitura sobre quais são os nós críticos para a abordagem regulatória acerca da desinformação.

Esta seção apresenta um panorama sobre os principais pontos identificados. A tabela anexa a este paper localiza como cada um dos casos responde a cada um desses nós críticos.

1) Definição e arbítrio sobre veracidade – a definição sobre veracidade, ou sobre o critério que separa a informação falsa da verdadeira, é um nó central dos debates de desinformação. Até hoje, em um cenário de prevalência do jornalismo profissional, essas definições, via de regra, se davam antes de tudo pelos meios de comunicação editores e, quando disputadas, pelo Poder Judiciário. Até recentemente, as plataformas não assumiam o lugar de editores de conteúdo e não arbitravam sobre veracidade de conteúdo postado por terceiros. Contudo, o volume e o impacto de fraudes informativas e conteúdos ilícitos passaram a levantar o debate sobre a necessidade de as plataformas, ou agências de checagem, terem o poder de arbitrar quanto à veracidade das informações. Esse modelo passou a ser adotado nos últimos anos e justifica, em vários casos, remoção de conteúdo ou redução do alcance. A opção de se trabalhar com o arbítrio privado sobre a verdade levanta preocupações em relação à liberdade de expressão dos usuários, já que passa a atribuir a um ator privado o papel de juiz. Além desses dois cenários, há também outros mais críticos em que o governo assume o papel de árbitro, como em Singapura, em que a lei deu a órgão da administração direta, ligada ao Poder Executivo, o papel de definir a veracidade de determinadas informações.

Nos textos analisados, embora fake news apareçam como expressão de fachada e, por mais que informações, conteúdos e declarações falsas apareçam como um dos principais alvos das normas jurídicas, nem sempre uma definição é apresentada. Ao mesmo tempo, algumas dessas leis e desses códigos não regulamentam necessariamente comunicações falsas, mas optam por enquadrar como crime especificamente discursos de ódio com intuito de evitar que extremismos de todo tipo prosperem online.

2) Responsabilidade pela aplicação – os modelos analisados combinam, em diferentes proporções, sistemas de autorregulação, corregulação e regulação pública. Na autorregulação, a definição de critérios e sua aplicação ficam por conta das empresas. Na cor-regulação, a definição de critérios se dá por lei, a aplicação inicial se dá pelas empresas e órgãos públicos supervisionam a aplicação. Na regulação pública, os critérios se dão por lei ou por normatização infralegal e a aplicação se dá diretamente por órgão público. Cada um desses modelos tem vantagens e desvantagens. Em todo caso, a ausência de critérios públicos de moderação de conteúdo impacta direitos fundamentais do usuário tanto quando a aplicação é feita pela própria empresa como quando ela é feita por entidade reguladora.

3) Atribuição e tipo de responsabilidade legal – alguns textos legais atribuem responsabilidade legal apenas aos usuários responsáveis por conteúdos falsos ou enganosos, outros atribuem responsabilidade também às plataformas. No caso da responsabilidade dos usuários, ela pode ser civil ou penal. No caso da responsabilidade das plataformas, ela pode se dar em três níveis: em primeiro lugar, sobre cada um dos casos individuais em que há violação de direitos ou prática de ilícito por usuários e não houve ação da plataforma. Em um segundo nível, sobre um dever de cuidado em geral, como propõem Alemanha e Reino Unido, sem responsabilidade por cada um dos casos. Ou, ainda, num terceiro nível, responsabilidade apenas nos casos em que a plataforma deixa de respeitar ordem judicial, modelo que prevalece hoje nos Estados Unidos, na União Europeia e no Brasil. Este tema será detalhado na seção seguinte.

4) Abordagem preferencial – os textos legais trazem diferentes abordagens do problema na tentativa de enfrentar a desinformação. Em alguns casos, o foco está nos conteúdos falsos ou enganosos. Uma segunda abordagem tenta incidir sobre o comportamento dos usuários, buscando impedir, por exemplo, os robôs não identificados no Twitter, a omissão da identidade do responsável por determinada página no Facebook ou a distribuição de disparos em massa no WhatsApp. Uma terceira abordagem é buscar afetar o financiamento da desinformação, por exemplo, por meio da transparência sobre a responsabilidade dos anúncios políticos. Uma quarta abordagem possível é quando se busca afetar a arquitetura das redes sociais, por exemplo, ao impor às empresas a obrigação de guarda de dados.

5) Alcance e jurisdição – as características que definem a Internet como de alcance global, seu caráter privado, e a propensão à inovação e a digitalização ainda se materializam perante os reguladores como dificuldades de aderência dos mecanismos tradicionais do Direito às interações virtuais.

É neste cenário que diversas regulações iniciam experiências de regulação de conteúdo em plataformas que utilizam conteúdo gerado por terceiros. Notável, certamente, é a tentativa de o Estado nacional fazer valer sua lei, sendo a obrigação de indicação de responsável legal pela empresa uma demanda que se repete em diversos diplomas legais e proposições regulatórias aqui analisadas. A estratégia visa enfrentar o fato de a Internet permitir a oferta de serviço global sem necessidade de presença formal nas localidades.

Outra estratégia para lidar com este desafio colocado pelo modelo descentralizado de prestação de serviços online é a regulação com extraterritorialidade, afirmando o poder do Estado de ir além de seus territórios em determinadas condições. Em Singapura, a lei para combater declarações consideradas falsas permite que uma corte cite uma pessoa que cometeu uma ofensa fora do país, como se o crime fosse perpetrado no território.

A Internet Society, em documento de análise de previsões legais com impacto extraterritorial, sugere que governos busquem dialogar com outros *stakeholders* para obter o resultado esperado (INTERNET SOCIETY, 2018). De fato, iniciativas multissetoriais e transnacionais poderiam servir de referência para ações mais efetivas. Um bom exemplo é a iniciativa Christchurch Call[1], encabeçada pelos governos da França e da Nova Zelândia, com apoio inicial de outros 15 países[2], mais a Comissão Europeia. A iniciativa foi criada em reação aos ataques terroristas de 15 de março de 2019 na comunidade muçulmana de Christchurch, na Nova Zelândia. O Christchurch Call consiste em uma série de propostas de compromissos voluntários para governos e redes sociais agirem de forma coordenada e sistemática para conter conteúdo extremista violento online e prevenir o abuso da Internet por organizações e indivíduos terroristas. Embora esteja baseado em compromissos genéricos, o chamado obteve adesão significativa. Em setembro de 2019, outros 31 países, mais a Unesco, aderiram, bem como os grandes provedores de serviços online como Amazon, Google, Facebook, YouTube, Twitter, Microsoft, Qwant, Line, DailyMotion e JeuxVídeos. Iniciativa semelhante poderia ser realizada para o tema da desinformação.

6) Grau de acompanhamento da ação das plataformas pelo poder público ou por pesquisadores independentes – diferentes textos legais estabelecem obrigações de transparência para as empresas, que servem não só para permitir ao usuário conhecer e acompanhar as políticas de moderação de conteúdo como para avaliar a eficácia da aplicação das leis e normas regulatórias. No caso da Alemanha, por exemplo, o Facebook foi multado porque, entre outras coisas, o número de denúncias registrado de violação à NetzDG foi muito abaixo das demais plataformas. No Brasil, durante o processo eleitoral de 2020, o Facebook afirmou remover mais de 140 mil conteúdos por violação de políticas de interferência eleitoral. Sem o acesso a dados por pesquisadores independentes torna-se difícil saber se as empresas estão aplicando critérios definidos em lei, se estão interferindo sobre conteúdo legítimo ou ainda se estão sendo negligentes.

Todos esses nós críticos contêm tensões entre direitos fundamentais, sendo os mais comuns: liberdade de expressão individual x acesso à informação (entendida como informação plural, diversa e confiável), e liberdade de expressão individual x privacidade e proteção de dados. Ao tratar da regulação para conter o fenômeno da desinformação, cabe tomar em conta a obrigação dos governos de garantir a plena fruição da liberdade de expressão e acesso à informação, sendo que tais direitos estão diretamente relacionados.

A Corte Interamericana de Direitos Humanos (CIDH) manifestou o entendimento de que "A liberdade de expressão é a pedra angular da própria existência de uma sociedade democrática. É essencial para a formação da opinião pública. É também uma condição *sine qua non* para partidos políticos, sindicatos, sociedades científicas e culturais e, em geral, aqueles que desejam influenciar a coletividade poderem se desenvolver plenamente. É, em suma, uma condição para a comunidade, ao exercer suas opções, estar suficientemente informada" (Corte IDH, 1985). Assim, é possível afirmar que uma sociedade que não está bem informada não é totalmente livre para se expressar. Neste sentido, a busca por um ciclo virtuoso entre liberdade de expressão e o direito de estar bem informado deve ser almejada pela política pública.

A compreensão de necessidade de regulação não significa que se depreenda de imediato o contorno adequado para alcançar o objetivo almejado. No caso da regulação da internet e do ambiente digital, vale relembrar a teoria desenvolvida por Lawrence Lessig (2006) de que, além da lei, do mercado e das normas sociais, o próprio código de programação (a arquitetura) tem fortes efeitos regulatórios.

---

1 <https://www.christchurchcall.com/call.html>

2 Fundadores, Nova Zelândia e França. Apoiadores fundadores: Austrália, Canadá, Comissão Europeia, França, Alemanha, Indonésia, Índia, Irlanda, Itália, Japão, Jordânia, Holanda, Nova Zelândia, Noruega, Senegal, Espanha, Suécia e Reino Unido. Apoiadores: Argentina, Áustria, Bélgica, Bulgária, Chile, Colômbia, Costa Rica, Chipre, Dinamarca, Finlândia, Geórgia, Gana, Grécia, Hungria, Islândia, Costa do Marfim, Quênia, Letônia, Lituânia, Luxemburgo, Maldivas, Malta, México, Mongólia, Polônia, Portugal, Romênia, Coréia do Sul, Eslovênia, Sri Lanka, Suíça, Unesco, Conselho da Europa. Provedores de serviços: Amazon, DailyMotion, Facebook, Google, Microsoft, Qwant, Twitter, YouTube, Line.

Tabela 1
Análise transversal – Principais tensões regulatórias

	Lei NetzDG (Alemanha)	Lei 2018-1202 (França)	Lei de Modernização das Eleições (Canadá)	Lei da Tecnologia da Informação (Índia)	Lei de Proteção contra Falsidades e Manipulação (Singapura)	Código de Prática sobre Desinformação (União Europeia)	Livro Branco sobre Danos Online (Reino Unido)	Projeto de lei de combate às *Fake news* 2630/20 (Brasil)
Aprovação	**2017**	**2018**	**2018**	**2000/2018**	**2019**	**2018**	**Em discussão**	**Em discussão**
Definição e arbítrio sobre veracidade	Cobra atuação direta das plataformas contra conteúdos 'evidentemente ilegais' (tipos penais já previstos no Código Penal alemão)	Empodera (e cobra) plataformas para atuar e arbitrar diretamente	Não. Foco é sobre transparência do financiamento de anúncios no processo eleitoral	Sim, pela Administração direta	Sim, pela Administração direta	Empodera plataformas para atuar e arbitrar diretamente	Empodera (e cobra) plataformas para atuar e arbitrar diretamente	Não cria regras sobre isso
Responsabilidade pela aplicação (governamental, autorregulação ou corregulação)	Corregulação (autorregulação regulada)	Corregulação (monitoramento pelo Conselho Superior do Audiovisual), com rito especial para o processo eleitoral	Autoridade eleitoral	Regulação governamental	Regulação governamental	Autorregulação	Corregulação (critérios em lei, aplicação pelas plataformas e supervisão pelo Ofcom)	Corregulação (regras definidas em lei aplicadas pelas plataformas), com supervisão de Conselho ligado ao Congresso
Atribuição e tipo de responsabilidade legal (indivíduos ou plataformas)	Impõe um tipo de 'dever de cuidado' às plataformas	Impõe um tipo de 'dever de cuidado' às plataformas	Impõe obrigação a campanhas, partidos, candidatos e apoiadores	Isenta plataformas de responsabilidade por conteúdos de terceiros, mas emenda proposta impõe responsabilidade de remoção e bloqueio	Prevê sanções duras a indivíduos e plataformas	Não há. Cria Código de Prática de adesão voluntária das plataformas, sem sanções	Impõe um 'dever de cuidado' às plataformas. Foco é na responsabilização das plataformas	Impõe às plataformas obrigações equivalentes a um dever de cuidado. Não impõe sanções a usuários
Abordagem preferencial	Conteúdo	Comportamento dos usuários e financiamento	Financiamento	Conteúdo e arquitetura das plataformas	Conteúdo e comportamento dos usuários	Conteúdo e comportamento dos usuários	Conteúdo e arquitetura das plataformas	Comportamento dos usuários e arquitetura das plataformas
Alcance e jurisdição	Nacional, com impacto extraterritorial	Nacional	Nacional	Nacional	Nacional, com impacto extraterritorial	Regional (União Europeia)	Nacional	Nacional
Transparência/ acompanhamento por pesquisadores independentes	Impõe significativas obrigações de transparência às plataformas	Sim, quanto a impulsionamentos, valor e dados utilizados. Além disso, propõe transparência dos algoritmos	Foco na transparência dos gastos em anúncios eleitorais nas plataformas	Não	Não	Sim, sobre anúncios políticos e algumas medidas gerais para as plataformas	Sim, órgão regulador pode solicitar relatório anual, informações e explicação sobre impacto dos algoritmos	Sim, impõe obrigação de transparência das plataformas, inclusive de impulsionamento e publicidade

4

RESPONSABILIDADE DOS INTERMEDIÁRIOS E DOS USUÁRIOS

O tema que foi historicamente central na organização do debate regulatório sobre conteúdo online é o de responsabilidade dos intermediários. A aprovação, em 1996, nos Estados Unidos, do Communications Decency Act (DCA), surgida após a tentativa de regulação da pornografia online, estabeleceu um modelo de regulação para plataformas digitais, sendo aqui considerados os negócios que conectam um ou mais lados de transações, utilizando conteúdo produzido por terceiros.

A Seção 230 do DCA passou a proteger websites e posteriormente plataformas: "Nenhum provedor ou usuário de um serviço de computação interativa deve ser tratado como o editor ou orador de qualquer informação fornecida por outro provedor de conteúdo de informação." A seção não apenas protege sites e plataformas de serem legalmente responsáveis pelo conteúdo de terceiros, mas permite "qualquer ação tomada voluntariamente de boa-fé para restringir o acesso ou disponibilidade de material que o provedor ou usuário considerar obsceno, lascivo, sujo, excessivamente violento, hostil ou de outra forma questionável", sem qualquer responsabilidade por suas decisões. Esta ampla autorização (especialmente marcada pela expressão 'de outra forma questionável') oferece a possibilidade de aplicação de suas regras comunitárias sem apresentar qualquer justificativa para remoções e sem condições.

O trecho da lei, que passou a ser citado como "a ferramenta mais importante já existente para a liberdade de expressão na Internet" ou "a lei mais importante da Internet" criou um padrão, que foi adotado de formas adaptadas, com algumas diferenças, na diretiva de comércio eletrônico da União Europeia, em 2000, e no Marco Civil da Internet do Brasil, em 2014, entre outros países. Os usuários e os serviços online utilizados não se confundem, de forma que se criou segurança jurídica para que os conteúdos postados não tivessem de ser constantemente analisados pelas empresas então nascentes e, em qualquer ameaça de risco jurídico, retirados do ar. Esta proteção permitiu que as plataformas crescessem sem serem incomodadas a todo momento por pessoas descontentes com o conteúdo publicado por seus usuários e sem gerar um efeito silenciador sobre a liberdade de expressão dos usuários.

Cabe destacar que a opção do legislador americano não foi capaz de criar um ambiente de total liberdade de expressão sem considerar a necessidade de conter determinados discursos. Pelo contrário, ao evitar a responsabilidade objetiva pelo conteúdo gerado por terceiros previa permitir que os intermediários aplicassem os melhores esforços para evitar conteúdo danoso e ilegal e não tivessem que escolher entre: (i) tentar moderar conteúdo de terceiros e ser responsabilizado de forma objetiva quando falhasse, ou (ii) abrir mão totalmente da tentativa de conter abusos.

No entanto, a escolha de regime de responsabilidade civil escolhido é apontada como o pilar do baixo investimento e reação lenta das plataformas ante as externalidades negativas de seus negócios, incluindo o fenômeno da desinformação em escala global, explorada especialmente por grupos políticos, que usam informações falsas ou enganosas para disputar as posições dos cidadãos.

Neste sentido, observa-se globalmente uma tentativa de encontrar um novo regime de responsabilidade para as plataformas digitais quanto ao conteúdo produzido por terceiros. No caso das propostas do Reino Unidos e Brasil e das leis do Canadá e Alemanha, é possível afirmar que há um esforço de apontar o que se espera dessas empresas em termos de aplicação de melhor esforço para conter conteúdo danoso, sem que sejam responsabilizadas a cada conteúdo produzido por seus usuários. Cabe destaque que algumas das regulações apontadas como destruidoras do padrão Seção 230 do CDA americano não o são. O que se cria é uma responsabilidade sistemática de tomar medidas preventivas e reações céleres diante de conteúdo ilegal ou danoso. Isso não significa que diante da inação quanto a um conteúdo específico exista uma mudança no regime de responsabilidade civil.

Parte dos governos, no entanto, busca soluções no sentido de promover maior proatividade das plataformas para evitar que suas atividades econômicas impactem direitos assegurados em leis e permitam a perpetração de crimes. Em termos gerais, este segundo modelo pode ser exemplificado pelas regulações indianas e de Singapura. As leis mencionadas, ao promoverem a gestão individual de conteúdo a partir de demandas de governo baseadas em conceitos genéricos de desinformação - sem necessidade de participação da Justiça

ou órgão independente - não observam os princípios de necessidade, proporcionalidade e legalidade que norteiam qualquer iniciativa legítima de cerceamento da liberdade de expressão pelos padrões internacionais, e podem ser enquadradas como normas de aplicação de censura estatal.

Embora a proteção das plataformas contra a responsabilização de conteúdo por terceiros seja importante para evitar o efeito silenciador, a questão da responsabilização dos indivíduos por conteúdo disseminado na rede está longe de ter encontrado um ponto de equilíbrio. Diversas análises críticas acerca da resposta a conteúdos de ódio, por exemplo, apontam para a baixa capacidade dos estados e dos sistemas de Justiça punir os perpetradores de crimes relacionados a calúnia, injúria, difamação, racismo, homofobia, transfobia, ameaça etc. Nesse sentido, segue relevante a busca de modelos que consigam proteger a liberdade de expressão ao mesmo tempo que oferecem mecanismos eficazes para defender os demais direitos humanos com os quais ela pode estar em colisão. Talvez o maior desafio seja como garantir que as respostas contra conteúdos ilegítimos sejam adequadas ao seu enorme volume e à sua rápida velocidade de disseminação.

5

TEMAS AUSENTES

Enquanto a seção anterior analisou os temas transversais presentes nos diferentes textos legais, essa seção se propõe a discutir os temas ausentes. Questões com potencial para incidir sobre a prática de desinformação, mas que têm sido negligenciadas ou secundarizadas nas definições regulatórias.

5.1 PROTEÇÃO DE DADOS

Quando olhamos apenas para o conteúdo da manipulação informativa, o que dá sentido a essas estratégias fica em segundo plano: o direcionamento dessas informações aos grupos mais propensos a compartilhá-las, auxiliando a viralização e apropriando-se dos critérios de visibilidade segmentada dos algoritmos nas plataformas. A mesma informação, lançada inicialmente em grupos com perfis diferentes, pode desencadear processos virais ou ser completamente ignorada (Margetts et al, 2016). É preciso levar em conta o fato de os conteúdos não transitarem de modo autônomo e a pluralidade de modos como as pessoas podem reagir ao contato com eles.

Considerando o fluxo dessas informações e a precisão de seus alvos, a proteção aos dados pessoais é uma das poucas ações capazes de interferir, a um só tempo, no tratamento de dados por agências que vendem serviços irregulares de segmentação para disparos criminosos no WhatsApp, na ampliação da visibilidade para nichos específicos seguindo o funcionamento dos algoritmos de plataformas como Facebook e na utilização desses dados para venda de propaganda direcionada no Instagram ou outras redes. Isso acontece porque a proteção reconhece que os cidadãos são titulares dos próprios dados (que neste caso não podem ser tratados para fins com os quais não consentiu) e exige medidas de proteção e transparência nos casos em que tratamentos são autorizados.

Neste ponto, cabe lembrar a comparação feita por Bimber et al (2012): do mesmo modo como o impacto dos automóveis no planejamento urbano faz com que todos aqueles que não possuem carros sejam afetados na distribuição das ruas, limitação das calçadas, velocidade dos fluxos da cidade, políticas de transporte etc., o impacto social da internet supera o simples acesso à rede. O mesmo vale para dados pessoais. Além de produtos de comunicação, os aparelhos e plataformas utilizando dados pessoais perpassam outras dimensões cotidianas - operações bancárias, relações profissionais, alimentação, locomoção - igualmente envolvidas na produção descentralizada de dados extremamente úteis à identificação de perfis e comunicação política direcionada.

5.2 ARQUITETURA DAS PLATAFORMAS E APLICATIVOS

O modelo de negócio das aplicações baseia-se na perspectiva de acumular milhões de pontos de dados sobre cada utilizador para viabilizar a criação de perfis de utilizadores, de modo a oferecer perfis hipersegmentados aos anunciantes. Assim, o modelo de negócio baseado na acumulação de dados gera fragmentação e os algoritmos e a inteligência artificial são guiados por valores mais relacionados com a uniformidade e a semelhança do que o pluralismo e a diversidade. Esta contradição está diretamente relacionada com o objetivo de manter o utilizador por mais tempo na plataforma. Ela provém da arquitetura e arranjos de infraestruturas das redes sociais, que geram impacto no fluxo e priorização do conteúdo de informação, ideias e opiniões. Assim, a fragmentação e a segmentação são fenômenos que são reforçados numa espiral e reforçam o que os estudiosos chamaram de filtros bolhas e câmaras de eco (PARISER, 2011).

Outro ponto que fica em segundo plano são as diferenças drásticas entre a economia política dos atores envolvidos nas plataformas e redes sociais e a economia política na comunicação broadcast. As plataformas unificam em uma mesma rede social online canais relacionados às diversas indústrias de comunicação broadcast - radiodifusão, cinematográficas, musicais, imprensa e publicações eletrônicas, jogos eletrônicos ou digitais, marketing, relações públicas, propaganda - ao mesmo tempo em que o conteúdo disponibilizado online é produzido de modo descentralizado e distribuído de modo segmentado (narrowcast). Diferentes indústrias entram em uma competição de nicho, disputando as recomendações dos algoritmos de visibilidade e a atenção de usuários, em um embate em que "influencers" individuais trazem elementos novos ao debate.

Por um lado, os algoritmos alteram a propensão a indicar vídeos de acordo com o engajamento dos usuários que tiveram contato com estes vídeos; por outro, aplicativos podem ser utilizados em campanhas canalizando acessos e engajamento em outras plataformas, traçando interconexões complexas entre usuários, diferentes plataformas e seus algoritmos. Diversas plataformas compõem um mesmo sistema de redes interconectadas, em que alterações em um ponto podem gerar processos adaptativos em outras redes e não se pode alterar um ponto esperando que o restante permaneça estático. A concentração de propriedade diz respeito, portanto, à concentração do poder de tomar decisões como critérios dos algoritmos, políticas de uso e privacidade dos dados (pontos marginais nas discussões sobre meios broadcast), e não a concentração dos atores aptos a divulgar conteúdo em um canal específico. Também concentra poder ao definir os limites e a responsabilização dos profissionais que escrevem algoritmos que tomarão decisões (quais conteúdos priorizar, quais devem ser exibidos o mínimo possível, qual conteúdo é uma ameaça em potencial), muitas vezes reproduzindo e acentuando desigualdades estruturais.

5.3 OPACIDADE E VIRALIZAÇÃO

Por outro lado, aplicativos de mensagem sem algoritmos de visibilidade possuem um papel central em campanhas de desinformação. Destacamos quatro aspectos: (i) na ausência de timeline, as postagens ficam armazenadas nos aparelhos dos usuários (enquanto retirar uma postagem com um milhão de compartilhamentos no Facebook deleta a fonte desses compartilhamentos, no WhatsApp mensagens com um milhão de encaminhamentos – para não falar de grupos – estão em um milhão de aparelhos diferentes, que podem inseri-las de volta em outras redes a qualquer momento; (ii) elas garantem que a informação será exposta a todos os integrantes de grupos em que essas informações chegam (não importando o quão desinteressados estejam em relação à política ou quão mal tenham reagido em contatos prévios), repetidamente; (iii) a combinação entre encaminhamentos e grupos viabiliza um aumento exponencial de visibilidade propenso à viralização, particularmente considerando a possibilidade de construção de grupos segmentados com base em dados pessoais retirados de outras redes (se cada integrante em um grupo com 256 pessoas encaminhar para um novo grupo cheio, 65,5 mil pessoas são alcançadas na primeira rodada de encaminhamentos e 16,7 milhões na segunda); (iv) por fim, o padrão de mensagens privadas (criptografadas e opacas por motivos de segurança) faz com que a investigação e monitoramento dessas estratégias sejam limitados, bem como a identificação de informações viralizando em nichos distantes dos chamados "grupos públicos".

Vale ressaltar que é a combinação desses diferentes aspectos que torna o uso desses aplicativos problemático, e não características como design privado, criptografia ou ausência de filtros. Diversas plataformas têm aplicativos com design privado e não tiveram papel de destaque nas eleições recentes, como Messenger do Facebook, Direct do Instagram, Messages do Twitter, a despeito de sua integração com essas plataformas.

6

CONCLUSÕES E RECOMENDAÇÕES

A análise sobre os instrumentos legais de combate à desinformação, já aprovados ou em estágio avançado no processo decisório, ajuda a iluminar os desafios de dar respostas regulatórias ao problema. Em primeiro lugar, pela ameaça que geram a direitos fundamentais. Há críticas contundentes direcionadas a quase ou todos os conteúdos normativos das legislações. Da mesma forma, por razões variadas, em muitos países, artigos foram posteriormente considerados inconstitucionais por serem abusivos e afrontarem a liberdade de expressão.

É possível situar o principal desafio em dois nós críticos: o primeiro é o de definir e arbitrar sobre 'a verdade'. Toda a tradição regulatória das comunicações no século XX evitou esse caminho e foi ajudada por isso pela centralidade que tinha o jornalismo profissional na esfera pública. Com a reorganização do ambiente informacional, impôs-se a necessidade de lidar com a escala e a velocidade de circulação de notícias falsas ou enganosas, que dificulta que todos os conflitos sejam levados à justiça. A falta de parâmetro em quase todas as leis e códigos estudados para guiar o entendimento do que pode ser considerado conteúdo danoso para as democracias (entre *fake news*, declarações falsas, informações enganosas em geral) é um dos principais motivos para a dificuldade de julgar e regular.

O segundo nó crítico mais importante é a responsabilidade dos intermediários. O modelo que isenta as plataformas e aplicativos por conteúdos postados por terceiros têm gerado incentivo para que as empresas não sejam suficientemente diligentes para enfrentar temas graves com discurso de ódio e desinformação. Ao mesmo tempo, não há clareza de o quanto é possível sair desse modelo para outro sem afetar negativamente a liberdade de expressão dos usuários. O caminho adotado pela Alemanha e pelo Reino Unido, em que a responsabilidade da plataforma não se dá sobre cada conteúdo individualmente, mas sobre processos de cuidado e proteção de deveres dos usuários, parece ser o mais promissor.

De toda forma, ainda é cedo para se afirmar com certeza que a lei alemã foi bem-sucedida na busca desse equilíbrio. O white paper do Reino Unido, por sua vez, nem sequer foi transformado em lei. O Reino Unido, na verdade, resolveu começar seus esforços regulatórios por mecanismos anti-concentração, que serão objeto de uma Unidade de Mercado Digital, a partir de abril de 2021.

Não há, até agora, um debate amadurecido e melhores práticas consolidadas. Os consensos que existem se dão em torno de temas óbvios, como transparência ou códigos de conduta, insuficientes para o enfrentamento da desinformação na gravidade que o tema alcançou.

Outro desafio evidente se dá pela tentativa de lidar com "desinformação" ou "fake news" ou "informações enganosas" como um tema único, quando ele, na verdade, é multifacetado. A maior parte dos casos estudados se preocupa em incidir sobre o que pode ser explicitamente nocivo e ilegal, mas a desinformação pode ganhar contornos muito distintos se conectado a processos eleitorais ou a discurso de ódio contra minorias, por exemplo. Regulações, que se estendem para todos os temas, políticos ou não e buscam proteger instituições de governo e símbolos nacionais, são as menos democráticas, como as da Índia e de Singapura.

Os processos de desenvolvimento de normas, diante do desafio colocado, têm se mostrado bastante distintos. Enquanto o Reino Unido realiza consulta pública de documento de definição do cenário, dilemas e desafios - e não da proposta legislativa -, a Alemanha aprovou a NetzDG em poucos meses e o Brasil pode aprovar uma lei para combater *fake news* sem realização de consulta pública ou audiência e durante o período de funcionamento em regime excepcional do Congresso, em decorrência das medidas para conter a pandemia do novo coronavírus.

A análise dos casos revela os desafios intrínsecos ao processo de construir soluções regulatórias que sejam, ao mesmo tempo, protetoras de direitos, mas eficazes na promoção do acesso à informação confiável. Esses desafios não devem ser vistos, contudo, como intransponíveis. É preciso ampliar os debates públicos e os esforços de formulação e apostar em saídas de caráter experimental e inovador, que sejam capazes de dar respostas que impeçam que processos de desinformação sigam afetando a proteção de direitos e o funcionamento de democracias em todo o mundo

REFERÊNCIAS BIBLIOGRÁFICAS

ADAMS, Paul. Geographies of Media and Communication: A Critical Introduction. Wiley-Blackwell, 2009

ALEMANHA. NetzDG. Network Enforcement Act, NetzDG. Bonn: 2017. Disponível em: <https://www.bmjv.de/DE/Themen/FokusThemen/NetzDG/NetzDG_EN_node.html>

ALEMANHA. Declaration on freedom of expression. [s.d.]. Disponível em: <https://deklaration-fuer-meinungsfreiheit.de/en/>. Acesso em: 28 nov. 2020.

ALLEN, J.; FLORES, N. The role of government in the Internet. p. 105, [s.d.]. BfJ - Federal Office of Justice Issues Fine against Facebook. Disponível em: <https://perma.cc/9G3V-SJRN>. Acesso em: 26 nov. 2020.

ARTIGO 19. The Global Expression Report 2019/2020: The state of freedom of expression around the world. Londres, 2020a. Disponível em: <https://artigo19.org/wp-content/blogs.dir/24/files/2020/10/GxR_Final_DigitalVersion_19Oct2020.pdf>. Acesso em: 22 fev. 2020.

ARTIGO 19. France: Analysis of draft hate speech bill. Disponível em: <https://www.article19.org/resources/france--analysis-of-draft-hate-speech-bill/>. Acesso em: 24 nov. 2020.

ARTIGO 19. France: Avia law is threat to online speech. Disponível em: <https://www.article19.org/resources/france-avia-law-is-threat-to-online-speech/>. Acesso em: 24 nov. 2020b.

ARTIGO 19. Análise legal da lei alemã NetzDG [s.d.]. Disponível em: <https://www.article19.org/wp-content/uploads/2017/09/170901-Legal-Analysis-German-NetzDG-Act.pdf>. Acesso em: 6 dez. 2020.

ARTIGO 19. Germany: Responding to 'hate speech'. Disponível em: <https://www.article19.org/resources/germany-responding-to-hate-speech/>. Acesso em: 6 dez. 2020.

ARTIGO 19. GxR_Final_DigitalVersion_19Oct2020.pdf. [s.d.]. Disponível em: <https://artigo19.org/wp-content/blogs.dir/24/files/2020/10/GxR_Final_DigitalVersion_19Oct2020.pdf>. Acesso em: 25 nov. 2020

ASSANGE, Julian, APPELBAUM, Jacob, MÜLLER-MAGUHN, Andy, ZIMMERMANN, Jéréme. Cypherpunks: liberdade e o futuro da internet. São Paulo: Ed. Boitempo, 2013.

BIMBER, Bruce, FLANAGIN, Andrew J, STOHL, Cynthia. Collective Action in Organizations: Interaction and Engagement in an Era of Technological Change. Cambridge University Press, 2012.

BRASIL. Projeto de Lei 2630/2020. Brasília: Senado Federal, 2020. Disponível em: <https://www.camara.leg.br/proposicoesWeb/fichadetramitacao?idProposicao=2256735>

BREEDEN, A. French Court Strikes Down Most of Online Hate Speech Law. The New York Times, 18 jun. 2020.

BUNDESAMT FUR JUSTIZ. Federal Office of Justice Issues Fine against Facebook. Disponível em: <https://www.bundesjustizamt.de/DE/Presse/Archiv/2019/20190702_EN.html;jsessionid=306BFD593DD710232937717A8D07F115.2_cid393?nn=3449818>. Acesso em: 22 fev. 2021.

CAMPOS, 2020. Lei alemã ou movimento global? O debate sobre regulação de redes contextualizado. Consultor Jurídico. Disponível em: <https://www.conjur.com.br/2020-nov-24/direito-digital-lei-alema-ou-movimento-global-contextualizando-debate-regulacao-redes>. Acesso em: 22 fev. 2021.

CANADÁ. Elections Modernization Act (BILL C-76). Otawa: Câmara dos Comuns & Senado, 13 de dezembro de 2018. Disponível em: <https://parl.ca/DocumentViewer/en/42-1/bill/C-76/royal-assent>. Acesso em 22 fev. 2021.

CDT. Overview of the NetzDG Network Enforcement Law. Center for Democracy and Technology, [s.d.]. Disponível em: <https://cdt.org/insights/overview-of-the-netzdg-network-enforcement-law/>. Acesso em: 26 nov. 2020

CEPS. The Impact of the German NetzDG law. CEPS, 2 ago. 2019. Disponível em: <https://www.ceps.eu/ceps-projects/the-impact-of-the-german-netzdg-law/>. Acesso em: 26 nov. 2020.

CHRISTCHURCH CALL. Disponível em em: <https://www.christchurchcall.com/call.html>. Acesso em 22 fev. 2021.

CNCDH. Disponível em: <https://perma.cc/HR2L-GH9Q>. Acesso em: 24 nov. 2020.

COMISSÃO EUROPEIA. Comunicação: Tackling online disinformation: a European Approach. Bruxelas: Comissão Europeia, 2018. Disponível em: <https://ec.europa.eu/digital-single-market/en/news/communication-tackling-online-disinformation-european-approach>.

CONSEIL CONSTITUTIONNEL, 2020. Décision n° 2020-801 DC du 18 juin 2020 - Communiqué de presse. Disponível em: <https://www.conseil-constitutionnel.fr/actualites/communique/decision-n-2020-801-dc-du-18-juin-2020-communique--de-presse>. Acesso em: 22 fev. 2021.

CORTE IDH. OC-5/85. Parecer consultivo de 13-11-1985 sobre o registro profissional obrigatório de jornalistas. San José: 1985.

ECHIKSON, William; KNODT, Olivia. Conter Extremism Project. Germany's NetzDG: A key test for combatting online hate. Disponível em: <https://www.counterextremism.com/sites/default/files/CEP-CEPS_Germany%27s%20NetzDG_020119.pdf>. Acesso em: 26 nov. 2020.

ECHIKSON, W.; KNODT, O. Germany's NetzDG: A Key Test for Combating Online Hate. Rochester, NY: Social Science Research Network, 22 nov. 2018a. Disponível em: <https://papers.ssrn.com/abstract=3300636>. Acesso em: 24 nov. 2020.

ELETRONIC FRONTIER FOUNDATION. Community Input on Christchurch call. 2019. Disponível em: <https://www.eff.org/files/2019/05/16/community_input_on_christchurch_call.pdf>. Acesso em: 24 nov. 2020.

EPRA. Online hate in France - the 'Avia Law': the end of an intensive legislative saga. Disponível em: <https://www.epra.org/news_items/online-hate-in-france-the-law-avia--an-intensive-legislative-saga-for-a-heavily-censored-law>. Acesso em: 24 nov. 2020.

EUROPEAN COMISSION. Law aimed at combating hate content on the internet. Disponível em: <https://ec.europa.eu/growth/tools-databases/tris/en/index.cfm/search/?trisaction=search.detail&year=2019&num=412&mLang=EN>. Acesso em: 24 nov. 2020.

FRANÇA. LOI n° 2018-1202 du 22 décembre 2018 relative à la lutte contre la manipulation de l'information. Paris: 2018. Disponível em: <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000037847559>

FRANÇA. Decisão n ° 2020-801 DC de 18 de junho de 2020 - Comunicado de imprensa do Conselho Constitucional. Disponível em: <https://www.conseil-constitutionnel.fr/actualites/communique/decision-n-2020-801-dc-du--18-juin-2020-communique-de-presse>. Acesso em: 24 nov. 2020.

Germany's balancing act: Fighting online hate while protecting free speech. Disponível em: <https://www.politico.eu/article/germany-hate-speech-internet-netzdg-controversial-legislation/>. Acesso em: 26 nov. 2020.

GLOBAL INTERNET FORUM TO COUNTER TERRORISM. June 2020 | Appointment of Executive Director and Formation of the Independent Advisory Committee. Disponível em: <https://gifct.org/about/story/#june-2020---appointment-of--executive-director-and-formation-of-the-independent-advisory-committee-1>. Acesso em: 22 fev. 2021.

GOLDSMITH, J.; WU, T. Who Controls the Internet? Illusions of a Borderless World.

HOWARD, Philip. N. New Media Campaigns and the Managed Citizen. New York: Cambridge University Press, 2006.

ICYMI: New Report on Germany's NetzDG Online Hate Speech Law Shows No Threat of Over-Blocking. Disponível em: <https://www.counterextremism.com/press/icymi-new-report-germany%E2%80%99s-netzdg-online-hate-speech-law-shows-no-threat-over-blocking>. Acesso em: 26 nov. 2020.

INDIA. Draft of The Information Technology [Intermediary Guidelines (Amendment) Rules] 2018. New Delhi: MEITY, 2018. Disponível em: <https://www.meity.gov.in/writereaddata/files/Draft_Intermediary_Amendment_24122018.pdf>

INTERNET SOCIETY. The Internet and extra territorial application of laws. 2018. Disponível em: <https://www.internetsociety.org/wp-content/uploads/2018/10/The-Internet-and-extra-territorial-application-of-laws.pdf>. Acesso em: 1 dez. 2020

JACOB, Lisa. Dalia Research. 87% of Germans Approve pf Social Media Regularion Law. Disponível em: <https://daliaresearch.com/blog/blog-germans-approve-of-social--media-regulation-law/>. Acesso em : 22 fev. 2021.

LA QUADRATURE DU NET. Loi haine : le Conseil constitutionnel refuse la censure sans juge. Disponível em: <https://www.laquadrature.net/2020/06/18/loi-haine-le-conseil-constitutionnel-refuse-la-censure-sans-juge/>. Acesso em: 24 nov. 2020.

LASSWELL, Harold. Propaganda technique in the world war. Peter Smith: New York, 1938.

LESSIG, Lawrence. Code: version 2.0. New York: Soho Books, 2006.

MARGETTS, Helen; JOHN, Peter; HALE, Scott A.; Yasse Ri, Taha. Political Turbulence: How Social Media Shape Collective Action. Princeton e Oxford: Princeton University Press, 2016.

PARISER, Eli. The filter bubble. New York: Penguin Books, 2011.

POPKIN, Samuel. The Reasoning Voter: Communication and Persuasion in Presidential Campaigns. University of Chicago Press, 1994.

REINO UNIDO. Online Harms White Paper. Disponível em: <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/793360/Online_Harms_White_Paper.pdf>. Acesso em: 22 fev. 2021.

RELEASE, P. Victory! French High Court Rules That Most of Hate Speech Bill Would Undermine Free Expression. Disponível em: <https://www.eff.org/press/releases/victory--french-high-court-rules-most-hate-speech-bill-would-undermine-free-expression>. Acesso em: 24 nov. 2020.

REPORTERES SEM FRONTEIRAS. Classificação Mundial da Liberdade de Imprensa 2020. Disponível em: <https://rsf.org/pt/classificacao%20>. Acesso em: 25 nov. 2020.

SCHULZ, Jacob. What's Going on With France's Online Hate Speech Law? Disponível em: <https://www.lawfareblog.com/whats-going-frances-online-hate-speech-law>. Acesso em: 24 nov. 2020.

STOLTON, S. EU Commission to introduce sanctions regime for illegal content in Digital Services Actwww.euractiv.com, 4 nov. 2020. Disponível em: <https://www.euractiv.com/section/digital/news/eu-commission-to-introduce-sanctions-regime-for-illegal-content-in-digital-services-act/>. Acesso em: 24 nov. 2020

TAYLOR WESSING. New law to fight online hate speech (Avia law) to reshape notice, take down and liability rules in France. Disponível em: <https://www.taylorwessing.com/en/insights-and-events/insights/2020/05/new-law-to-fight-online-hate-speech-in-france>. Acesso em: 24 nov. 2020.

TWOREK, H.; LEERSSEN, P. Transatlantic Working Group. An Analysis of Germany's NetzDG Law. Disponível em: <https://www.ivir.nl/publicaties/download/NetzDG_Tworek_Leerssen_April_2019.pdf>. Acesso em: 22 fev. 2021.

UNIÃO EUROPEIA. Código de Prática sobre Desinformação. Bruxelas: UE, 2018. Disponível em: <https://ec.europa.eu/digital-single-market/en/news/code-practice-disinformation>. Acesso em: 21 fev. 2021.

UNIVERSAL RIGHTS GROUP. France's watered-down anti-hate speech law enters into force. Universal Rights Group, 16 jul. 2020. Disponível em: <https://www.universal-rights.org/blog/frances-watered-down-anti-hate-speech-law-enters-into-force/>. Acesso em: 24 nov. 2020

ZIPURSKY, R. Nuts About NETZ: The Network Enforcement Act and Freedom of Expression. p. 51, [s.d.].

AUTORES

João Brant é pesquisador e consultor em políticas de comunicação, doutor em Ciência Política (USP).

João Guilherme Bastos dos Santos é pesquisador no Instituto Nacional de Ciência e Tecnologia em Democracia Digital (INCT.DD), doutor em Comunicação (UERJ).

Tatiana Dourado é pesquisadora na Diretoria de Análise de Políticas Públicas da FGV e membro do INCT.DD, doutora em Comunicação (UFBA).

Marina Pita é pós-graduanda em Direito Digital (UERJ) e mestranda em Comunicação (UnB).

FICHA TÉCNICA

Friedrich-Ebert-Stiftung (FES) Brasil
Av. Paulista, 2001 - 13° andar, conj. 1313
01311-931 • São Paulo • SP • Brasil

Responsáveis:
Christoph Heuser, representante da FES no Brasil
Gonzalo Berrón, diretor de programas
https://brasil.fes.de

Contato:
fesbrasil@fes.org.br

ISBN 978-65-87504-16-2

REGULAÇÃO DE COMBATE À DESINFORMAÇÃO

Estudo de oito casos internacionais e recomendações para uma abordagem democrática

A prática da desinformação foi definida pela UNESCO (Organização das Nações Unidas para a Educação, a Ciência e à Cultura) e pelo PNUD (Programa das Nações Unidas para o Desenvolvimento) como "conteúdo falso, manipulado ou enganoso, criado e disseminado intencionalmente ou não, e que pode causar danos potenciais à paz, aos direitos humanos e ao desenvolvimento sustentável". De fato, seus efeitos são sentidos em muitos campos: em campanhas eleitorais, nos temas de saúde pública (como ficou evidente na pandemia de Covid-19), na disseminação de discurso de ódio contra grupos sociais ou de ataque à reputação de ativistas, e em todas as disputas relevantes no campo socioambiental, apenas para citar os exemplos mais evidentes.

Os oito casos analisados possuem diferentes objetos de regulação. Os textos que mais restringem a livre expressão costumam ter como foco a ideia de informação, declaração ou fato falso. Nenhum deles, porém, define critérios públicos de moderação de conteúdo, o que fica ou a cargo das empresas de tecnologia ou a cargo direto dos governos. Os sistemas analisados combinam autorregulação, corregulação e regulação pública, mas não há, até agora, um debate amadurecido e melhores práticas consolidadas. Os consensos que existem se dão em torno de temas óbvios, como transparência ou códigos de conduta, insuficientes para o enfrentamento da desinformação na gravidade que o tema alcançou.

Embora a proteção das plataformas contra a responsabilização de conteúdo por terceiros seja importante para evitar o efeito silenciador, a questão da responsabilização dos indivíduos por conteúdo disseminado na rede está longe de ter encontrado um ponto de equilíbrio. Diversas análises críticas acerca da resposta a conteúdos de ódio, por exemplo, apontam para a baixa capacidade dos estados e dos sistemas de Justiça punir os perpetradores de crimes relacionados a calúnia, injúria, difamação, racismo, homofobia, transfobia, ameaça etc. Nesse sentido, segue relevante a busca de modelos que consigam proteger a liberdade de expressão ao mesmo tempo que ofereçam mecanismos eficazes para defender os demais direitos humanos com os quais ela pode estar em colisão.

Para mais informações sobre o tema, acesse:
https://brasil.fes.de

FRIEDRICH EBERT STIFTUNG