Padrões de Direitos Humanos como Linhas de Base para a Regulação e Prestação de Contas das Plataformas

UMA CONTRIBUIÇÃO PARA O DEBATE BRASILEIRO

ELECTRONIC FRONTIER FOUNDATION

accessnow

Há pelo menos uma década, instituições de direitos humanos têm reconhecido o potencial da Internet para concretizar uma série de direitos humanos. As tecnologias digitais se mostraram ferramentas imensamente transformadoras para permitir às pessoas se manifestarem contra atos arbitrários de poderes públicos e privados, empoderando a expressão de grupos historicamente vulneráveis, marginalizados e silenciados, catalisando a organização e a participação cívica e facilitando formas inovadoras de construir e compartilhar conhecimento coletivamente. Desde então, o direito de buscar, receber e difundir informações tem possibilitado o exercício de outros direitos e fortalecido o ecossistema da Internet, mas não sem retrocessos e desafios críticos.

A discussão atual sobre a regulação de plataformas no Brasil, com destaque ao projeto de lei 2630 (PL 2630) e os casos constitucionais pendentes no Supremo Tribunal Federal (STF), demonstra que muito se está buscando fazer para enfrentar esses desafios, mas mostra também que elaborar respostas adequadas não é tarefa trivial. Devemos ter a capacidade de modelar essas respostas, protegendo o potencial positivo das tecnologias digitais e o papel essencial que a liberdade de expressão, incluindo o acesso à informação, desempenha na preservação de sociedades democráticas.

Um breve histórico

O PL 2630, também conhecido como "PL das Fake News", foi proposto no Senado brasileiro em 2020. A pressão de organizações da sociedade civil, com destaque à *Coalizão Direitos na Rede,* para melhorar o texto, bem como o trabalho delas em conjunto com o relator do projeto de lei na Câmara dos Deputados, foram fundamentais para neutralizar ameaças como a obrigação de rastreabilidade de mensagens criptografadas de ponta-a-ponta. Até então, os grupos brasileiros de direitos digitais também haviam enfatizado que a regulação deveria se concentrar em processos de moderação de conteúdo (por exemplo, transparência e regras de devido processo) em vez de restrição de certos tipos de conteúdo. Após a divulgação de um novo texto preliminar no início de 2022, o projeto de lei permaneceu parado na Câmara dos Deputados até o início de 2023.

Na esteira da tentativa fracassada da extrema direita no início deste ano de derrubar o novo governo do presidente Lula da Silva e de um pico de ataques violentos em escolas brasileiras, o PL 2630 se consolidou como o caminho legislativo para enfrentar preocupações mais abrangentes sobre o uso de tecnologias digitais em contextos de agitação social. Para isso, o Poder Executivo propôs ao relator do PL um novo texto que introduziu várias mudanças, considerando leis como a alemã *NetzDG*, o *Digital Services Act* (Lei de Serviços Digitais - DSA) da União Europeia (UE) e iniciativas de lei como o controverso *Online Safety Bill* do Reino Unido.

A versão mais recente publicada do PL incorpora algumas dessas propostas, como regras de avaliação de risco, obrigações de dever de cuidado e novas exceções à regra geral de responsabilidade de intermediários de internet em vigor no Brasil. De acordo com o artigo 2º do projeto de lei, suas regras se aplicam às redes sociais, mecanismos de busca e serviços de mensageria instantânea constituídos na forma de pessoa jurídica e com

mais de dez milhões de usuários mensais no Brasil. Embora a DSA seja frequentemente mencionada como uma inspiração e precedente democrático que fundamenta a nova proposta, o texto reformulado do projeto de lei apresenta diferenças importantes com a lei europeia e ainda falha em garantir freios e contrapesos suficientes considerando o contexto e o marco institucional brasileiros.

Em paralelo, o STF tem casos pendentes sobre responsabilidade de intermediários de internet (repercussão geral 533 e 987) e sobre bloqueios de *sites* e aplicações *online* ordenados por autoridades judiciais (ADI 5527 e ADPF 403). Atualmente, o regime geral de responsabilidade de intermediários no Brasil é estabelecido pelo artigo 19 do Marco Civil da Internet (Lei nº 12.965/2014). De acordo com o artigo 19, as aplicações de internet podem ser responsabilizadas por conteúdos de seus usuários apenas quando deixam de cumprir uma decisão judicial ordenando a remoção de conteúdo infringente. Há exceções em que uma notificação extrajudicial pode tornar as plataformas responsáveis por conteúdo de terceiros, como violação de direitos autorais, divulgação não autorizada de imagens privadas que contenham nudez ou atividade sexual e conteúdo que envolva abuso sexual infantil.

Alguns ministros do STF expressaram sua opinião de que o regime geral do Marco Civil precisa de uma atualização para endurecer as regras de responsabilidade de intermediários, e os casos constitucionais pendentes podem ser uma maneira de fazê-lo se o Congresso demorar para lidar com a questão. O papel cada vez mais poderoso das principais aplicações de internet tem motivado debates e iniciativas para revisar os atuais regimes de responsabilidade de intermediários em diferentes países. No entanto, há perguntas-chave a fazer, ferramentas a considerar e lições aprendidas a ter como base antes de se introduzir mudanças que possam afetar seriamente expressões protegidas e a capacidade das pessoas de fortalecerem suas vozes e direitos fazendo uso de tecnologias digitais.

Por sua vez, a decisão do STF sobre bloqueios de *sites* e aplicações *online* foi interrompida desde 2020, quando o Ministro Alexandre de Moraes pediu vistas do processo, devolvendo-o apenas em março deste ano. Esses casos se referem aos bloqueios do WhatsApp no Brasil em 2015 e 2016, envolvendo a questão de se autoridades poderiam exigir que um provedor de aplicações comprometesse suas implementações de privacidade e segurança por padrão, ou seja, a criptografia de ponta-a-ponta, para divulgar dados de comunicações do usuário no âmbito de uma investigação criminal. O julgamento dos casos começou em 2020 com votos importantes dos Ministros Edson Fachin e Rosa Weber apoiando as proteções de privacidade e segurança incorporadas na arquitetura dos sistemas digitais e rejeitando a interpretação da lei brasileira para permitir a determinação judicial de bloqueios com o objetivo de comprometer tais proteções. Infelizmente, os possíveis resultados de retomar esse julgamento no contexto atual são imprevisíveis. Seguindo seu papel pioneiro de reconhecer a proteção de dados pessoais como um direito fundamental na Constituição Federal brasileira, é essencial que o STF endosse os votos dos Ministros Rosa Weber e Edson Fachin em favor de proteções robustas de privacidade e segurança por padrão.

Apesar das movimentações do Poder Executivo e do STF para mudanças no atual marco regulatório brasileiro, os atores políticos concordaram, pelo menos por enquanto, que o

Congresso é o local adequado para um debate democrático sobre regulação de plataformas. Nós também concordamos. É pertinente, assim, analisar o projeto de lei em discussão à luz do debate atualmente em curso no país. Embora o PL contenha elementos positivos, devemos destacar pontos ainda a serem melhorados.

Pontos de preocupação importantes

O PL 2630 pretende fortalecer os direitos de usuárias e usuários frente ao poder de grandes aplicações de internet, como Facebook, Youtube e Twitter. No entanto, há pontos de preocupação cruciais que o debate sobre regulação de plataformas e o PL 2630 devem considerar atentamente. Outros grupos na região, como Derechos Digitales, levantaram pontos de atenção. Como passamos a elaborar neste texto, há um conjunto de questões que as partes interessadas devem examinar e encaminhar antes de aprovar uma nova lei. Os mais relevantes são:

- Neutralizar os riscos de abuso de regulações baseadas em conteúdo, abandonando as obrigações de dever de cuidado, privilegiando avaliações de impacto sistêmico e explicitando que a atuação diligente das plataformas não significa monitoramento e filtragem geral de conteúdos dos usuários.
- Garantir freios, contrapesos e garantias de devido processo robustos para a aplicação de regras específicas a situações de conflito e risco iminente.
- Conceber de forma criteriosa e garantir os meios apropriados para estabelecer uma estrutura adequada de supervisão independente, autônoma, participativa e multissetorial para a regulação em debate.
- Estabelecer garantias claras contra o aumento da vigilância e os riscos de segurança relacionados.
- Abster-se de conceder proteções especiais a declarações de autoridades estatais, que têm responsabilidades especiais conforme padrões internacionais de direitos humanos.
- Garantir sanções de acordo com padrões de direitos humanos e garantias de devido processo legal, particularmente quando isso envolver o bloqueio de aplicações *online*.

O último ponto se refere às sanções administrativas que podem ser aplicadas caso os provedores de aplicação sujeitos ao projeto de lei descumpram as suas regras. A "suspensão temporária das atividades" está entre as sanções possíveis. Na prática, isso significa que uma autoridade governamental administrativa teria o poder de bloquear um *site* ou aplicativo por inteiro. De forma geral, o bloqueio de *websites* no Brasil acontece após uma ordem judicial, embora o Ministério da Justiça tenha afirmado recentemente que órgãos administrativos de proteção do consumidor teriam esses poderes de acordo com as penas de suspensão tradicionais definidas no Código de Defesa do Consumidor. Padrões internacionais de direitos humanos apontam que o bloqueio de *sites* e aplicativos inteiros é uma medida extrema com desafios técnicos, grandes riscos de abuso e impactos significativos em direitos fundamentais. Em 2021, o Conselho de Direitos Humanos da ONU reiterou a adoção de uma resolução condenando inequivocamente o recurso à interrupção do acesso à Internet e medidas de censura

online, o que inclui o bloqueio a mídias sociais, para arbitrariamente impedir ou prejudicar o acesso ou a disseminação de informações *online*. Destacamos anteriormente tais preocupações no contexto do PL 2630. Enquanto nas versões anteriores do PL apenas a maioria absoluta de um órgão judicial colegiado poderia aplicar tal sanção de bloqueio, a proposta atual dá esse poder a uma autoridade administrativa não especificada. Os legisladores brasileiros devem reconhecer os perigos do uso arbitrário de bloqueios *online* e recuar.

Além disso, a aplicação legítima de possíveis sanções se relaciona diretamente ao conjunto de regras e à estrutura de supervisão do projeto de lei. Os outros pontos de preocupação que mencionamos acima destacam lacunas remanescentes relevantes nesta frente. Eles serão discutidos na próxima seção.

De 2011 a 2023: Lidar com os desafios atuais a partir de princípios e garantias existentes

Desde a declaração conjunta de 2011 sobre Liberdade de Expressão e Internet dos Relatores Especiais para a Liberdade de Expressão, as instituições de direitos humanos vêm ressaltando que as iniciativas governamentais que buscam regular as comunicações *online* devem preservar e se adaptar às características únicas da Internet. Isso porque tais iniciativas devem ao mesmo tempo ser eficazes e respeitar as características da Internet que potencializam o exercício de direitos e liberdades fundamentais. Quaisquer restrições devem seguir o "teste de três partes", ou seja, devem ser claramente estabelecidas por lei, estritamente necessárias e proporcionais para alcançar um objetivo legítimo em uma sociedade democrática. Preocupações importantes em torno da fragmentação da Internet, censura colateral, remoção excessiva de expressão legítima e, mais recentemente, complexidades inerentes à moderação de conteúdo em escala, levaram especialistas ao longo dos anos a evitar regulações específicas de conteúdo. Os riscos de aplicação e interpretação arbitrárias de regras que restringem conteúdos em contextos não democráticos ou conflituosos adicionam outras camadas a esse conjunto de preocupações.

A seguir, detalhamos os nossos demais pontos de atenção já apresentados.

Obrigações de dever de cuidado preocupantes

A evolução das versões do PL 2630 foi uma expressão da opção por uma abordagem baseada em processos, em vez de uma focada em conteúdo, no âmbito de uma iniciativa de regulação com o objetivo de promover maiores compromissos das plataformas *online*.

No entanto, após alterações no início deste ano, o projeto de lei agora contém uma lista de práticas ilícitas, ligadas a conteúdos ilícitos, que as aplicações de internet "devem atuar diligentemente para prevenir e mitigar (...), envidando esforços para aprimorar o combate à disseminação de conteúdos ilegais gerados por terceiros". Tal previsão diz respeito às obrigações de *dever de cuidado*, que o projeto de lei não define, porém, ainda assim, operacionaliza sua aplicação. A lista de tais práticas ilícitas, prevista no artigo 11, aponta para disposições em seis leis diferentes que abarcam cerca de 40 infrações penais – cada uma contendo um conjunto de elementos que devem estar presentes para que a conduta seja ilegal. Algumas infrações também têm causas que excluem certas condutas de serem a base de um crime. Por exemplo, tanto a Lei Antiterrorismo (Lei nº 13.260/2016) quanto os crimes contra o Estado Democrático de Direito estabelecidos no Código Penal não se aplicam a manifestações políticas críticas baseadas em direitos constitucionais. De acordo com o artigo 11 do PL, caberia à aplicação de internet considerar todos esses elementos e avaliar se a conduta ou o conteúdo visível através de suas plataformas constituem uma atividade criminosa.

Em alguns casos, é ainda mais difícil entender o que exatamente o provedor de aplicações deve verificar, ou se é algo que ele realmente deve verificar, apesar de sua inclusão na lista de infrações penais do artigo 11. Por exemplo, o artigo 11 se refere genericamente aos crimes contra crianças e adolescentes da Lei nº 8.069/1990. Entre esses crimes está a falha do médico, enfermeiro ou dirigente de estabelecimento de saúde em identificar corretamente o recém-nascido e a mãe parturiente no momento do parto (artigo 229 da Lei nº 8069/1990). Qual é o dever de cuidado esperado das plataformas de internet aqui? Esta regra é um exemplo de uma disposição abrangida pelo artigo 11 que não parece ter qualquer relação com as plataformas *online*. O artigo 11 também não é muito claro sobre como e quais instituições avaliarão o cumprimento das obrigações de dever de cuidado por parte das aplicações de internet. Ele afirma que a avaliação de cumprimento não focará em casos isolados e incluirá informações que as aplicações de internet fornecerão às autoridades sobre seus esforços para prevenir e mitigar as práticas listadas, bem como a análise dos relatórios da plataforma e como respondem a notificações e reclamações.

Dentro do mesmo PL, o artigo 45 estipula que "quando o provedor tomar conhecimento de informações que levantem suspeitas de que ocorreu ou que possa ocorrer um crime que envolva ameaça à vida, ele deverá informar imediatamente da sua suspeita às autoridades competentes". Embora um crime envolvendo uma ameaça à vida seja definitivamente uma emergência e uma situação terrível, o artigo 45 estabelece um novo papel de policiamento para aplicações de internet que, mesmo dentro desse escopo estrito, podem dar margem a resultados controversos, potencialmente afetando, por exemplo, mulheres no Brasil que buscam informações *online* sobre aborto seguro.

As obrigações de dever de cuidado estabelecidas no PL 2630 se sustentam em uma abordagem regulatória que reforça as plataformas digitais como pontos de controle sobre a expressão e as ações *online* das pessoas. Elas exigem que as aplicações de internet ajam como juízes quanto à legalidade de atos ou conteúdos com base em uma lista de delitos criminais complexos, como se fosse simples programar ferramentas e processos de moderação de conteúdo para reconhecer cada elemento que constitui cada delito. Pelo contrário, estas análises são com frequência desafiadoras até mesmo para juízes e

tribunais. Em muitos casos, pessoas divulgam conteúdo sensível precisamente para denunciar a violência institucional, as violações de direitos humanos e a perpetração de crimes em situações de conflito. O compartilhamento de vídeos em redes sociais que expõem casos de discriminação contribui para responsabilizar os ofensores. Durante a onda de protestos no Chile, as plataformas de internet restringiram indevidamente conteúdo que denunciava a dura repressão policial às manifestações, por o terem considerado como conteúdo violento. No Brasil, vimos preocupações semelhantes, por exemplo, quando o Instagram censurou imagens do massacre da comunidade do Jacarezinho em 2021, que foi a operação policial mais letal na história do Rio de Janeiro. Em outras geografias, a missão de restringir o conteúdo extremista já removeu vídeos que documentavam violações de direitos humanos em contextos de conflito em países como Síria e Ucrânia.

Como a Relatoria Especial para a Liberdade de Expressão da Comissão Interamericana de Direitos Humanos (CIDH) destacou, enquanto atores privados, as aplicações de internet *"não têm a capacidade de ponderar direitos e interpretar a lei em conformidade com os padrões em matéria de liberdade de expressão e outros direitos humanos"*, particularmente quando deixar de restringir conteúdos específicos pode ocasionar sanções administrativas ou responsabilidade legal.

Não é que as aplicações de internet não devam fazer esforços para evitar a prevalência de conteúdo pernicioso em suas plataformas, ou que não queremos que elas façam um trabalho melhor ao lidar com conteúdo capaz de causar sérios danos coletivos. Concordamos que elas podem fazer melhor, especialmente por meio da consideração da cultura e realidades locais. Também concordamos que suas políticas devem se alinhar a padrões internacionais de direitos humanos e que devem considerar os impactos potenciais de suas decisões em direitos humanos, de forma a prevenir e mitigar possíveis danos.

No entanto, não devemos misturar a garantia desses compromissos com o reforço das plataformas digitais como pontos de controle sobre a expressão e as ações *online* das pessoas. Este é um caminho perigoso considerando o poder que já está nas mãos das grandes plataformas e a crescente intermediação de tecnologias digitais em tudo o que fazemos. A abordagem do artigo 11 também é problemática na medida em que estabelece esse controle com base em uma *lista* de práticas potencialmente ilegais que a correlação de forças política pode mudar e expandir a qualquer momento ou levar a uma aplicação oportunista ou abusiva para restringir o acesso à informação e silenciar críticas ou vozes dissidentes.

Pelo contrário, compromissos de maior diligência e prestação de contas pelas plataformas priorizam uma abordagem sistêmica e baseada em processos pela qual o provedor de aplicações avalia e elabora respostas para prevenir e mitigar os impactos negativos de suas atividades aos direitos humanos. Isso é consistente com os Princípios Orientadores da ONU sobre Empresas e Direitos Humanos. O próprio PL 2630 contém disposições sobre avaliação de risco sistêmico e medidas de mitigação relacionadas às atividades das empresas. Os legisladores brasileiros devem priorizar essa abordagem em detrimento das obrigações relativas ao "dever de cuidado".

Além disso, o conceito de dever de cuidado, como vemos atualmente no debate brasileiro, apresenta um outro risco. Ele pode ensejar interpretações de que as aplicações de internet devem realizar um monitoramento geral do conteúdo de terceiros que elas hospedam. Tais interpretações não são explicitamente negadas no texto do PL 2630, como são, por exemplo, na DSA da UE.

Repelir regras e interpretações que possam levar a obrigações de monitoramento de conteúdo

Os Relatores Especiais para a Liberdade de Expressão afirmaram também: *"No mínimo, não se deve exigir que os intermediários controlem o conteúdo gerado por usuários."* E que: *"Os sistemas de filtragem de conteúdo que sejam impostos por um governo e não sejam controlados pelo usuário final não representam uma restrição justificada à liberdade de expressão."*

Há pelo menos duas razões principais pelas quais as obrigações gerais de controle de conteúdo não são uma boa ideia. Em primeiro lugar, tais obrigações são talvez a expressão máxima do tratamento de aplicações de internet como uma força de policiamento de tudo o que fazemos e dizemos *online*, com consequências nocivas para a liberdade de expressão e acesso à informação, e infringindo expectativas de privacidade. Se as práticas comerciais de aplicações de internet frequentemente geram preocupações semelhantes, a resistência da sociedade à vigilância corporativa impulsionou regulações de privacidade e proteção de dados, bem como mudanças nas políticas das empresas em favor da privacidade de usuárias e usuários. Em segundo lugar, o controle geral e a filtragem de conteúdos relacionada falham constantemente, e o fato de ter um desempenho deficiente causa ainda mais preocupações para direitos humanos. Dado o grande volume de novos conteúdos que as pessoas postam e compartilham em plataformas *online* a cada minuto, a moderação de conteúdo depende cada vez mais de ferramentas automatizadas, refletindo suas limitações e falhas. Regulações ou interpretações que obrigam a adoção dessas ferramentas e vinculam tal obrigação a sanções ou responsabilização de aplicações de internet ampliam o potencial de erros e de aplicação problemática da lei.

Apenas em termos de probabilidade, quando um sistema que já é propenso a cometer erros é ampliado em escala para moderar conteúdos que são gerados em uma taxa de muitos milhões a bilhões de entradas por dia, mais erros ocorrerão. E quando os modelos de aprendizagem são empregados para educar a inteligência artificial (IA) dentro desses métodos, são poucas as chances de esses modelos reconhecerem e corrigirem esses erros. Na maioria das vezes, essas tecnologias reproduzem discriminação e vieses. São propensas a censurar conteúdo lícito, não ofensivo e relevante. Embora defendamos e continuaremos a defender a análise humana em processos de moderação de conteúdo, ter moderadores humanos suficientes trabalhando em condições adequadas para evitar restrições indevidas de conteúdo será um desafio contínuo.

Os sistemas de IA geralmente empregados na moderação de conteúdo incluem algoritmos de reconhecimento de imagem e modelos de processamento de linguagem natural. Quanto às complexidades do treinamento de modelos de linguagem de IA, os especialistas ressaltam que a linguagem depende muito de contextos culturais e sociais e varia consideravelmente entre grupos demográficos, temas de conversa e tipos de plataformas. Além disso, o treinamento de algoritmos de processamento de linguagem exige definições claras e precisas do conteúdo alvo, o que é muito difícil de alcançar com termos complexos normalmente implicados na caracterização de uma prática criminosa ou ilícita. Mesmo que, no geral, consideremos que o estágio atual das ferramentas de processamento de linguagem natural disponíveis mostra um desempenho eficaz em inglês, elas apresentam variações significativas em termos de qualidade e precisão para outros idiomas. Elas também podem reproduzir discriminação nos dados, afetando desproporcionalmente comunidades marginalizadas, como pessoas LGBTQIA+ e mulheres. Modelos de linguagem multilíngue também têm suas limitações, pois podem não refletir bem a linguagem do cotidiano usada por falantes nativos e não levar em conta contextos específicos.

Por sua vez, apesar dos avanços atuais na tecnologia, as ferramentas de reconhecimento de imagem também têm suas limitações. Um bom exemplo está relacionado ao reconhecimento de imagens sexuais. Uma vez que a fronteira exata em relação a imagens sexuais ofensivas e não ofensivas é objeto de discordância, a tendência natural dos sistemas que construímos para reconhecê-las automaticamente e removê-las das plataformas *online* estará alinhada às estimativas mais conservadoras para minimizar os riscos legais. Isso significa que a expressão que é de outra forma protegida, legal e, muitas vezes proveniente de minorias sexuais, será considerada inadequada. Um caso marcante de censura *online* privada no Brasil reflete precisamente esse problema. Em 2015, o Facebook bloqueou uma foto do início do século XX de um casal indígena parcialmente vestido, postada pelo Ministério da Cultura para divulgar o lançamento do acervo digital *Portal Brasiliana Fotográfica* logo antes do Dia dos Povos Indígenas no Brasil.

Da mesma forma, e à medida que nos aproximamos de sistemas sofisticados de IA capazes de determinar com precisão imagens sexuais de outros materiais, nos deparamos com o antigo problema da arte *versus* pornografia. A arte clássica que retrata a forma nua continua a ser sinalizada como imprópria por algoritmos de moderação, apesar do consenso esmagador de que ela está firmemente na categoria "arte", e não na qualificação como ilegal ou contrária aos padrões da comunidade. A arte contemporânea confunde ainda mais esses limites, muitas vezes intencionalmente. Nossa capacidade de expressão como seres humanos está em constante mudança, o que continuará a ser um desafio para os desenvolvedores de sistemas de computadores construídos para reconhecer e categorizar o conteúdo gerado por pessoas, o que, em escala, produzirá ainda mais erros.

Uma taxa considerável de erros também pode acontecer em sistemas de reconhecimento de imagem baseados em hashes. Erros comuns enfrentados por esse tipo de tecnologia, como as chamadas "colisões", ocorrem porque duas imagens diferentes podem ter o mesmo valor hash, criando falsos positivos, onde uma imagem é identificada incorretamente como algo que não é. Isso pode ocorrer por vários motivos, por exemplo,

se as imagens forem muito semelhantes, se a função hash não é muito boa em distinguir entre imagens diferentes ou se a imagem foi corrompida ou manipulada. O oposto também pode ocorrer, ou seja, manipular imagens infratoras para que a função hash não as reconheça e sinalize. Além das questões de eficiência, esses sistemas comprometem as proteções inscritas na arquitetura de plataformas digitais que, por padrão, garantem a inviolabilidade das comunicações, privacidade, segurança e proteção de dados, como é o caso da criptografia de ponta-a-ponta.

Quando os sistemas de moderação são dimensionados para tamanhos desproporcionalmente grandes, o alcance de obrigações de monitoramento e denúncia anexadas a eles, se existentes, é dimensionado da mesma maneira. Isso pode ser e tem sido moldado como os olhos e ouvidos de forças arbitrárias e não democráticas.

A regulação de plataformas não deve incentivar interpretações ou regulamentação adicional que exijam o controle geral e filtragem de conteúdo. O PL 2630 deve ser mais explícito para repelir tais interpretações, e o debate regulatório no Brasil sobre compromissos de diligência e prestação de contas das aplicações de internet deve rejeitar essas obrigações por não serem respostas necessárias e proporcionais.

Freios, contrapesos e garantias de devido processo robustos para a aplicação de medidas excepcionais em situações de crise

O PL 2630 estabelece obrigações especiais para quando há um risco iminente de dano ou negligência de um provedor de aplicações (Artigos 12-15). Ao avaliar esta seção do projeto de lei, é fundamental recordar a Declaração Conjunta de 2015 sobre situações de crise. Entre outras recomendações, ela destaca que os "*[e]stados não devem responder a situações de crise com a adoção de restrições adicionais à liberdade de expressão, salvo o estritamente justificado pela situação e pelas leis internacionais de direitos humanos. [...] Medidas administrativas que restrinjam a liberdade de expressão deveriam ser impostas unicamente quando justificadas em virtude do teste de três partes para tais restrições.*"

A intenção desta seção do projeto de lei é ser o fundamento jurídico para restringir as liberdades fundamentais durante situações de crise. Porém, sua redação atual não contém precisão e clareza suficientes, bem como freios e contrapesos adequados para fundamentar uma intervenção necessária e proporcional.

De acordo com o PL 2630, a decisão de implementação do protocolo de segurança especificará, entre outros, os provedores impactados, o prazo do protocolo (até 30 dias, que pode ser prorrogado) e uma lista de quesitos relevantes que devem ser abordados pelos provedores por meio de medidas de mitigação eficazes e proporcionais durante o período do protocolo. Enquanto o protocolo vigorar e para os tipos de conteúdo especificados na decisão de implementação, os provedores afetados estão sujeitos à

responsabilidade solidária pelo conteúdo gerado pelo usuário, desde que os provedores tenham conhecimento prévio de tal conteúdo. Uma simples notificação do usuário, utilizando o mecanismo de notificação que o artigo 16 exige que as aplicações de internet forneçam, é suficiente para configurar esse conhecimento prévio. O projeto de lei, portanto, cria um mecanismo excepcional de notificação e retirada a ser aplicado enquanto o protocolo vigorar e relacionado a certos tipos de conteúdo (conforme a "delimitação temática" do protocolo).

Mecanismos de notificação e retirada causam muitas preocupações, pois podem alimentar a utilização de sistemas de notificação como arma para censurar reportagens críticas, críticas políticas e vozes de grupos marginalizados. Com muita frequência levam a remoções excessivas. A Relatoria Especial da CIDH para a Liberdade de Expressão observou que eles criam incentivos para a censura privada, pois colocam "*os intermediários privados em posição de ter que tomar decisões sobre a licitude ou ilicitude*" dos conteúdos gerados por usuárias e usuários. Tais intermediários não vão *"necessariamente considerar o valor da liberdade de expressão ao tomar decisões sobre conteúdos produzidos por terceiros que possam comprometer sua responsabilidade"*. A própria experiência brasileira nos tribunais mostra como a questão pode ser complicada. Pesquisa do InternetLab baseada em decisões judiciais envolvendo liberdade de expressão *online*, divulgada cinco anos após a aprovação do Marco Civil, mostrou que os tribunais de apelação brasileiros negaram pedidos de remoção de conteúdo em mais de 60% dos casos. Na audiência pública que o STF realizou para receber contribuições sobre seus casos envolvendo responsabilidade de intermediários, a Associação Brasileira de Jornalismo Investigativo (ABRAJI) apresentou dados sobre solicitações de remoção apresentadas judicialmente entre 2014 a 2022. De acordo com a ABRAJI, em algum momento do processo judicial, os juízes concordaram com os pedidos de remoção de conteúdo em cerca de metade dos casos, sendo que alguns destes foram revertidos posteriormente.

No entanto, o mecanismo de notificação e retirada do PL 2630 ligado a um protocolo de segurança parece desempenhar um papel moderador em meio a uma pressão crescente do Poder Executivo e do STF para expandir as exceções à regra geral do Marco Civil sobre responsabilidade de intermediários. O fato de que esse mecanismo seria limitado no tempo e no escopo poderia ajudar com algumas das preocupações acima, assim como a aplicação das regras do artigo 18, que incluem o direito dos usuários de recorrer de decisões de moderação de conteúdo. Porém, a dinâmica geral do protocolo de segurança ainda apresenta sérios problemas. Uma preocupação primordial é que as situações de crise não se tornem permanentes, por meio da extensão da duração ou reiterando a ocorrência de medidas que, por definição, são restritas a circunstâncias excepcionais. São necessários controles claros e eficazes para que uma disciplina legal para situações de crise não se transforme na regulação por padrão.

Aqui estão as principais questões e possíveis mitigações que os legisladores brasileiros devem considerar:

- O artigo 12 define situação de crise de uma forma extremamente ampla. A iminência dos riscos estabelecidos no artigo 7º, que inclui uma série de temas (por exemplo, a disseminação dos conteúdos ilícitos listados no artigo 11 e riscos

à liberdade de expressão, saúde pública e ao Estado Democrático de Direito), **ou** a "negligência ou insuficiência da ação do provedor" é suficiente para desencadear a implementação do protocolo de segurança. Os critérios para tipificar o que constitui tal insuficiência ou negligência dependem de regulamentação que ainda não existe. No entanto, o artigo 12 não relaciona a ação negligente do provedor aos riscos estabelecidos no artigo 7º. Uma insuficiência ou negligência da aplicação de internet relacionada com *qualquer* questão **ou** um risco iminente estabelecido no artigo 7º é suficiente para configurar uma situação de crise. Isso também significa que, mesmo que os provedores estejam tomando medidas importantes de boa-fé para lidar com os riscos iminentes do artigo 7º, eles ainda podem estar sujeitos às medidas excepcionais do protocolo de segurança. No mínimo, **a disposição deve combinar ambos os requisitos, usando "*e*" em vez de "*ou*" em sua redação**. Há, porém, outras preocupações fundamentais quanto ao protocolo de segurança.

- Versões anteriores do projeto de lei qualificavam a situação de risco iminente do protocolo. Havia uma referência a "riscos iminentes de danos à dimensão coletiva dos direitos fundamentais". Este é um qualificador importante, especialmente porque o artigo 7º ainda é bastante amplo nos riscos que enumera. Embora sua lista possa funcionar para orientar as avaliações de impacto de grandes provedores, ela traz preocupações sobre possíveis interpretações abusivas e usos oportunistas no contexto de um protocolo de segurança que define obrigações excepcionais para aplicações de internet. Portanto, deve haver **um risco de dano à dimensão coletiva dos direitos fundamentais** para permitir que uma autoridade implemente esse protocolo de segurança. Além disso, o projeto de lei deve ser explícito em estabelecer que a avaliação da autoridade deve seguir padrões estritamente necessários e proporcionais ao tomar tal decisão.
- O projeto de lei é omisso sobre qual autoridade tem o poder de declarar uma situação de crise e estabelecer os termos do protocolo de segurança. Tratamos do desenho de supervisão do projeto de lei no próximo item, e o fato de que o PL atualmente carece de uma estrutura de supervisão democrática adequada é uma grande preocupação quando da implementação de um protocolo de segurança. A Declaração Conjunta de 2015 afirma que "*[m]edidas administrativas que limitem diretamente a liberdade de expressão, incluindo sistemas que regulem meios de comunicação, deveriam ser sempre aplicadas por um órgão independente. Também deveria ser possível recorrer contra a aplicação de medidas administrativas a um tribunal independente ou outro órgão adjudicatório.*" **A este respeito, e com base em importantes salvaguardas relacionadas, o mecanismo do protocolo de segurança deve contar com freios e contrapesos robustos, incluindo:** *(i)* uma entidade governamental independente ou estrutura de supervisão que avalia a situação de crise com base em critérios claros e transparentes e determina a *implementação* ou *prorrogação* do protocolo de segurança por meio de decisão fundamentada dentro de processo administrativo público respeitando as garantias de devido processo legal; *(ii)* um referendo ou consulta prévia de um conselho participativo multissetorial como parte do processo de decisão (para implementar ou prorrogar o protocolo); *(iii)* assim como o processo administrativo, não apenas um resumo, mas a própria resolução que implementa ou prorroga o protocolo de segurança é pública; *(iv)* o direito a um reexame judicial; *(v)* transparência adequada e contínua sobre as medidas dos provedores

que derivam do protocolo de segurança e sobre atividades governamentais de supervisão.
- Por último, o artigo 16, que estabelece o mecanismo de notificação pelo usuário, deixa definições essenciais para regulamentação posterior. Ele deve, pelo menos, esclarecer que as notificações de usuários devem indicar especificamente a localização do material supostamente ilegal e explicar por que o usuário o considera ilegal. O projeto de lei também deveria explicitar que as garantias de devido processo que o artigo 18 assegura para os usuários que têm seu conteúdo restringido permanecem aplicáveis no contexto de um protocolo de segurança, abarcando os provedores e tipos de conteúdo afetados e todo o período em que o protocolo vigorar.

Estrutura de supervisão independente e participativa

O projeto de lei estipula obrigações para aplicações de internet e poderes para uma autoridade administrativa não especificada para supervisionar o cumprimento das regras do PL 2630. A aplicação do projeto de lei sem uma estrutura de supervisão genuinamente independente e democrática compromete seus objetivos. Até agora, o texto da proposta não logra garantir a base para tal estrutura, dando uma margem maior à aplicação arbitrária do PL 2630, em vez de estabelecer as bases para evitar tais abusos. Embora os projetos de lei de iniciativa do Poder Legislativo estejam limitados para a criação de novas entidades no âmbito da administração federal, essa é uma equação política que o Congresso e o governo federal devem resolver, em debate com a sociedade civil, antes de aprovar o PL 2630.

A Anatel, agência reguladora de telecomunicações, tem trabalhado para se encaixar como resposta. A agência já existe e conta com atributos essenciais assegurados por lei, como independência administrativa, ausência de subordinação hierárquica, estabilidade de seus diretores e autonomia financeira. No entanto, sua experiência e competências dizem respeito a serviços e infraestruturas de telecomunicações, não a aplicações de internet e atividades de moderação de conteúdo. Além disso, o histórico da Anatel deixa muito a desejar, tanto no cumprimento de suas competências como agência reguladora das telecomunicações quanto na garantia de participação significativa da sociedade civil em suas decisões.

A *Coalizão Direitos na Rede* enfatizou um conjunto de deficiências da Anatel em uma declaração pública divulgada no início deste ano. Entre elas, a coalizão critica o favorecimento da Anatel às grandes operadoras de telecomunicações no leilão das faixas do espectro para a prestação do 5G. Também aponta falhas quanto à eficiência e transparência da fiscalização da Anatel, com base em relatórios do Tribunal de Contas da União (TCU). Por outro lado, a *Coalizão Direitos na Rede* defende uma nova agência de supervisão autônoma apoiada por um conselho participativo e multissetorial.

Isso se alinha à Declaração Conjunta de 2019 dos Relatores Especiais para a Liberdade de Expressão, que apoia *"mecanismos de supervisão independentes, multissetoriais e transparentes para lidar com regras de conteúdo privado que podem ser inconsistentes com o direito internacional de direitos humanos e ingerir no direito dos indivíduos de desfrutar da liberdade de expressão".*

A Comissão Especial de Direito Digital da Ordem dos Advogados do Brasil (OAB) também propôs uma estrutura de supervisão mais elaborada. Ela envolveria três frentes: *(i)* uma entidade fiscalizadora e deliberativa formada por representantes dos três poderes do governo (Legislativo, Executivo, Judiciário), das autoridades brasileiras de concorrência e proteção de dados, Anatel e OAB; *(ii)* uma entidade autorreguladora responsável por tratar de casos específicos de moderação de conteúdo e *(iii)* o Comitê Gestor da Internet do Brasil (CGI.br), que já desempenha um papel fundamental na publicação de estudos, diretrizes e recomendações para o desenvolvimento da Internet no país. Um ponto essencial é que qualquer projeto deve preservar o papel e a natureza atuais do CGI.br.

As propostas da *Coalizão Direitos na Rede* e da Comissão Especial da OAB refletem a necessidade de freios e contrapesos robustos, incluindo uma participação significativa da sociedade civil, no projeto de supervisão do PL 2630. Isso ainda está faltando, e preencher essa lacuna fundamental exige um debate comprometido e participativo.

Garantias claras contra o aumento da vigilância e dos riscos de segurança relacionados.

Dadas as novas obrigações que o PL 2630 estabelece para os provedores de aplicação, incluindo regras específicas para situações de crise, é importante deixar claro no PL que nenhuma de suas disposições implicará mudanças nos sistemas das plataformas para introduzir vulnerabilidades de segurança ou comprometer as proteções de privacidade por padrão. Isso é de vital importância para preservar as implementações de criptografia de ponta-a-ponta em aplicações de internet e prevenir intuitos de enfraquecer os princípios e proteções fundamentais da criptografia.

Nesse sentido, a Declaração Conjunta de 2016 dos Relatores Especiais para a Liberdade de Expressão que trata dos esforços do governo para combater o extremismo violento enfatiza que os Estados não devem adotar e devem revisar leis e políticas que envolvam medidas que enfraqueçam as ferramentas de segurança digital existentes. O artigo 8º do PL 2630 já estipula que as medidas que os provedores implementem em conformidade com o projeto de lei devem preservar a segurança da informação e a proteção de dados pessoais. Isso é bom, mas a disposição deve ir além para repelir explicitamente aplicações da lei que busquem introduzir vulnerabilidades nos sistemas das plataformas ou fazer com que as aplicações de internet adotem outras medidas que possam aumentar sistematicamente o risco de incidentes de segurança.

Além disso, o projeto de lei contém regras que ampliam as obrigações de retenção de dados existentes. Neste ponto, a Declaração Conjunta de 2015 sobre situações de crise afirma que "*requisitos para reter ou práticas de retenção de dados pessoais de forma indiscriminada com o fim de aplicação da lei ou segurança não são legítimos. Pelo contrário, os dados pessoais deveriam ser retidos por motivos de aplicação da lei ou segurança apenas de forma limitada e direcionada e de uma maneira que represente um equilíbrio adequado entre as necessidades de aplicação da lei e de segurança e os direitos à liberdade de expressão e privacidade*".

A previsão mais problemática relacionada às obrigações de guarda de dados se encontra no artigo 46 do PL 2630. O texto exige que as aplicações de internet guardem os metadados associados a todos os conteúdos que forem removidos ou desativados como consequência das regras do PL 2630 ou por ordens judiciais. Embora possa parecer, à primeira vista, uma medida "direcionada", relacionada a um conteúdo potencialmente ofensivo, o volume de conteúdos restringidos enquadrados pela regra tende a ser enorme pela própria natureza e dinâmica da criação de conteúdos por usuárias e usuários em grandes plataformas. Se faz sentido guardar o conteúdo restringido por um período específico, a redação do projeto de lei é muito ampla quanto aos metadados que as aplicações de internet teriam que armazenar junto com esse conteúdo.

De acordo com o Artigo 46, a obrigação de guarda inclui "quaisquer dados e metadados conexos removidos" juntamente com o conteúdo, bem como "os respectivos dados de acesso à aplicação, como o registro de acesso, endereço de protocolo de internet, incluindo as portas de origem, além de dados cadastrais, telemáticos, outros registros e informações dos usuários que possam ser usados como material probatório, inclusive as relacionadas à forma ou meio de pagamento, quando houver". O período de armazenamento é de 6 meses, podendo ser prorrogado.

A Autoridade Nacional de Proteção de Dados (ANPD) emitiu uma declaração criticando a natureza vaga de disposições do projeto de lei que estabelecem a coleta de dados pessoais para fins de investigação criminal, com referências específicas à redação do Artigo 46. De acordo com a ANPD, "[o] PL nº 2630/20 estabelece obrigações de guarda de dados para fins de investigação criminal, valendo-se, para tanto, de expressões vagas e imprecisas, o que pode levar a uma ampliação desproporcional da coleta de dados pessoais ou, ainda, ao rastreamento e à vigilância abusivas sobre titulares de dados pessoais". A autoridade brasileira de proteção de dados destaca que as autoridades governamentais devem observar a necessidade de definir as finalidades específicas para o tratamento de dados pessoais, limitar esse tratamento ao estritamente necessário para atingir essas finalidades, adotar medidas de segurança proporcionais aos riscos envolvidos e garantir ampla transparência das operações realizadas com os dados pessoais. Nesse sentido, a ANPD recomenda que os legisladores revisem o texto do projeto de lei para indicar expressa e explicitamente quais dados podem ser coletados.

Considerando os princípios de finalidade, necessidade e prevenção assegurados na Lei de Proteção de Dados Pessoais brasileira, a guarda padrão de metadados associados a conteúdos restringidos estabelecida no PL não deve ir além das regras de retenção de dados já estipuladas no Marco Civil. Com a guarda de registros de acesso a aplicações prevista no Marco Civil, que inclui o endereço IP do usuário, as autoridades podem

iniciar uma investigação e, no âmbito de seus procedimentos, solicitar informações adicionais ou outras verificações conforme necessário e dependendo de cada caso.

Rever a problemática imunidade de autoridades públicas

O artigo 33, parágrafo 6º, do projeto de lei amplia a imunidade que a Constituição brasileira garante aos parlamentares por suas opiniões, palavras e votos no exercício de seus mandatos a conteúdos publicados por "agentes políticos" em redes sociais e plataformas de mensageria privada. O termo "agentes políticos" no artigo parece abranger todas as autoridades eleitas nos poderes Executivo e Legislativo nos níveis federal, estadual e municipal, bem como ministros de estado, secretários estaduais e municipais e os dirigentes de entidades governamentais em geral. Se esta disposição for aprovada, este grande conjunto de autoridades estaria imune à responsabilização civil e criminal pelo conteúdo que publicam *online*.

O projeto de lei confere proteções especiais ao discurso de autoridades públicas, enquanto os padrões interamericanos de liberdade de expressão reconhecem que estas autoridades, pelo contrário, têm obrigações especiais por suas declarações. Tais obrigações incluem o dever de garantir que suas declarações não sejam uma ingerência arbitrária, direta ou indireta, nos direitos daqueles que contribuem para o discurso público com a expressão e difusão de seus pensamentos, o dever de garantir que suas declarações não se configurem como violações de direitos humanos e o dever de razoavelmente verificar os fatos nos quais suas declarações se baseiam.

Tendo em vista esses deveres, a Declaração Conjunta de 2021 dos Relatores Especiais para a Liberdade de Expressão, ao abordar preocupações crescentes com a disseminação da desinformação, enfatizou que os Estados devem *"a) [a]dotar políticas que estabeleçam a imposição de medidas disciplinares às pessoas que exercem funções públicas que, atuando ou sendo percebidas como atuando no exercício de suas funções, realizem, patrocinem, incentivem ou sigam disseminando declarações que elas saibam ou deveriam razoavelmente saber que são falsas; b) [g]arantir que as autoridades públicas façam todo o possível para difundir informações precisas e confiáveis, incluindo a respeito das suas atividades e de assuntos de interesse público.*

O projeto de lei, cujas raízes se baseiam em preocupações semelhantes, contém uma previsão que parece ignorar o papel que as autoridades públicas proeminentes desempenham na criação, financiamento e disseminação de conteúdo nocivo *online*. Esta previsão contradiz os objetivos pretendidos pelo PL 2630, e os legisladores brasileiros devem rejeitar o seu texto.

Conclusão

Quaisquer leis e regulações que busquem fortalecer os direitos de usuárias e usuários frente a aplicações de internet dominantes devem se guiar por esses princípios e garantias, em vez de descartá-los. Não podemos oferecer respostas aos desafios decorrentes da inter-relação constante, mas em contínua mudança, entre as tecnologias digitais e a sociedade, se em cada etapa desse caminho desconsiderarmos bases relevantes já estabelecidas, fundadas em padrões internacionais de direitos humanos. Empoderar usuárias e usuários perante o enorme poder corporativo das plataformas de internet envolve ainda medidas mais estruturais e econômicas, que estão em grande medida negligenciadas no debate atual, como promover a interoperabilidade das redes sociais. Esperamos que as preocupações e princípios que articulamos aqui possam contribuir para o debate atualmente em curso no Brasil.