

CURSO GRATUITO

ESTRUTURA E FUNCIONAMENTO DA INTERNET

ASPECTOS TÉCNICOS,
POLÍTICOS E REGULATÓRIOS



FICHA TÉCNICA

CENTRO DE ENSINO E PESQUISA EM INOVAÇÃO

CEPI FGV Direito SP

Coordenação

Alexandre Pacheco da Silva

Marina Feferbaum

Pesquisadores envolvidos

Ana Carolina Rodrigues

Ana Paula Camelo

Guilherme Forma Klafke

João Pedro Favaretto Salvador

Jordan Vinícius de Oliveira

Laurianne-Marie Schippers

Victor Doering Xavier da Silveira

Tatiane B. Guimarães

Saiba mais: <http://bit.ly/cepidireitosp>

INTERNET SOCIETY CAPÍTULO BRASIL

ISOC Brasil

Presidente

Flavio Rech Wagner

Saiba mais: <https://isoc.org.br/>

APOIO

Twitter Brasil

Coordenadora de Políticas Públicas do Twitter Brasil

Natália Neris

DIAGRAMAÇÃO

Laurianne-Marie Schippers



Este trabalho está licenciado sob uma licença *Creative Commons*.

Como citar este trabalho

CEPI FGV DIREITO SP; ISOC BRASIL. Estrutura e funcionamento da Internet: aspectos técnicos, políticos e regulatórios – Curso livre. São Paulo: FGV Direito SP; ISOC Brasil, 2022.

As opiniões expressas nesta publicação são de responsabilidade exclusiva dos(as) autores(a)s, não refletindo necessariamente a opinião institucional do CEPI FGV Direito SP, da ISOC Brasil e/ou do Twitter Brasil.

Realização



Apoio



Estrutura e funcionamento da internet [recurso eletrônico] : aspectos técnicos, políticos e regulatórios / organização, Centro de Ensino e Pesquisa em Inovação da Escola de Direito de São Paulo da Fundação Getulio Vargas, Internet Society-Capítulo Brasil. - São Paulo : CEPI-FGV Direito SP/ISOC Brasil, 2022.

112 p.

Inclui bibliografia.

ISBN: 978-65-87355-38-2

1. Internet - aspectos morais e éticos. 2. Internet - aspectos políticos. 3. Liberdade de expressão. 4. Desinformação. 5. Inteligência artificial. 6. Proteção de dados. I. Internet Society-Capítulo Brasil. II. Fundação Getulio Vargas. Centro de Ensino e Pesquisa em Inovação. III. Escola de Direito de São Paulo da Fundação Getulio Vargas.

CDU 004.738.5

Ficha catalográfica elaborada por: Cristiane de Oliveira CRB SP-008061/O
Biblioteca Karl A. Boedecker da Fundação Getulio Vargas - SP

Realização



Apoio



ESTRUTURA E FUNCIONAMENTO DA INTERNET: ASPECTOS TÉCNICOS, POLÍTICAS

Realização: CEPI FGV Direito SP | ISOC Brasil

Apoio: Twitter Brasil

Artigos elaborados como produtos do curso livre (2021)

Os textos apresentados a seguir foram elaborados como produtos do curso livre **Estrutura e funcionamento da Internet: aspectos técnicos, políticas (2021)**, que teve como objetivo aprofundar o conhecimento sobre a estruturação, os princípios norteadores e o funcionamento da Internet e engajar e capacitar profissionais de múltiplas áreas para que possam incidir nos debates sobre políticas públicas para o setor no Brasil.

A partir de encontros e debates online semanais, entre setembro e novembro de 2021, os(as) alunos(as) desenvolveram um ensaio como trabalho final que consta nesta publicação. Os textos foram comentados por pesquisadores e pesquisadoras do CEPI e a versão final revisada pelos alunos(as).

Refletindo a proposta e dinâmica do curso estruturada a partir do tema da Internet Aberta, os textos exploram suas dimensões jurídicas, tecnológicas, sociológicas, de políticas públicas, entre outras.

Na primeira seção, apresentamos as perspectivas do CEPI, da ISOC Brasil e do Twitter Brasil que viabilizaram este projeto. Na sequência, serão apresentados os trabalhos que foram organizados na seguinte ordem: (i) Abordagens Conceituais; (ii) Moderação de Conteúdo; (iii) Desinformação; (iv) Algoritmos e uso ético de Inteligência Artificial; (v) Privacidade e Proteção de Dados Pessoais; (vi) Soberania Digital. Dentro dessas seções, por sua vez, os trabalhos foram organizados em ordem alfabética dos(as) autores(as).

Boa leitura!

Realização



Apoio



SUMÁRIO

SOBRE O CURSO: CEPI FGV DIREITO SP	5
SOBRE O CURSO: INTERNET SOCIETY	10
SOBRE O CURSO: TWITTER BRASIL	13
ABORDAGENS CONCEITUAIS	15
NOTAS INICIAIS SOBRE OS CONCEITOS DE PROVEDORES A PARTIR DA DISTINÇÃO DO PROJETO DE LEI Nº 2630	16
MODERAÇÃO DE CONTEÚDO	22
FAKE, VERDADE OU IRONIA? O SUBJETIVISMO LINGÜÍSTICO E CULTURAL COMO DESAFIO À MODERAÇÃO DE CONTEÚDO	23
FACEBOOK PAPERS E A TRANSPARÊNCIA DE ALGORITMOS A FAVOR DO DEBATE REGULATÓRIO: ESTUDO DE CASO	30
LIBERDADE DE OPINIÃO AINDA É PERMITIDA NA INTERNET?	36
MODERAÇÃO DE CONTEÚDO NA INTERNET – UMA BREVE AVALIAÇÃO DOS TERMOS DE USO DO INSTAGRAM	44
DESINFORMAÇÃO	51
O IMPACTO DA “PL DAS FAKE NEWS” NA LIBERDADE DE EXPRESSÃO.....	52
FORMAS DE COMBATE À DESINFORMAÇÃO SOBRE A PANDEMIA DE COVID-19 ADOTADAS PELO TWITTER	62
ALGORITMOS E USO ÉTICO DE INTELIGÊNCIA ARTIFICIAL	71
UTILIZAÇÃO DE INTELIGÊNCIA ARTIFICIAL NA TOMADA DE DECISÕES EM DIFERENTES CENÁRIOS: ASPECTOS POSITIVOS E NEGATIVOS, O PAPEL DA ATUAÇÃO HUMANA NESSE CONTEXTO E O GRAU DE TRANSPARÊNCIA NAS DECISÕES	72
TRANSPARÊNCIA E EXPLICABILIDADE DE ALGORITMOS: DA CLUSTERIZAÇÃO AO INDIVÍDUO	79
TRANSPARÊNCIA DE ALGORITMOS E GIG ECONOMY: DISCUSSÕES ACERCA DA VIABILIDADE DE REALIZAÇÃO DE PERÍCIA ALGORÍTMICA	88
PRIVACIDADE E PROTEÇÃO DE DADOS PESSOAIS	96
OS DADOS SÃO DE QUEM?	97
SOBERANIA DIGITAL	105
FLUXOS TRANSNACIONAIS E INTERNET ABERTA NO SÉCULO XX	106

SOBRE O CURSO

CEPI FGV DIREITO SP

Esta publicação nasce em um momento importante de debates e reflexões sobre o que entendemos e defendemos como Internet aberta: uma Internet conectada globalmente, segura e confiável para todos.

Em 2009, o Comitê Gestor da Internet (CGI) lançava 10 Princípios para a governança e uso da Internet, naquela época tidos como princípios inéditos em âmbito local (Brasil) e global, com impactos de longo prazo, dentre eles na redação do Marco Civil da Internet e da Lei de Proteção de Dados brasileira (Possebon, 2019).

A importância e atualidade do Decálogo do CGI se materializa na defesa da: Liberdade de expressão, defesa da privacidade e respeito aos direitos humanos; Governança democrática e coletiva, multilateral e transparente; Acesso universal para desenvolvimento social e humano; Diversidade cultural; Inovação; Neutralidade da rede, não admitindo qualquer tipo de discriminação ou favorecimento político, comercial, religioso ou cultural; Estabilidade, segurança e funcionalidade a partir de padrões abertos e globais; além da garantia de um ambiente legal e regulatório que preserve a sua estrutura e funcionamento como espaço de colaboração (CGI.br, 2009).

Nos últimos anos, uma série de desafios nos parecem demandar especial atenção e defesa desses princípios e da concepção do que é e porque a Internet deve continuar sendo aberta, segura e global, tanto do ponto de vista técnico, político quanto regulatório.

O combate à desinformação e o polêmico PL 2630; a disseminação de discurso de ódio e extremismos versus os impasses em termos de moderação de conteúdo; o crescente vazamento de dados pessoais de órgãos públicos e empresas privadas; os desafios técnicos e éticos da inteligência artificial; o conflito entre Rússia e Ucrânia que levou ao pedido de banimento da Rússia da Internet (Mendes, 2022) são apenas alguns de vários destaques recentes que nos levam a refletir e mobilizar debates sobre a Internet que queremos.

Como parte desse contexto complexo e da necessidade de compreensão e participação ativa e qualificada no debate e da construção dessa Internet, nos sentimos motivados e desafiados a olhar para esses e outros temas com o objetivo de aprofundar o conhecimento sobre a estruturação, os princípios norteadores e o funcionamento da Internet Aberta.

Esta publicação marca um pouco desse processo, celebrando a parceria entre o Centro de Ensino e Pesquisa em Inovação da Escola de Direito de São Paulo da Fundação Getulio Vargas (CEPI FGV Direito SP), o capítulo brasileiro da Internet Society (ISOC Brasil) e o Twitter Brasil, que resultou na realização da segunda edição do curso Estrutura e Funcionamento da Internet: aspectos técnicos, políticos e regulatórios.

O curso foi desenvolvido para abranger diversos aspectos relacionados à infraestrutura e ao funcionamento da Internet como "a rede das redes", considerando suas dimensões técnica, regulatória e social.

Compreender os princípios que orientam o desenho estrutural e o funcionamento da Internet; desenvolver uma perspectiva crítica sobre o funcionamento e o ecossistema de governança da Internet, abarcando dimensões técnicas e não técnicas; avaliar iniciativas regulatórias e legislativas, políticas e técnicas que possam apresentar riscos ou proporcionar condições para a operação da Internet como uma rede aberta e descentralizada, foram alguns dos principais objetivos do projeto que contou com a participação de grandes especialistas nos temas em debate.

O formato online e a premissa de diversificação do perfil dos(as) alunos(as) é um elemento que prezamos por destacar como forma de expandir essa oportunidade para pessoas que se encontram em outros estados e que não se restringem a advogados(as) e juristas. Alunos e alunas de diferentes áreas de formação e atuação, de todas as regiões do Brasil e com diferentes níveis de envolvimento com o tema do curso, foram essenciais para o sucesso do projeto.

O encontro e interação entre estudantes e profissionais nos parecem fundamentais para uma adequada compreensão das complexidades que conformam os desafios endereçados pelo curso. Como resultado prático, essa combinação tem garantido não apenas a qualidade, mas o caráter de difusão dos aprendizados mobilizados pelo grupo.

Os encontros conjugaram aulas gravadas, compartilhamento de leituras preparatórias e encontros online ao vivo (via plataforma Zoom) durante três meses, com encontros semanais (entre 22/09/2021 e 24/11/2021). Para cada aula, foram convidados especialistas e profissionais de notório saber sobre os temas em pauta para proporcionar um espaço rico de trocas e aprendizado a partir de diferentes perspectivas.

Além da metodologia participativa em um ambiente virtual, orientada por métodos ativos e coletivos de debate e construção do conhecimento, a edição 2021 teve como grande diferencial a discussão de casos desenvolvidos especialmente para o curso a fim de aprofundar a compreensão e discussão dos impactos e desafios atrelados às políticas de remoção de conteúdos e ao processo tomada de decisão; à identificação e responsabilização das iniciativas que ameaçam alguns princípios para a governança e uso da Internet; e ao bloqueio e/ou banimento de aplicativos em diferentes situações com base em argumentos de soberania nacional.

A participação ativa de todos os envolvidos, desde os parceiros, especialistas que aceitaram os convites e os(as) alunos(as) que se dedicaram nos 10 encontros desta edição, nos leva a apresentar com orgulho esta publicação.

Princípios e fundamentos da Internet Aberta; responsabilidade de intermediários; transparência e explicabilidade de algoritmos; moderação de conteúdo; criptografia, desinformação e discurso de ódio; privacidade e proteção de dados são alguns dos tópicos que exploramos de forma relacional nas aulas e que você (leitor(a)) encontrará refletidos nos ensaios a seguir.

Os textos foram elaborados como produtos finais do curso pelos alunos e alunas e contaram com os comentários de pesquisadoras e pesquisadores do CEPI FGV Direito SP com o objetivo de aprofundar o conhecimento e refletir sobre políticas públicas para o setor no Brasil.

A partir da experiência do CEPI, esta tem sido uma oportunidade ímpar de congregar iniciativas de ensino e pesquisa na interface entre Direito, Tecnologia e Sociedade de forma gerar e multiplicar conhecimento por meio de bens públicos e interação com grandes nomes que também se engajam e buscam contribuir para um processo de tomada de decisão responsável no que tange à estrutura e ao funcionamento da Internet ao longo da história.

Por isso, não menos importante, deixamos nossos agradecimentos ao Professor Flávio Wagner e ao João Paulo de Vasconcelos Aguiar, pelo envolvimento e contribuições essenciais à realização do projeto em termos de forma e conteúdo; ao Capítulo Brasileiro da Internet Society e ao Twitter Brasil pelo apoio institucional; aos pesquisadores e pesquisadoras do CEPI pela dedicação e comprometimento nas diferentes fases do projeto; e aos alunos e alunas que se dedicaram e contribuíram para o sucesso desse projeto.

Fica o desejo de promover novos e melhores cursos que contribuam com o debate público sobre a governança da Internet e seus principais desafios. Ansiosos pelos próximos! Por isso, deixamos a sugestão de leitura e o convite para participar dessa construção conosco.

Desejamos a todos e todas uma boa leitura!

Alexandre Pacheco da Silva

Prof. da FGV Direito SP e Coordenador do CEPI FGV Direito SP

Ana Paula Camelo

Líder de pesquisa e gestora de projetos do CEPI FGV Direito SP

Referências bibliográficas

CGI.BR. **Princípios para a governança e uso da Internet**. [S.l.], [2009]. Disponível em: <https://principios.cgi.br/>. Acesso em: 06 jun. 2022.

INTERNET SOCIETY. **Quick Analysis** – The Impact of Effort to Disconnect Russia from the Internet. [S. l.]. Publicado em: 18 mar. 2022. Disponível em: <https://www.internetsociety.org/resources/doc/2022/quick-analysis-the-impact-of-efforts-to-disconnect-russia-from-the-internet/>.

MENDES, F. Rússia pode ser banida da internet como conhecemos? Entenda. **Tilt Uol**. São Paulo, 10 mar. 2022. Disponível em: <https://www.uol.com.br/tilt/noticias/redacao/2022/03/10/ucrania-queria-remover-russia-da-internet.htm>.

POSSEBON, S. 10 anos do decálogo da Internet: os princípios ainda estão atuais? **Teletime**, [s.l.], 05 jun. 2019. Disponível em: <https://teletime.com.br/05/06/2019/10-anos-do-decalogo-da-internet-os-desafios-atuais-e-futuros/>. Acesso em: 06 jun. 2022.

SOBRE O CURSO

INTERNET SOCIETY

Um dos projetos estratégicos da Internet Society em escala global, iniciado em 2020, é denominado 'Internet Way of Networking' (IWN, no acrônimo em inglês, ou o 'Modo Internet de Interconectividade', em uma tradução livre para o português). Este projeto tem por objetivo promover a defesa de um conjunto de propriedades críticas fundamentais que são inerentes ao desenho estrutural da Internet e que explicam o seu sucesso ao longo de várias décadas, mesmo com profundas transformações tanto nas tecnologias sobre as quais ela se apoia como nos seus usos. O 'Internet Way of Networking' é fundamentado em uma rede aberta que é baseada em padrões técnicos comuns, um sistema de roteamento distribuído e um conjunto de identificadores únicos e que é neutra para qualquer tecnologia e aplicação. Em complemento a estas propriedades críticas, a Internet Society oferece o 'Internet Impact Assessment Toolkit' (IIAT, no acrônimo em inglês), um conjunto de ferramentas a serem usadas na análise de eventuais propostas de legislações, políticas públicas, normas e novas tecnologias que possam de alguma maneira afetar essas propriedades críticas da Internet. Em 2021, este conjunto de ferramentas foi estendido com a definição de um conjunto de 11 'habilitadores', que correspondem a características importantes que garantem que a Internet permaneça aberta, globalmente conectada, segura e confiável, os quatro grandes objetivos aspiracionais que, segundo a ISOC, são os pilares da Internet. A ISOC Brasil, o capítulo brasileiro da Internet Society, está engajada fortemente nesse projeto, contribuindo para a consolidação tanto do conjunto de propriedades críticas do IWN como do IIAT.

Ainda em 2020, a ISOC Brasil procurou a FGV, através do seu Centro de Ensino e Pesquisa em Inovação (CEPI), para propor uma parceria no oferecimento de um curso online aberto a pessoas de todo o país, com ampla diversidade regional, setorial e de perfis pessoais e profissionais, tendo por objetivo a discussão dos aspectos técnicos, políticos e regulatórios das propriedades críticas do 'Internet Way of Networking', especialmente considerando o contexto brasileiro - tanto em perspectiva histórica, quanto a partir da realidade atual. Esta proposta procurou estimular a discussão de aspectos tão diversos como os princípios de projeto que

estruturam a Internet, a responsabilidade dos intermediários, a neutralidade da rede, o combate à desinformação e ao discurso de ódio, o papel da criptografia, a localização de dados e os riscos de fragmentação da Internet, a partir de medidas que, adotadas localmente, podem impactar o seu ecossistema em uma perspectiva mais ampla. Renomados especialistas brasileiros, de diferentes áreas, foram convidados a contribuir, participando de entrevistas pré-gravadas que foram disponibilizadas aos participantes e/ou participando de debates ao vivo durante as aulas.

Este curso seguiu um modelo que a ISOC Brasil tem adotado em outras iniciativas, que é baseado na parceria com entidades brasileiras de renome e experiência em diferentes temas a partir da mobilização de recursos da ISOC, da ISOC Foundation e da própria entidade parceira. O curso foi planejado e formatado num trabalho conjunto entre o CEPI, a ISOC Brasil e a ISOC, de maneira a assegurar o alinhamento de seu conteúdo aos objetivos estratégicos do capítulo no desenvolvimento de atividades relacionadas ao IWN. Todo o processo foi conduzido de maneira entusiasmada e competente pela equipe do CEPI.

Com o grande sucesso da primeira edição do curso em 2020, a ISOC Brasil, a ISOC e o CEPI iniciaram o planejamento da segunda edição, que se realizou em 2021 e agora se conclui com a publicação deste volume, contendo os ensaios preparados pelos participantes do curso.

De grande importância para o sucesso desta segunda edição foi a inclusão de um novo parceiro estratégico, o Twitter Brasil, que aportou sua grande experiência e interesse no tema da Internet Aberta, que é uma das propriedades críticas conforme propugnado pelo paradigma do 'Internet Way of Networking' e que ganhou especial destaque na programação do curso.

Como na primeira edição, em 2021 mais uma vez o curso alcançou e ultrapassou amplamente seus objetivos inicialmente propostos. Além dos objetivos precípuos (de um lado, a discussão das propriedades críticas da Internet e de sua relação com aspectos políticos e regulatórios no contexto brasileiro e, de outro lado, a formação de profissionais de todo o país neste tema tão relevante), o curso tem outros desdobramentos muito importantes.

Em primeiro lugar, o curso obviamente fortalece a parceria entre a ISOC, a ISOC Brasil, o CEPI da FGV e o Twitter Brasil, abrindo caminho para novas iniciativas conjuntas no futuro.

Para a Internet Society, o curso é um exemplo para seus capítulos em outros países, ilustrando de que forma um capítulo pode alavancar recursos financeiros da ISOC, da ISOC Foundation e de entidades e empresas parceiras e, em trabalho conjunto com entidades locais, fomentar dentro do país a discussão de questões fundamentais para o contínuo desenvolvimento da Internet.

Além disto, o curso deu origem a esta publicação coletiva, reunindo ensaios de todos os seus participantes e que cobrem um amplo espectro de temas e de perspectivas. Tal publicação servirá certamente como referência importante para todos os estudiosos desses temas aqui no Brasil. Adicionalmente, tal resultado pode ser mostrado aos mais de 100 capítulos nacionais da Internet Society como um exemplo de engajamento de cada país na discussão dos grandes projetos estratégicos da entidade.

Finalmente, estendendo um trabalho de divulgação já feito em 2020, uma parte expressiva do curso está sendo disponibilizada aos demais capítulos da ISOC, em particular da América Latina e Caribe, e à comunidade em geral. Todas as entrevistas gravadas com especialistas ao longo do curso, assim como as duas mesas de debates finais, que trataram da Internet Aberta em perspectivas global e nacional, foram legendadas em espanhol e em inglês, graças a um importante apoio da ISOC. Com isso, a ISOC Brasil, a ISOC, o CEPI e o Twitter Brasil trazem uma grande contribuição para a discussão do 'Internet Way of Networking' em âmbito global, reforçando o caráter de pioneirismo dessa parceria e servindo de exemplo a outros países e regiões.

Que o legado desta parceria e a leitura desta publicação sirvam como um ponto de partida e de estímulo a todas as pessoas empenhadas em promover e defender uma Internet aberta, globalmente interconectada, segura e confiável sempre e em todos os lugares.

Flávio Rech Wagner

Presidente da ISOC Brasil, capítulo brasileiro da Internet Society

SOBRE O CURSO

TWITTER BRASIL

O trabalho do Twitter é manter as pessoas informadas sobre o que está acontecendo no mundo, e nosso propósito é servir a conversa pública. Nós fazemos isso servindo a nossa audiência global e focando nas necessidades das pessoas que usam o nosso serviço. No Twitter, nós estamos comprometidos em prover um serviço que encoraja e facilita o debate aberto e democrático e o mais diverso e plural possível. Estamos também comprometidos em defender o que consideramos ser uma rede aberta, sustentável no longo prazo e que priorize, ao mesmo tempo, o equilíbrio entre segurança e liberdade de expressão e informação a pessoas de todo o mundo.

Para tanto, estabelecemos cinco princípios orientadores para a regulamentação da Internet:

1. A Internet Aberta é global, deve estar disponível para todos e ser construída com base em princípios e na proteção dos direitos humanos.
2. A confiança é essencial e pode ser construída com transparência, tratamento justo e proteção da privacidade.
3. Os algoritmos de recomendação e classificação devem estar sujeitos ao controle e escolha humana.
4. Competição, liberdade de escolha e inovação são fundamentos da Internet Aberta e devem ser protegidos e expandidos, de forma a garantir que os usuários não acabem sendo prejudicados por leis e regulamentos.
5. A moderação de conteúdo é mais do que apenas deixar ou retirar do ar. A regulamentação deve permitir uma gama de intervenções, ao mesmo tempo que apresenta definições e limites substantivos e consistentes com os padrões de direitos humanos.

Entendemos que a efetividade na defesa desses princípios é tanto maior a medida que possamos trabalhar com parceiros que possuem objetivo semelhante, como a ISOC Brasil - entidade da sociedade civil que exemplarmente vêm produzindo sobre o tema- e também com parceiro no âmbito acadêmico que além de produzir, forma novos pesquisadores no campo e assim, contribui para a qualificação do debate público. Este é o caso do Centro de Ensino e Pesquisa em Inovação (CEPI) da Fundação Getúlio Vargas.

Desde 2020, o CEPI vem promovendo um curso online, aberto e gratuito para profissionais e pesquisadores interessados no debate sobre direitos digitais. A iniciativa tem o mérito de prezar pela excelência dos professores convidados e oportunizar, inclusive àqueles que não estão nos grandes centros urbanos, a oportunidade de troca sobre o tema. Considerando o potencial da iniciativa - também no que se refere à pluralização de ideias no debate - é que o Twitter Brasil entendeu ser fundamental apoiar sua segunda edição cujo tema foi "Estrutura e Funcionamento da Internet: Aspectos Técnicos, Políticos e Regulatórios".

Sentimos muito orgulho de apoiar a iniciativa já no momento em que participamos dos debates (de altíssimo nível) nas dez aulas do curso. Tivemos também a oportunidade de discutir, do ponto de vista do Twitter, sobre a complexidade e os desafios da manutenção de uma Internet Aberta por meio da participação em um painel com a presença da nossa líder da equipe de Políticas Públicas do Twitter para os Estados Unidos, Canadá e América Latina, Jessica Herrera-Flanigan.

Conhecer os frutos destes ciclos tais como a formação dos/as estudantes, a disponibilização de grande parte do conteúdo também em espanhol e inglês e também esta publicação - que reflete a qualidade de pontos de vista de profissionais e pesquisadores das cinco regiões do país-, reforça nosso orgulho e a nossa certeza de que a construção da Internet Aberta que desejamos tem o potencial de se realizar a partir do trabalho coordenado e do diálogo produtivo entre entes da sociedade civil, academia e indústria.

Time de Políticas Públicas Twitter Brasil

Realização



Apoio



CURSO GRATUITO

**ESTRUTURA E
FUNCIONAMENTO
DA INTERNET**

ASPECTOS TÉCNICOS,
POLÍTICOS E REGULATÓRIOS

ESTRUTURA E FUNCIONAMENTO DA INTERNET

ASPECTOS TÉCNICOS, POLÍTICOS E REGULATÓRIOS

ABORDAGENS CONCEITUAIS

Realização



Apoio



NOTAS INICIAIS SOBRE OS CONCEITOS DE PROVEDORES A PARTIR DA DISTINÇÃO DO PROJETO DE LEI Nº 2630

André Lucas Fernandes¹

RESUMO: O presente ensaio se volta a analisar a semântica conceitual atribuída à palavra “provedor” no Marco Civil da Internet e no Projeto de Lei das Fake News. Toma como referencial teórico as balizas da história dos conceitos, conforme trabalho de Reinhart Koselleck. Conclui-se, de forma breve, por uma ruptura no sentido conceitual, cujas implicações dogmáticas podem ser negativas.

PALAVRAS-CHAVE: História dos conceitos; Provedores; Internet; Marco civil da internet; PL 2630

A guisa de introdução

O presente texto se insere no tipo de esforço ideativo que Ortega y Gasset chama “ensaio”. O ensaio é a atividade de fazer ciência sem prova expressa. Há, portanto, dado objetivo, referencial teórico e metodologia – não há, por outro lado, dado apresentado para além da menção (o dado aqui é o texto da lei, Marco Civil da Internet, e o texto do PL nº 2630, recentemente aprovado pelo GT-NET, da Câmara dos Deputados). Na atividade do ensaio não há, também, esforço metafísico. E justamente por não afastar com um marcador subjetivista, ou incondizente com a realidade fática, que o presente texto aponta para um referencial teórico-metodológico da história dos conceitos, conforme orientações de Reinhart Koselleck (2006, 2014, 2020).

¹ Doutorando em Direito, na Universidade Católica de Pernambuco. Mestre em Direito pela UFPE. Diretor do Instituto de Pesquisa em Direito e Tecnologia do Recife – IP.rec. Advogado, Head de Direito e Tecnologia no Buonora & Oliveira Advocacia.

Sobre história dos conceitos

A história dos conceitos é uma metódica histórica voltada a superar limitações do paradigma da história das ideias – especialmente no refinamento das noções de linguagem (como semântica e pragmática) e a associação correspondente entre eventos/fatos e sua representação conceitual.

Entre as ferramentas de trabalho da história dos conceitos, alguns elementos analíticos ganham maior destaque, como a noção de “espaço de experiência”, que se refere ao acúmulo de eventos e sentidos dados aos eventos e “horizonte de expectativa”, ou seja, o conjunto de projeções que se faz, socialmente, com base em um contexto específico de experiência.

Correlacionados a este movimento de passado-a-futuro, temos a noção de sincronia histórica e de diacronia histórica. A sincronia é como o recorte do “ponto” no fluxo da história, tem contexto tempo-espacial e é onde, por excelência, ocorre o evento novo. A diacronia, por outro lado, é o recorte pelo próprio fluxo geracional, analisando os movimentos correspondentes entre fatos e conceitos/linguagem. Uma diacronia só existe separada das sincronias a partir de uma análise teórica, como evento do mundo as duas dimensões não são separáveis.

Um conceito, portanto, é um ente histórico, que se situa no tempo e espaço e faz conexão entre campos de sentido (campos semânticos). O conceito não é uma entidade abstrata à qual se atribui sentido, fora de uma história. A tensão entre evento (fato) e a linguagem (dimensão do conceito) não pode ser reduzida por imposições “ex machina”.

O conceito de intermediários

Com base na história dos conceitos é possível comparar, no tempo e no espaço, se uma ideia, sua representação e os eventos apontados por essa relação são mantidos ao longo do tempo, ou se modificam. Na forma lógica: ou (1) o evento modifica e com ele a linguagem muda, ou (2) o evento se mantém e com ele a linguagem permanece, ou (3) o evento permanece e a linguagem muda, ou (4) o evento permanece e a linguagem permanece.

No debate evolutivo sobre a Internet e sua regulação pelos Estados, um conceito surge e passa a se destacar como articulador central de sentidos: “provedor”. De acordo com Marcel Leonardi, que posteriormente foi referenciado de forma indireta por decisões do Superior Tribunal de Justiça, provedores seriam compreendidos da seguinte forma:

Provedor de serviços de Internet é o gênero do qual as demais categorias (provedor de backbone, provedor de acesso, provedor de correio eletrônico, provedor de hospedagem e provedor de conteúdo) são espécies. O provedor de serviços de Internet é a pessoa natural ou jurídica que fornece serviços relacionados ao funcionamento da Internet, ou por meio dela. (LEONARDI, 2005, p. 19)

Essa classificação não é a adotada no Marco Civil da Internet (Lei nº 12.965/14), para o qual os provedores se dividem em duas espécies de natureza mais geral – provedores de conexão e de aplicação.

No Marco Civil, conexão é “a habilitação de um terminal para envio e recebimento de pacotes de dados pela internet, mediante a atribuição ou autenticação de um endereço IP”, enquanto aplicação é “conjunto de funcionalidades que podem ser acessadas por meio de um terminal conectado à internet”. Nota-se uma clara opção do legislador em diferenciar a camada da conexão, e seus intermediários, das demais camadas da rede (todas colocadas sob o aspecto de aplicação) com o objetivo de proteger e dar concreção à ubiquidade e neutralidade da rede.

No PL nº 2630 (2020), conhecido como “PL das *Fake News*”, há uma espécie de tripartição, ou talvez quadripartição, da ideia de provedores. De acordo com o artigo segundo, o gênero “provedor”, a partir de seu uso, poderia ser especificado como “de redes sociais”, “ferramentas de busca”, “mensageria instantânea” e, numa hipótese de exceção “enciclopédia online”.

A provisão de aplicação e conexão está, conceitual e juridicamente, ligada à topologia da rede, ao nível, no qual a atividade é realizada. Na nova lei, que possui pretensão legiferante global, como uma regulação complementar ao Marco Civil da Internet (MCI), a classificação muda o seu critério. Se o campo semântico do conceito “provedores” estava ligado à arquitetura da rede, agora há uma propositura de deslocamento voltada aos tipos de serviços prestados.

Do ponto de vista jurídico, em sistema lógico (MIRANDA, 1954), há efetiva antinomia, dando abertura a uma desordem semântica que poderá ter reflexos práticos relevantes – especialmente para o resguardo de uma internet livre e aberta. A falta de unidade do conceito dá espaço ao arbítrio judicial, baseada em interpretações criativas da lei e da Constituição Federal.

Em sendo considerado o MCI como regra geral, anterior, aplicando-se o princípio básico de solução de antinomias e hierarquia das normas no sistema, tem-se, como resultado, a não correspondência entre os campos semânticos de uma mesma palavra “provedores”. Não há, efetivamente, correspondência com a regra geral, anterior do Marco Civil. Se se interprete que o “PL das *Fake News*” aponta para subespécies de provedores de aplicação, não se fundamenta, com efetiva clareza, o motivo de destaque destas subespécies e qual a razão diferenciadora para que elas existam – com uma regra mais especial, do que aquela apresentada no Marco Civil.

Em sendo consideradas ambas as leis (MCI e futura Lei de Liberdade e Transparência) como especiais, criar-se-á nítida antinomia de tipo especial e complexo: aqui, mais uma vez, a ideia seria que apenas as subespécies definidas seriam objeto de incidência da norma.

Numa linguagem aristotélica, – ou, de outra forma, a partir das regras de lógica formal, inseridas na estrutura do ordenamento jurídico, que é sistema de desdobramento dedutivo de normas – não é coerente o apontamento da diferença específica (mensageria, busca, redes sociais) do gênero próximo provedor. Sendo que o próprio ordenamento já classifica o gênero “provedor” a partir da dicotomia *quase* exaustiva de “conexão” e “conteúdo”. O mesmo raciocínio se aplica ao eventual argumento de subespécie: qual a diferença específica dos provedores apontados na futura lei que exija uma distinção e, em existindo, quais as diferenças específicas de cada um dos três/quatro conceitos entre si e eventuais futuras distinções.

Etimologicamente, “provedor” alude a fazer provisão, sob a perspectiva de antecipar eventos futuros. Quem provê, acautela, separa, guarda e fornece. Quem fornece, como provedor, não simplesmente fornece, ainda que de fornecedor, em sentido comum, se trate. Ser provedor, no sentido etimológico e no sentido jurídico do ordenamento brasileiro voltado a regular a Internet, é agir sob cálculo e precaução para abastecer a sociedade com um bem, que no caso é o fluxo de dados

– da sua dimensão de infraestrutura até sua dimensão lógica. Aos primeiros cola-se a diferença específica da “conexão”, aos segundos a diferença específica “aplicação”.

Do ponto de vista técnico, ressalta o casuísmo da classificação legal proposta no Projeto de Lei, sustentada em critério artificial de “serviço prestado” – o serviço já é a provisão em si. Afinal, a cada novo serviço de aplicação ter-se-ia uma nova espécie? Em um ambiente de inovação, o tentar dizer tudo, de forma específica, é tão danoso quanto o dizer tudo sob a perspectiva genérica: nada se diz.

O tema tem reflexos na aplicação pelo Judiciário: um sistema de diferenças específicas exige do intérprete que não ignore as partes gerais de outras normas (MCI) que se apliquem de forma suplementar ao que não está disposto na diferença posta (PL *Fake News*). Em termos de fundamentação, a diferença artificialmente criada impõe um ônus insolúvel ao intérprete e favorece o casuísmo ao criar exceções de exceções, sem o devido sistema. O que já é uma grande chaga do direito brasileiro, ganharia mais um capítulo.

Direito é processo de adaptação social, e enquanto processo, a criação artificiosa de categorias, contra a prática estabilizada, não é, necessariamente um expediente de bom combate a violências ou comportamentos antiéticos. Se um conceito tem história, campo semântico, pragmática e inter-relação com outros conceitos, a eventual alteração *ab ovo* de uma estrutura conceitual significa apenas uma coisa: inadaptação e aplicação defeituosa.

Considerações finais

Como se percebe, estudar o aspecto tecnológico, jurídico e social a partir de uma abordagem histórico-conceitual permite explicitar artificialidades sistêmicas que não encontram substrato ôntico, e, portanto, não encontram respaldo na realidade (o conceito é limítrofe na tensão entre linguagem e fato/evento). O processo legislativo tem dessas coisas: falta de referencialidade a campos semânticos necessários e previamente existentes numa prática (como o caso dos conceitos técnicos) e tentativas de criação “do nada” de estruturas conceituais para regular fatos, sem, efetivamente, conversar com esses fatos.

Ao tempo em que denunciam processos de inadequação legislativa, a história dos conceitos permite perquirir alternativas viáveis. Ao tempo em que alerta sobre uma dificuldade no processo de aplicação da decisão, permite determinar quais os elementos de uma fundamentação conceitual adequada. Há que se pensar conceitualmente ao tentar regular a realidade acelerada da tecnologia e da inovação.

Referências bibliográficas

BRASIL. Câmara dos Deputados. **Projeto de Lei nº 2.630**, de 03 de julho de 2020. Lei Brasileira de Liberdade, Responsabilidade e Transparência na Internet. Brasília: Câmara dos Deputados, 2020. Disponível em: https://www.camara.leg.br/proposicoesWeb/prop_mostrarintegra?codteor=2115425. Acesso em: 12 dez. 2021.

KOSELLECK, R. **Futuro Passado**: Contribuição à semântica dos tempos históricos. Rio de Janeiro: Contraponto: PUC-Rio, 2006.

_____. **Estratos do tempo**: estudos sobre história. Rio de Janeiro: Contraponto: PUC-Rio, 2014.

_____. **História de conceitos**. Rio de Janeiro: Contraponto: 2020.

LEONARDI, M. **Responsabilidade Civil dos Provedores da Internet**. São Paulo: Editora Juarez de Oliveira, 2005.

MIRANDA, F. C. P. **Tratado de Direito Privado**. Tomo 1. Rio de Janeiro: Borsoi, 1954.

ESTRUTURA E FUNCIONAMENTO DA INTERNET

ASPECTOS TÉCNICOS, POLÍTICOS E REGULATÓRIOS

MODERAÇÃO DE CONTEÚDO

FAKE, VERDADE OU IRONIA? O SUBJETIVISMO LINGUÍSTICO E CULTURAL COMO DESAFIO À MODERAÇÃO DE CONTEÚDO

Bruno Ett Bicego¹

RESUMO: Este artigo busca investigar em qual medida o subjetivismo linguístico e cultural pode ser um desafio na moderação de conteúdo. Para se entender o real sentido de algum conteúdo, pode ser necessário investigar uma série de elementos contextuais sobre sua publicação. Além disso, nota-se que muitos termos tradicionalmente utilizados como ofensivos, podem ser ressignificados por comunidades e passarem a ser utilizados com um novo sentido positivo. A identificação do sentido de um conteúdo se torna ainda mais difícil na moderação de conteúdo por meio de ferramentas de inteligência artificial. Nesse cenário, para a busca de uma moderação de conteúdo mais justa, além do contínuo aperfeiçoamento de técnicas de identificação de conteúdo e capacitação dos moderadores, três caminhos parecem fundamentais: transparência, diálogo e a construção de uma arquitetura online voltada para a garantia de direitos dos usuários.

PALAVRAS-CHAVE: Moderação de Conteúdo; Subjetivismo Linguístico; Subjetivismo Cultural; Garantia de Direitos Online; Liberdade de Expressão.

Introdução

“Tomem cachaça com limão e mel para se blindarem do coronavírus”. Mensagens como essa talvez tenham passado pelas redes sociais de muitas pessoas. Eu, pelo menos, recebi algumas vezes tais “conselhos”, no entanto, sempre com um teor claramente humorístico ou irônico, referindo-se à ideia de que o álcool nos ajudaria a esquecer de nossos problemas (algo como: “o Brasil nos obriga a beber”).

¹ Pesquisador no Centro de Ensino e Pesquisa em Inovação da FGV Direito SP. Graduando em Direito pela Pontifícia Universidade Católica de São Paulo (PUC-SP). *Alumnus* da Escola de Formação Pública da Sociedade Brasileira de Direito Público (SBDP).

No entanto, muitas dessas histórias foram difundidas como verdadeiros conselhos médicos. Surgiram correntes de desinformação afirmando que uísque ou vodca seriam eficazes para matar o coronavírus ainda na garganta (PENNAFORT, 2020) ou ainda que a ingestão de aproximadamente 40ml a 50ml de vodca por dia seria uma boa forma de “envenenar” o vírus (ALONSO, 2020).

Apenas para ilustrar a complexidade que a interpretação de certo conteúdo pode gerar, pensemos em algumas categorias para enquadrar eventuais publicações sobre o consumo de destilados durante a pandemia: (i) *posts* que abordam ou incentivam o consumo de álcool como uma forma de lazer; (ii) *posts* humorísticos que retratam ou incentivam o consumo de álcool como uma forma de “superar a situação, esquecendo dos problemas”; (iii) *posts* com imprecisões científicas, que espalham desinformação sobre o consumo de vodca ou cachaça como formas de matar ou envenenar o coronavírus dentro do corpo; (iv) *posts* humorísticos que ironizam tais desinformações sobre “vodca ou cachaça como remédios contra o coronavírus” e, justamente fazendo uso dessa ironia, incentivam o consumo de álcool para “combater o coronavírus”.

Percebemos que a linguagem humana não é simples. Para entender o real sentido de algum conteúdo, pode ser necessário saber o contexto da publicação, o autor, o histórico de publicações do autor, o público-alvo daquele discurso, os elementos visuais utilizados, entre muitas outras coisas. E mesmo com diversas dessas informações contextuais, pode ser impossível determinar se algo é, ou não, irônico ou dotado de caráter humorístico. A simples frase “tomem cachaça com limão e mel para se blindarem do coronavírus” pode ter um significado interpretativo que vai desde uma desinformação que coloca em risco a saúde das pessoas, até uma propaganda bem-humorada de uma loja de destilados.

Interpretação linguística e moderação de conteúdo

Nesse cenário, conforme escreve Grimmelmann (2018), a dificuldade em se distinguir entre o incentivo a uma certa atividade ou a uma paródia dessa atividade é um grande desafio quando se pretende pensar e construir melhores diretrizes de moderação de conteúdo. A situação pode se tornar ainda mais delicada quando o conteúdo sob análise se relaciona com questões sensíveis, como racismo. Segundo explica Tushnet (2019), os moderadores muitas vezes não são do país

onde estão analisando o conteúdo e podem não possuir o conhecimento cultural e circunstancial necessário para distinguir entre uma publicação de alguém denunciando uma injúria racial ou alguém que, se valendo dessa injúria, está causando ainda mais danos.

Se, para seres humanos, a identificação e interpretação de certo conteúdo já é um desafio, a moderação de conteúdo a partir de sistemas de inteligência artificial pode apresentar ainda mais dificuldades. Em média, ferramentas de moderação automática de conteúdo cometem mais erros do que seres humanos, o que já levou empresas, que antes vinham apostando em ferramentas de inteligência artificial, a recuarem e investirem mais na moderação humana (VINCENT, 2020).

Nota-se, conforme explicam Rodriguez e Kurtz (2020), que a moderação de conteúdo por meio de ferramentas de inteligência artificial pode apresentar fortes diferenças a depender do idioma em questão, uma vez que especificidades da língua inglesa, por exemplo, poderiam ser mais bem identificadas, dado um melhor treinamento dos sistemas nesse idioma, enquanto outras regiões poderiam sofrer com um maior número de remoções injustas de conteúdo, visto que a análise realizada seria de menor qualidade.

Para ilustrar a complexidade que a moderação automática de conteúdo pode gerar, vale observar a pesquisa desenvolvida por Gomes, Antonialli e Oliva (2019). Essa pesquisa testou o programa “Perspective”, desenvolvido por uma empresa do grupo Alphabet, que possui como finalidade avaliar, por meio de um sistema de inteligência artificial, qual o nível de toxicidade de um conteúdo. Para os desenvolvedores do programa, a definição de conteúdo tóxico seria um “comentário rude, desrespeitoso ou não-razoável que provavelmente fará com que você abandone uma discussão” (GOMES; ANTONIALLI; OLIVA, 2019). O Perspective classifica automaticamente os conteúdos textuais entre os extremos “muito saudável” e “muito tóxico” apresentando os resultados em uma probabilidade de 0% a 100% daquele conteúdo ser tóxico.

A partir do uso do Perspective, os pesquisadores testaram os níveis de toxicidade de *tweets* postados por famosas *drag queens* dos Estados Unidos e compararam os resultados com uma análise de *tweets* de outros usuários famosos do Twitter, em especial pessoas da extrema direita. Os resultados das análises indicam que um número significativo de perfis de *drag queens* foram erroneamente considera-

dos como potencialmente mais tóxicos que perfis de supremacistas brancos (GOMES; ANTONIALLI; OLIVA, 2019). Ademais, a partir de testes realizados com palavras isoladas, os pesquisadores perceberam que expressões usualmente encontradas em *tweets* de *drag queens* foram consideradas fortemente tóxicas como “BITCH – 98.18%; FAG – 91.94%; SISSY – 83.20%; GAY – 76.10%; LESBIAN – 60.79%; QUEER – 51.03%; TRANSVESTITE – 44.48%” (GOMES; ANTONIALLI; OLIVA, 2019).

A partir dessas observações, os pesquisadores indicam que existem vieses no Perspective, uma vez que as palavras acima foram consideradas tóxicas, independentemente do contexto em que se encontravam (GOMES; ANTONIALLI; OLIVA, 2019):

Isso significa que, independentemente do contexto, palavras como gay, lesbian e queer já são consideradas como significativamente tóxicas, o que aponta a existência de vieses no Perspective. Adicionalmente, muito embora outras palavras como fag, sissy e bitch possam ser usualmente consideradas “tóxicas”, estudos queer sobre linguística indicam como o uso de palavras como essas por membros da comunidade LGBTQ desempenham um papel relevante e positivo.

O uso de ferramentas de moderação automática de conteúdo, como o Perspective ou similares, pode colocar em risco o próprio exercício da liberdade de expressão de membros de comunidades específicas, uma vez que tais ferramentas de inteligência artificial ainda não estariam suficientemente desenvolvidas para analisar o contexto social e cultural de um discurso (GOMES; ANTONIALLI; OLIVA, 2019):

Considerando os códigos de comunicação da comunidade LGBTQ, particularmente das drag queens, o uso do Perspective – bem como de outras tecnologias similares – para moderação de conteúdo nas plataformas de internet poderia prejudicar o exercício da liberdade de expressão de membros da comunidade. Isso porque a inteligência artificial ainda não consegue analisar o contexto social do discurso, muitas vezes falhando em reconhecer o seu valor – como na prática da comunidade LGBTQ de reapropriação de palavras consideradas tóxicas e de seu uso para desenvolver uma “casca grossa” nos membros da comunidade com o objetivo de protegê-los contra quem deseja atacá-los.

A análise do caso do Perspective ilustra a dificuldade de se interpretar certo discurso. Comunidades podem ressignificar palavras tradicionalmente utilizadas como ofensivas e passar a usá-las como um linguajar próprio e positivo. No entanto, paralelamente, outras pessoas podem continuar a usar esses termos de uma maneira pejorativa, fazendo com que uma mesma expressão circule pelas

mídias com significados diametralmente opostos: tanto positivos, como fortemente insultuosos.

Realizar essa avaliação é uma tarefa complexa e, se feita de maneira incorreta, pode colocar em xeque direitos fundamentais dos usuários. Por um lado, quando certo conteúdo é removido injustamente, pode ocorrer a restrição indevida da liberdade de expressão. Em casos mais sérios, quando ocorre o over-removal (remoção excessiva de conteúdo) pode-se prejudicar e invisibilizar uma comunidade inteira no ambiente online. Por outro lado, as plataformas digitais também atuam dentro de uma lógica de responsabilidade corporativa (KLONICK, 2018), na qual elas devem agir proativamente para coibir correntes de discurso de ódio e desinformação, sob pena de permitirem a perpetuação de discursos nocivos a certas comunidades ou à própria democracia.

Perspectivas para a garantia de direitos online

Observados alguns desafios que o subjetivismo linguístico e cultural pode gerar na moderação de conteúdo, vale apontar algumas direções que podem ajudar na proteção de direitos online e na construção de uma moderação de conteúdo mais justa. Além do contínuo aperfeiçoamento de técnicas de identificação de conteúdo e capacitação dos moderadores, três caminhos parecem fundamentais: (i) transparência, (ii) diálogo e (iii) a construção de uma arquitetura online voltada para a garantia de direitos dos usuários.

Seja na moderação de conteúdo realizada por seres humanos, seja na moderação por meio de ferramentas de inteligência artificial, é fundamental que as plataformas digitais atuem com transparência. Um exemplo é a adoção de relatórios periódicos de transparência, com informações sobre quais tipos de conteúdo foram removidos, sua quantidade e os motivos que levaram a sua exclusão. Países e blocos econômicos como União Europeia (EUROPEAN COMMISSION, 2017) e Reino Unido (REINO UNIDO, 2020) já se manifestaram reiterando a importância de tais relatórios.

A construção de um diálogo transparente e aberto é necessária para se pensar em soluções e melhorias na moderação de conteúdo. Empresas, academia, sociedade civil e órgãos reguladores devem trabalhar em conjunto, de modo a criar

um leque de visões complementares que possam auxiliar no desenvolvimento de políticas e princípios para a moderação de conteúdo online.

Por fim, a arquitetura do ambiente online deve ser construída de maneira a gerar mais segurança e garantir ao máximo proteção de direitos dos usuários. Um exemplo são os mecanismos para denunciar conteúdos supostamente danosos, que devem ser estruturados para que a outra parte também possa realizar seu direito de defesa e não ter seu conteúdo removido, caso ele não viole as políticas de comunidade da plataforma.

A moderação de conteúdo é extremamente complexa e os desafios só aumentam quando pensamos no subjetivismo linguístico e cultural que permeiam todo o diálogo humano. Somente com transparência e um diálogo contínuo será possível aumentar a garantia de direitos aos usuários de plataformas digitais.

Referências bibliográficas

ALONSO, L. Países recomendam masturbação e vodca como medidas para contenção do coronavírus. **Folha de S. Paulo**, 2020. Disponível em: <https://www1.folha.uol.com.br/mundo/2020/03/paises-recomendam-masturbacao-e-vodca-como-medidas-para-contencao-do-coronavirus.shtml>. Acesso em: nov. 2021

EUROPEAN COMMISSION. Tackling Illegal Content Online: Towards an enhanced responsibility of online platforms. **European Commission**, 2017. Disponível em: <https://digital-strategy.ec.europa.eu/en/library/communication-tackling-illegal-content-online-towards-enhanced-responsibility-online-platforms>. Acesso em: nov. 2021.

GOMES, A.; ANTONIALLI, D.; OLIVA, T. D. Drag queens e Inteligência Artificial: computadores devem decidir o que é 'tóxico' na internet? **InternetLab**, 2019. Disponível em: <https://www.internetlab.org.br/pt/liberdade-de-expressao/drag-queens-e-inteligencia-artificial-computadores-devem-decidir-o-que-e-toxico-na-internet>. Acesso em: nov. 2021.

GRIMMELMANN, J. The Platform is the Message. **Cornell Legal Studies Research Paper** No. 18-30, 2018. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3132758. Acesso em: jun. 2021.

KLONICK, K. The New Governors: The People, Rules, and Processes Governing Online Speech. **Harvard Law Review**, Cambridge, vol. 131:1598, 2018. Disponível em: https://harvardlawreview.org/wp-content/uploads/2018/04/1598-1670_Online.pdf. Acesso em: nov. 2021.

PENNAFORT, R. É #FAKE que beber vodca ou uísque mate o coronavírus na garganta. **G1**, 2020. Disponível em: <https://g1.globo.com/fato-ou-fake/noticia/2020/07/10/e-fake-que-beber-vodca-ou-uisque-mate-o-coronavirus-na-garganta.ghtml>. Acesso em: nov. 2021.

REINO UNIDO. Online Harms White Paper: Full Government Response to the consultation. **Government of the United Kingdom**, 2020. Disponível em: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/944310/Online_Harms_White_Paper_Full_Government_Response_to_the_consultation_CP_354_CCS001_CCS1220695430-001_V2.pdf. Acesso em: nov. 2021

RODRIGUES, G.; KURTZ, L. Transparência sobre moderação de conteúdo em políticas de comunidade. **Instituto de Referência em Internet e Sociedade**, Belo Horizonte, 2020. Disponível em: <https://irisbh.com.br/wp-content/uploads/2020/10/Transpar%C3%Aancia-sobre-modera%C3%A7%C3%A3o-de-conte%C3%BAdo-em-pol%C3%ADticas-de-comunidade-IRIS.pdf>. Acesso em: nov. 2021.

TUSHNET, R. Content Moderation in an age of extremes. **Case Western Reserve Journal of Law, Technology & the Internet**, vol. 10, n. 1, 2019. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3503169. Acesso em: 31 mai. 2021.

VINCENT, J. YouTube brings back more human moderators after AI systems over-censor. **The Verge**, 2020. Disponível em: <https://www.the-verge.com/2020/9/21/21448916/youtube-automated-moderation-ai-machine-learning-increased-errors-takedowns>. Acesso em: nov. 2021.

FACEBOOK PAPERS E A TRANSPARÊNCIA DE ALGORITMOS A FAVOR DO DEBATE REGULATÓRIO: ESTUDO DE CASO

Giovanna Wolf Tadini¹

RESUMO: Com o fortalecimento de discussões sobre desinformação e discurso de ódio em plataformas digitais, gigantes de tecnologia passaram a ser cobradas pela saúde dos espaços virtuais. Sendo assim, este artigo objetiva discutir de que modo informações sobre o funcionamento de redes sociais – principalmente detalhes relacionados ao trabalho dos algoritmos de classificação do feed de notícias e a mecanismos de moderação de conteúdo – poderiam contribuir para o avanço do debate regulatório no setor de tecnologia, melhorando, assim, a segurança das plataformas digitais. Esta pesquisa caracteriza-se como sendo do tipo bibliográfica e documental, de natureza qualitativa, e adota como método o estudo de caso. O caso em questão, que ficou conhecido como Facebook Papers, teve início em setembro de 2021 com o vazamento de documentos internos da Meta e ganhou nova dimensão quando Frances Haugen, ex-funcionária da empresa, identificou-se como a fonte dos vazamentos. O artigo conclui que o episódio descrito trouxe uma nova camada de informação a apontamentos já feitos na literatura sobre possíveis efeitos tóxicos do Facebook, como o favorecimento de conteúdos inflamatórios no feed de notícias. O caso reforça a importância da transparência de algoritmos na construção de plataformas digitais saudáveis, por meio de regulações condizentes com as entranhas da operação de empresas de tecnologia.

PALAVRAS-CHAVE: Facebook Papers; Transparência; Algoritmos; Regulação; Rede Social

¹ Jornalista formada pela Escola de Comunicações e Artes da Universidade de São Paulo. Repórter de Tecnologia e Inovação do jornal O Estado de S. Paulo.

Introdução

Com o fortalecimento de discussões sobre desinformação e discurso de ódio em plataformas digitais, gigantes de tecnologia passaram a ser cobradas pela saúde dos espaços virtuais. Nesse processo, as companhias entraram na mira de agentes regulatórios no mundo todo, que estudam como garantir a segurança dos usuários em plataforma digitais e também como preservar a concorrência em um mercado cujas empresas criaram impérios nas últimas décadas.

De acordo com Almeida, Filgueiras e Doneda (2021), plataformas digitais como Facebook, Twitter, YouTube, WhatsApp e Instagram têm sido culpadas por causar turbulências sociais e políticas em diferentes partes do globo.

Seus motores são executados por algoritmos proprietários que classificam grandes quantidades de conteúdo dos usuários e organizam sua exposição online. Como as plataformas e serviços de rede social permeiam a vida de 3,8 bilhões de pessoas ou 49% da população mundial, é esperado que eles criem mecanismos para garantir a disponibilidade de conteúdos seguros e proteger os usuários e comunidades contra conteúdos nocivos (ALMEIDA; FILGUEIRAS; DONEDA, 2021, p. 13).

A Meta, por ser dona de vários aplicativos (Facebook, WhatsApp e Instagram), está na berlinda dos debates regulatórios. Além disso, a companhia de Mark Zuckerberg esteve no centro de escândalos como o caso Cambridge Analytica, em 2018, em que foram usados indevidamente dados de 87 milhões de usuários da rede social para influenciar as eleições americanas.

Assim, este artigo objetiva discutir de que modo informações sobre o funcionamento de redes sociais como o Facebook – principalmente detalhes relacionados ao trabalho dos algoritmos de classificação do feed de notícias e a sistemas de moderação de conteúdo – poderiam contribuir para o avanço do debate regulatório no setor de tecnologia, melhorando, assim, a segurança das plataformas digitais.

Em termos metodológicos, esta pesquisa caracteriza-se como sendo do tipo bibliográfica e documental, de natureza qualitativa, e adota como método o estudo de caso. O artigo toma como base o episódio dos Facebook Papers, que teve início em setembro de 2021 com o vazamento de documentos internos da empresa de Mark Zuckerberg.

Facebook Papers

Os arquivos da Meta vieram à tona em setembro de 2021 com reportagens do jornal americano Wall Street Journal. Uma das pesquisas internas da empresa trazida pela publicação mostrou que 32% das meninas que se sentiam mal com o próprio corpo tinham essa situação agravada ao acessar o Instagram (WELLS; HORWITZ; SEETHARAMAN, 2021).

As denúncias ganharam nova dimensão quando Frances Haugen, ex-funcionária da Meta, veio a público se identificar como a fonte dos vazamentos. O pacote de arquivos, apelidado de Facebook Papers pela cobertura midiática, foi enviado à Securities and Exchange Commission, órgão regulador das empresas listadas em bolsa nos Estados Unidos, e analisado por um consórcio internacional de veículos jornalísticos. A partir do dia 22 de outubro, a imprensa passou a divulgar quase que diariamente reportagens com conteúdos de documentos internos da Meta.

Uma das matérias, publicada no Washington Post em 26 de outubro de 2021, mostra como a fórmula do Facebook usada para hierarquizar publicações no feed de notícias estimulou o ódio e a desinformação na plataforma. De acordo com o jornal, a partir de 2017, os algoritmos de classificação da rede social passaram a tratar reações de emojis, incluindo o botão de “raiva”, com valores cinco vezes maiores que o botão “curtir”. Os próprios pesquisadores do Facebook suspeitaram que o mecanismo levaria conteúdos mais emocionais e provocativos aos feeds de notícias dos usuários (MERRILL; OREMUS, 2021).

Ainda sobre o funcionamento de algoritmos, uma reportagem da NBC News, publicada em 22 de outubro, trouxe detalhes do relatório “Carol’s Journey to QAnon”, elaborado pela empresa em 2019. Pesquisadores da Meta criaram uma conta fictícia de uma usuária chamada Carol Smith, descrita no perfil como uma mãe politicamente conservadora residente do Estado da Carolina do Norte. A pesquisa mostra que, embora a pessoa imaginária nunca tenha expressado interesse em teorias da conspiração, em apenas dois dias o Facebook recomendou que ela participasse de grupos dedicados ao QAnon, movimento conspiratório criado pela extrema-direita americana pró-Trump (ZADROZNY, 2021).

Além disso, os Facebook Papers mostram que os sistemas da empresa são “mal equipados” para combater conteúdo nocivo viral. De acordo com reportagem do jornal O Estado de S. Paulo, publicada em 14 de novembro, a empresa de Mark

Zuckerberg atua de forma reativa ao mau uso da plataforma e não garante o mesmo nível de moderação de conteúdo para todos os seus usuários ao redor do mundo (ARIMATHEA; ROMANI; WOLF, 2021).

Debate regulatório

Os Facebook Papers atestaram, com documentos internos da empresa, questionamentos levantados há algum tempo. Já era sabido, por exemplo, que, além de gerar cliques e visualizações, “o feed do Facebook privilegia conteúdos inflamatórios, criando um ciclo de estímulo-resposta onde a expressão de indignação se torna mais fácil e até normalizada” (MUNN, 2020, p. 8). Também já eram discutidas questões problemáticas nos modelos de moderação de conteúdo da rede social, incluindo decisões tomadas pelo Facebook em Mianmar e a circulação de conteúdos maliciosos relacionados à covid-19, como mostram Sablosky (2021) e Velásquez *et al.* (2021), respectivamente.

Com os arquivos vazados, porém, há detalhes sobre o design da plataforma. Trata-se de um avanço na transparência de algoritmos, amplamente discutida na literatura, como defendem Katzenbach e Ulbricht (2019).

Viés e justiça, transparência e agência humana são questões importantes que devem ser abordadas sempre que os sistemas algorítmicos estão profundamente integrados aos processos organizacionais, independentemente do setor ou aplicação específica (KATZENBACH; ULBRICHT, 2019, p. 11).

Tendo em vista a importância de empresas como o Facebook se envolverem mais ativamente em “cidadania corporativa” (GRYGIEL; BROWN, 2019, p. 13), detalhes sobre o desenho da plataforma e seus algoritmos se mostram um caminho de resposta para a pressão regulatória. É nisso que se baseiam os argumentos da delatora Frances Haugen (HAO, 2021), que, em depoimento no Senado americano em 5 de outubro de 2021, defendeu que a saída para consertar o Facebook está na mudança dos algoritmos – e nas informações sobre eles. Ela afirma que não é necessário desmembrar a Meta, mas sim transformar a lógica de classificação algorítmica – uma das bandeiras de Frances é a volta do feed de notícias em ordem cronológica. Esse movimento de transparência poderia inclusive reverberar em toda a indústria da tecnologia (ARIMATHEA; WOLF, 2021).

Considerações finais

O episódio dos Facebook Papers trouxe uma nova camada de informação a apontamentos já feitos na literatura sobre possíveis efeitos tóxicos do Facebook, como o favorecimento de conteúdos inflamatórios no feed de notícias. O caso reforça a importância da transparência de algoritmos na construção de plataformas digitais saudáveis, por meio de regulações condizentes com as entranhas da operação de empresas de tecnologia.

Referências bibliográficas

ALMEIDA, V.; FILGUEIRAS, F.; DONEDA, D. The ecosystem of digital content governance. **IEEE Internet Computing**, v. 25, n. 3, p. 13-17, maio/jun. 2021. Disponível em: <https://ieeexplore.ieee.org/document/9351745>. Acesso em: 30 nov. 2021.

ARIMATHEA, B.; ROMANI, B.; WOLF, G. Facebook Papers: sistemas da empresa são 'mal equipados' para combater conteúdo nocivo viral. **O Estado de S. Paulo**, São Paulo, 14 nov. 2021. Link Estadão. Disponível em: <https://link.estadao.com.br/noticias/empresas,facebook-papers-sistemas-da-empresa-sao-mal-equipados-para-combater-conteudo-nocivo-viral,70003896808>. Acesso em: 30 nov. 2021.

ARIMATHEA, B.; WOLF, G. Nova crise do Facebook deve forçar indústria da tecnologia a sair das sombras. **O Estado de S. Paulo**, São Paulo, 10 out. 2021. Link Estadão. Disponível em: <https://link.estadao.com.br/noticias/empresas,nova-crise-do-facebook-deve-forcar-industria-da-tecnologia-a-sair-das-sombras,70003863836>. Acesso em: 30 nov. 2021.

GRYGIEL, J.; BROWN, N. Are social media companies motivated to be good corporate citizens? Examination of the connection between corporate social responsibility and social media safety. **Telecommunications Policy**, v. 43, n. 5, p. 445-460, 2019. Disponível em: <https://doi.org/10.1016/j.telpol.2018.12.003>. Acesso em: 30 nov. 2021.

HAO, K. The Facebook whistleblower says its algorithms are dangerous: here's why. **MIT Technology Review**, Cambridge, 5 out. 2021. Disponível em:

<https://www.technologyreview.com/2021/10/05/1036519/facebook-whistleblower-frances-haugen-algorithms/>.

Acesso em: 30 nov. 2021.

KATZENBACH, C.; ULBRICHT, L. Algorithmic governance. **Internet Policy Review**, v. 8, n. 4, p. 1-18, 2019. Disponível em: <http://policyreview.info/concepts/algorithmic-governance>. Acesso em: 30 nov. 2021.

MERRILL, J. B.; OREMUS, W. Five points for anger, one for a 'like': how Facebook's formula fostered rage and misinformation. **The Washington Post**, Washington, 26 out. 2021. Technology. Disponível em: <https://www.washingtonpost.com/technology/2021/10/26/facebook-angry-emoji-algorithm/>. Acesso em: 30 nov. 2021.

MUNN, L. Angry by design: toxic communication and technical architectures. **Humanities and Social Sciences Communications**, v. 7, n. 1, artigo 53, p. 1-11, 2020. Disponível em: <https://www.nature.com/articles/s41599-020-00550-7>. Acesso em: 30 nov. 2021.

SABLOSKY, J. Dangerous organizations: Facebook's content moderation decisions and ethnic visibility in Myanmar. **Media, Culture & Society**, v. 43, n. 6, p. 1017-1042, 2021. Disponível em: <https://journals.sagepub.com/doi/abs/10.1177/0163443720987751>. Acesso em: 30 nov. 2021.

VELÁSQUEZ, N. *et al.* Online hate network spreads malicious COVID-19 content outside the control of individual social media platforms. **Scientific Reports**, v. 11, n. 1, artigo 11549, 2021. Disponível em: <https://www.nature.com/articles/s41598-021-89467-y>. Acesso em: 30 nov. 2021.

WELLS, G.; HORWITZ, J.; SEETHARAMAN, D. Facebook knows Instagram is toxic for teen girls, company documents show. **The Wall Street Journal**, Nova York, 14 set. 2021. Disponível em: <https://www.wsj.com/articles/facebook-knows-instagram-is-toxic-for-teen-girls-company-documents-show-11631620739>. Acesso em: 30 nov. 2021.

ZADROZNY, B. 'Carol's Journey': what Facebook knew about how it radicalized users. **NBC News**, Nova York, 22 out. 2021. Tech News. Disponível em: <https://www.nbcnews.com/tech/tech-news/facebook-knew-radicalized-users-rcna3581>. Acesso em: 30 nov. 2021.

LIBERDADE DE OPINIÃO AINDA É PERMITIDA NA INTERNET?

Humberto Sturiale Sartiní¹

RESUMO: Apresentar alguns pontos das políticas de utilização das redes sociais com o intuito de cercear o direito de liberdade de expressão e opinião dos usuários.

PALAVRAS-CHAVE: Liberdade de opinião; Internet; Redes sociais; Transparência; Conflito.

Liberdade de opinião

Contextualizando liberdade de opinião

No ano de 1948, quando foi adotada e proclamada pela Assembleia Geral das Nações Unidas (resolução 217 A III) a Declaração Universal dos Direitos Humanos (ORGANIZAÇÃO DAS NAÇÕES UNIDAS, 2021), já externava a preocupação de assegurar o direito à liberdade de opinião ao ser humano, conforme visto no artigo 19:

Todo ser humano tem direito à liberdade de opinião e expressão; esse direito inclui a liberdade de, sem interferência, ter opiniões e de procurar, receber e transmitir informações e ideias por quaisquer meios e independentemente de fronteiras.

Mas afinal de contas, o que é liberdade de opinião? É um direito inerente à pessoa humana que garante a permissão de exteriorização do pensamento, independente da forma e canal utilizado para esse fim. É importante ressaltar que existem limites para essa liberdade, como descrito no artigo 5º, inciso X da Constituição:

¹ Profissional da Internet com mais de 20 anos de experiência, tendo passagem por provedores, operadoras e empresa de consultoria de Internet. Bacharel em Informática, formado pela UEPG, especialização em Gestão Pública Municipal pela UFPR. Possui diversas certificações e cursos relativos à área de atuação.

X - são invioláveis a intimidade, a vida privada, a honra e a imagem das pessoas, assegurado o direito a indenização pelo dano material ou moral decorrente de sua violação;

Com o conceito apresentado fica a pergunta: atualmente na Internet é possível exercer essa liberdade de opinião?

Utilização da Internet no Brasil

O projeto TIC Domicilio (Cetic.BR, 2021), do ano de 2020, nos mostra que no Brasil mais de 81% da população possui acesso à internet, o que representa mais de 152 milhões de usuários. Dentre as atividades mais realizadas foram elencadas:

- Envio de mensagens instantâneas (93% dos usuários)
- Conversa por chamada de voz ou vídeo (80% dos usuários)
- Uso de redes sociais (72%)

Com o conhecimento desses números, fica mais claro de entender o poder de influência e persuasão que as redes sociais, utilizadas por 72% dos usuários de internet, pode possuir sobre as pessoas.

Nesse mesmo estudo foi apresentado que nos últimos anos, com destaque nos últimos anos devido a pandemia de Covid 19, vem crescendo o uso da Internet pelos brasileiros e por consequência o uso das redes sociais. Mais especificamente do Brasil onde ocorre uma polarização política cada vez maior, as redes sociais, além de outros meios, refletem esse cenário.

Redes sociais

Com essa massificação das redes sociais, que usualmente são ou fazem parte de grandes conglomerados empresariais, tornaram-se uma nova fonte de informação para os usuários. É importante ressaltar que muito da informação que divulgada nessas plataformas não necessariamente possuem uma fonte ou a geração ocorreu por alguma fonte fidedigna. Esse cenário criou um espaço para as redes sociais iniciarem a moderação do conteúdo publicado.

No estudo Transparência na moderação de conteúdo – Tendências regulatórias nacionais, desenvolvido pelo Instituto de Referência em Internet e Sociedade (IRIS), explora de uma forma prática os padrões, lacunas e tendências nas obrigações de transparência sobre moderação de conteúdo, utilizando não somente o Brasil como modelo, mas também diversos outros países.

Um ponto importante apresentado na pesquisa são os 04 fatores que impactam diretamente na moderação de conteúdo, que são “Filosofia da empresa”, “Conformidade regulatória”, “Maximização de lucros” e “Opinião pública”.

Os fatores apresentados podem ser conflitantes entre si em algumas áreas, porém um que poderia trazer o maior impacto é a filosofia da empresa. Esse tipo de conceito é muito difícil de ser entendido e apresentado aos usuários, uma vez que varia de cada uma das plataformas e sempre é necessário lembrar que o interesse dos fundadores, conselhos ou outros responsáveis pelas empresas devem ser atendidos.

Para não terem associadas suas imagens ao conteúdo categorizado por essas empresas como danoso, criaram algumas práticas de intervenção no conteúdo, tais como:

- Remoção: remoção completo do conteúdo publicado;
- Indisponibilização: permitir o acesso de forma limitada temporal ou geograficamente;
- Restrição: ocultar o conteúdo ou criar restrição por algum parâmetro;
- Sinalização: sinalizar que determinado conteúdo está associado a algo;
- Ranqueamento: apresentar conteúdos mais “relevantes” para os usuários;

Essas práticas não devem ser consideradas negativas, porém o ponto importante a ser questionado é a transparência que cada plataforma tratar o conteúdo. Importante ressaltar que em muitos casos temos termos genéricos ou abrangentes para servir como base dessas práticas.

Casos práticos

Se o usuário quiser tentar entender como as grandes empresas lidam com as práticas de intervenção é possível acessar páginas com explicação de como operar, porém nesses sites existem textos conflitantes e com termos genéricos.

Na página de transparência do Facebook temos a seguinte frase:

Reconhecemos que as pessoas às vezes compartilham conteúdos que incluem discursos de ódio de outra pessoa para condená-la ou aumentar a conscientização. Em outros casos, o discurso que poderia violar nossas normas pode ser usado de forma autorreferencial ou empoderadora. (FACEBOOK, [2021?])

No Instagram, que também possui uma página explicando a transparência aplicado, temos o seguinte texto:

Queremos promover uma comunidade diversificada e positiva. Removemos conteúdo que contenha ameaças reais ou discurso de ódio, conteúdo que ataque indivíduos privados com a intenção de degradá-los ou constrangê-los. Também removemos informações pessoais com o intuito de chantagear ou assediar alguém e mensagens indesejadas enviadas repetidamente. Geralmente, permitimos discussões fortes sobre pessoas que são noticiadas na mídia ou que possuem um público mais amplo devido à profissão ou às atividades de sua escolha. (INSTAGRAM, [2021?])

O Twitter, outra grande plataforma de rede social, no texto sobre elaboração de políticas e filosofias de medidas corretivas, temos:

Para ajudar a garantir que as pessoas tenham a oportunidade de ver todos os lados de uma questão, pode haver ocasiões raras em que permitimos que um conteúdo ou comportamento controverso, que de outra forma violaria as Regras do Twitter, permaneça em nosso serviço porque acreditamos que exista um interesse público legítimo em sua disponibilidade. Cada situação é avaliada caso a caso e por fim decidida por uma equipe interfuncional. (TWITTER, [2021?])

Por outro lado, temos ferramentas que possuem uma política mais permissiva de publicações, como são os casos do Telegram e GETTR. Na página inicial do GETTR é possível visualizar a temática da referida rede social:

Uma nova plataforma de rede social fundamentada nos princípios da liberdade de expressão, pensamento independente e na rejeição de censuras políticas e da “cultura de cancelamento”. (GETTR, [2022?])

Já o Telegram, o qual vem sofrendo tentativas de bloqueio por parte de autoridades brasileiras, também possui uma política mais aberta para as publicações, inclusive permitindo grupos maiores que seus concorrentes e funções complementares que diversas outras plataformas não possuem. Essa ferramenta, que começou a ser mais utilizada no país depois do “Apagão do WhatsApp”, em outubro de 2021 (G1, 2021), está sofrendo grandes pressões de determinados grupos para que seja bloqueada.

Um dos pontos preocupantes, é que nas notícias veiculadas informam que o Telegram também é bloqueado em 11 países e quando procuramos a relação de países, que se diga de passagem é muito difícil de encontrar, temos diversos países ditatoriais ou que possuem um forte desejo de controlar o pensamento e a expressão de seu povo, tais como: Irã, China, Bahrein, Cuba e outros países que estavam passando por forte pressão popular para que os direitos humanos fossem válidos nesses países. (WIKIPEDIA, 2022).

Aqui é válido um questionamento que todos devem se fazer: é isso que queremos para o Brasil?

É de extrema importância lembrar que no Marco Civil da Internet (Brasil, 2014) temos diversos pontos explícitos que visam garantir a liberdade de expressão, como visto abaixo:

Art. 2º A disciplina do uso da internet no Brasil tem como fundamento o respeito à liberdade de expressão, bem como:

II - os direitos humanos, o desenvolvimento da personalidade e o exercício da cidadania em meios digitais;

III - a pluralidade e a diversidade;

Art. 3º A disciplina do uso da internet no Brasil tem os seguintes princípios:

I - garantia da liberdade de expressão, comunicação e manifestação de pensamento, nos termos da Constituição Federal;

Art. 8º A garantia do direito à privacidade e à liberdade de expressão nas comunicações é condição para o pleno exercício do direito de acesso à internet.

Não se pode confundir liberdade de expressão como um salvo conduto que irá permitir a propagação de informações não verdadeiras, ofensas, ataques a honra ou a dignidade da pessoa humana.

Também pode-se pensar que o uso gratuito das redes sociais dá o direito as referidas empresas de executarem as ações da forma que melhor convir, sem a obrigatoriedade de atender os termos de uso das plataformas ou legislações específicas de cada localidade. É necessário lembrar que “quando um serviço é gratuito, você é o produto” nos remete ao caráter de relação de consumo entre as plataformas e os usuários.

Diante desses pontos fica claro que as referidas empresas estão criando situações em que o cerceamento prévio da liberdade de expressão, seja através de algoritmos, agências de checagem, denúncias de usuários ou outros meios, caracteriza uma forma de não permitir uma manifestação de pensamento plural e diversa do viés das empresas.

Conclusão

A liberdade de opinião é um direito adquirido da pessoa humana e deve ser respeitado em qualquer cenário. Em alguns casos a opinião apresentada em uma rede social pode ir contra as políticas daquela plataforma e com isso ser moderada.

É louvável que as redes sociais possuam políticas de moderação de conteúdo para que seja possível aplicar ações visando a diminuição ou extinção de danos a qualquer pessoa, porém a forma que utilizam essas ferramentas não são transparentes.

Para garantir uma maior liberdade de opinião é necessário que as plataformas possuam uma transparência maior em suas políticas, para evitar entendimentos incorretos, dúbios ou que se aplicam em certos cenários que não possuem definição.

Referências bibliográficas

BRASIL. Constituição (2014). **Lei nº 12965**, de 23 de abril de 2014. Lei: Estabelece princípios, garantias, direitos e deveres para o uso da Internet no Brasil. Brasília, Disponível em: http://www.planalto.gov.br/ccivil_03/ato2011-2014/2014/lei/l12965.htm. Acesso em: 16 jan. 2022.

FACEBOOK. **Discurso de ódio**. [2021?]. Disponível em: <https://transparency.fb.com/pt-br/policias/community-standards/hate-speech/>. Acesso em: 11 dez. 2021.

G1. Telegram registra mais de 70 Milhões de usuários novos durante pane do Whatsapp. **G1**, 05 out. 2021. Disponível em: <https://g1.globo.com/tecnologia/noticia/2021/10/05/telegram-registra-mais-de-70-milhoes-de-usuarios-novos-durante-pane-do-whatsapp.ghtml>. Acesso em: 06 fev. 2022.

GETTR. **Página Inicial**. [2022?] Disponível em: <https://www.gettr.com/onboarding>. Acesso em: 06 fev. 2022.

INSTAGRAM. **Diretrizes da Comunidade**. [2021?]. Disponível em: https://help.instagram.com/477434105621119/?helpref=hc_fnav. Acesso em: 11 dez. 2021

KURTZ, L. P.; DO CARMO, P. R. R.; VIEIRA, V. B. R. **Transparência na moderação de conteúdo: tendências regulatórias nacionais**. Belo Horizonte: Instituto de Referência em Internet e Sociedade, 2021. Disponível em: <https://bit.ly/3xjAUka>. Acesso em: 11 dez. 2021.

NÚCLEO DE INFORMAÇÃO E COORDENAÇÃO DO PONTO BR (NIC.br). Publicado em 18/08/2021. Pesquisa sobre o uso das tecnologias de informação e comunicação nos domicílios brasileiros: pesquisa TIC Domicílios (Edição COVID-19 - Metodologia adaptada), ano 2020. Disponível em: <https://cetic.br/pt/arquivos/domicilios/2020/domicilios/>. Acesso em: 08 dez. 2021

ORGANIZAÇÃO DAS NAÇÕES UNIDAS. **Declaração Universal dos Direitos Humanos**, 1948. Disponível em: <https://www.unicef.org/brazil/declaracao-universal-dos-direitos-humanos>. Acesso em: 11 dez. 2021.

TWITTER. **Abordagem de elaboração de políticas e filosofia de medidas corretivas do Twitter.** [2021?]. Disponível em: <https://help.twitter.com/pt/rules-and-policies/enforcement-philosophy>. Acesso em: 11 dez. 2021.

WIKIPEDIA. **Government censorship of Telegram Messenger.** 2022. Disponível em: https://en.wikipedia.org/wiki/Government_censorship_of_Telegram_Messenger. Acesso em: 06 fev. 2022.

MODERAÇÃO DE CONTEÚDO NA INTERNET – UMA BREVE AVALIAÇÃO DOS TERMOS DE USO DO INSTAGRAM

Rodrigo Jae Hyun Yu¹

RESUMO: A moderação de conteúdo tem tido cada vez mais relevância, principalmente pelo crescimento do uso das redes sociais, com seu potencial de influenciar opiniões, divulgação de informação em massa, entre outros fatores. Sendo assim, o tema tem sido objeto de regulação e estudo por parte de governos e órgãos dedicados ao assunto. Como condição para a utilização das redes sociais, os provedores dessas aplicações de internet exigem a adesão dos usuários aos Termos de Uso da rede, que governam a utilização dos serviços, incluindo a delimitação dos materiais que poderão ser compartilhados nelas. Esses Termos devem ser claros sobre suas políticas de moderação de conteúdo, identificando os materiais passíveis de remoção, bem como os procedimentos e mecanismos de recursos para usuários que tiveram conteúdos removidos com base nesses Termos. Assim, neste pequeno ensaio, será feita uma breve avaliação dos Termos de Uso de uma das maiores redes sociais da atualidade, o Instagram, a fim de se observar quão transparente suas regras são em relação a esses quesitos.

PALAVRAS-CHAVE: Moderação; Conteúdo; Internet; Instagram; Diretrizes

Introdução

A recente edição da Medida Provisória nº 1.068 em setembro de 2021 acendeu discussões sobre a moderação de conteúdo feita por provedores de aplicações de internet, como as redes sociais. A medida, que modificava o Marco Civil da Internet, foi rejeitada alguns dias depois e trazia entre suas principais previsões a

¹ Aluno da Graduação da Escola de Direito de São Paulo da Fundação Getúlio Vargas. Este ensaio foi produzido como fechamento do Curso Livre “Estrutura e Funcionamento da Internet: Aspectos Técnicos, Políticos e Regulatórios”, promovido pelo Centro de Ensino e Pesquisa em Inovação (CEPI) da FGV Direito SP, com apoio do Twitter Brasil.

de que “em observância à liberdade de expressão, comunicação e manifestação de pensamento, a exclusão, a suspensão ou o bloqueio da divulgação de conteúdo gerado por usuário somente poderá ser realizado com justa causa e motivação” (art. 8-C). O dispositivo ainda elencava, em seu inciso II, as hipóteses do que pode ser considerado como justa causa. Entre elas, alguns exemplos são: conteúdos que contenham nudez ou representações de atos sexuais (alínea a), que fazem apologia ao crime (alíneas b e c) ou à violência (alínea d e f), que possuam material discriminatório e preconceituoso (alínea d), que ameacem a segurança pública (alínea h), que violem propriedade intelectual (alínea i); além da hipótese de cumprimento de determinações judiciais (inciso IV) e do requerimento de ofendidos (ou representantes legais e até mesmo herdeiros), quando se verificar violação à privacidade, intimidade ou honra, entre outras causas (inciso III).

Vale apontar que mesmo antes da edição dessa Medida Provisória, o Marco Civil já possuía previsões sobre moderação de conteúdo de provedores de aplicações de internet, dentro da seção sobre responsabilidade por danos decorrentes de conteúdo gerado por terceiros. Em seu art. 19, sob a premissa de que a liberdade de expressão deve ser assegurada e a censura impedida, é estabelecido que o provedor de aplicações só será responsabilizado se não tomar providências dentro do prazo para remover conteúdo apontado como infringente por determinação judicial. O art. 21, por sua vez, determina que o provedor de aplicações de internet que disponibilize conteúdo gerado por terceiros (o que é o caso das redes sociais),

será responsabilizado subsidiariamente pela violação da intimidade decorrente da divulgação, sem autorização de seus participantes, de imagens, de vídeos ou de outros materiais contendo cenas de nudez ou de atos sexuais de caráter privado quando, após o recebimento de notificação pelo participante ou seu representante legal, deixar de promover, de forma diligente, no âmbito e nos limites técnicos do seu serviço, a indisponibilização desse conteúdo.

Tendo em vista a relevância desse tema, deve-se fazer uma análise da atuação desses provedores de aplicações de internet, mais especificamente, das redes sociais (em que a adição de conteúdo por usuários é mais acentuada), na moderação dos conteúdos divulgados em suas plataformas, observando o tipo de prática exercida por elas, bem como sua interação com a legislação nacional e recomendações de órgãos como o Internet Society. Assim, nesse ensaio será objeto de estudo o Instagram, uma das maiores redes sociais atualmente, que conta com cerca de 500 milhões de usuários diários (DEAN, 2021), e sua política de moderação de conteúdo.

1. Moderação de conteúdo nas redes sociais

De modo geral, as redes sociais, assim como outros serviços digitais, apresentam Termos de Uso com os quais os usuários devem concordar para utilizar os serviços prestados. Esses Termos costumam descrever algumas regras que regerão os limites do comportamento dos usuários, identificar seus direitos, entre outros fatores, de modo que a construção desses termos de maneira apropriada é essencial para o bom funcionamento e organização da rede social e também para a garantia dos direitos dos usuários, evitando abusos. Para isso, vale a observação de recomendações feitas por órgãos relacionados ao tema como a Internet Society, bem como a legislação do país em que a rede social for utilizada.

1.1. Recomendações – Internet Society

No Decálogo de Recomendações sobre o Modelo Brasileiro de Responsabilidade de Intermediários (ISOC, 2021), o Capítulo Brasil do Internet Society trouxe como um dos principais apontamentos a necessidade de os provedores de aplicações de internet garantirem o acesso à informação e medidas de devido processo em seus **termos de uso e serviço**. Assim, esses termos devem prever (i) “informações mais claras sobre regras de moderação de conteúdos, que proporcionem ao usuário o direito de conhecer os motivos da remoção e quais cláusulas foram violadas com o conteúdo de sua autoria”, além de (ii) “mecanismos claros e robustos de recursos para usuários que tenham tido seus conteúdos ou perfis removidos com base em violações aos termos de uso e políticas de moderação”.

1.2. Políticas do Instagram

Os Termos de Uso do Instagram (INSTAGRAM, [2021? a]) (bem como outras políticas da plataforma), com os quais todo e qualquer usuário da plataforma deve concordar para poder utilizar os serviços oferecidos, estabelece algumas diretrizes que definem que tipos de atitudes um usuário pode ou não ter. Entre elas: se passar por outras pessoas; fazer algo ilícito ou fraudulento; violar outras regras da plataforma (Diretrizes da Comunidade, Termos da Plataforma do Facebook e Políticas do Desenvolvedor, Diretrizes de Música, Políticas de Conteúdo de Marca); publicar informações privadas ou confidenciais de outra pessoa sem permissão;

violam direitos de outra pessoa (incluindo direitos de propriedade intelectual), entre outros. Essas outras regras e políticas, como as Diretrizes da Comunidade, estabelecem exigências adicionais, como o cumprimento da lei e a proibição de materiais que contenham nudez (com exceção de casos em que o conteúdo compartilhado tem natureza de protesto, pinturas, esculturas ou diz respeito a situações relacionadas à saúde).

Os Termos definem que, se o Instagram acreditar que o conteúdo publicado na plataforma viola qualquer uma das políticas da rede, ou se o Instagram for autorizado ou obrigado por lei, o material compartilhado poderá ser removido, ou a conta do usuário pode ser encerrada, caso haja violações repetidas ou seja criada uma “exposição legal” para a plataforma.

1.2.1. Procedimentos de remoção

Quanto à forma pela qual o Instagram decide quais conteúdos devem ser removidos, conforme a Central de Ajuda da rede (INSTAGRAM, [2021?b]), é usada uma combinação de tecnologia e de denúncias feitas por usuários para determinar se o conteúdo viola as Diretrizes da Comunidade e, depois, a tecnologia ou uma equipe de análise remove o conteúdo. Pouco se diz sobre a tecnologia utilizada e sobre os critérios que ela leva em conta para tomar suas decisões. Assim, eventuais vícios nas tomadas de decisão de sistemas de inteligência artificial podem ocorrer, gerando resultados parciais ou enviesados (KURTZ, 2021), sem que possam ser verificados e fiscalizados, o que seria negativo tanto no âmbito da moderação de conteúdo quanto em relação aos direitos fundamentais das pessoas, já que essa falta de transparência abre espaço para diversas formas de discriminação.

1.2.2. Apelação de decisões tomadas

Em relação aos procedimentos para apelação de decisões tomadas pela plataforma, a Central de Ajuda da plataforma (INSTAGRAM, [2021?c]) informa que, caso o usuário discorde da decisão tomada, é possível que ele solicite uma análise da decisão. Assim, a decisão é avaliada novamente (não se diz propriamente

como, os critérios utilizados, nem mesmo se essa reavaliação é feita por tecnologia ou por humanos). É informado, ainda, que caso o usuário discorde da decisão reavaliada, ela pode ser submetida a um Comitê de Supervisão através de uma apelação, que deve ser qualificada (ou seja, nem todos os casos podem ser submetidos). Um regulamento interno sobre o Comitê (FACEBOOK, 2022) é disponibilizado para informações sobre os procedimentos específicos, em inglês.

1.3. Relação com legislação

Quanto à sua relação com a legislação, o Instagram inclui em suas Diretrizes da Comunidade a exigência de cumprimento da lei por parte dos usuários, além de apontar em seus Termos de Uso a possibilidade de remoção de conteúdo quando exigido por lei. Isso revela a submissão à jurisdição e conformidade regulatória da plataforma, como apontado pelo Instituto de Referência em Internet e Sociedade (KURTZ, 2021), pela qual existe uma pressão para o cumprimento das normas vinculantes.

Conclusão

Assim, em relação às recomendações feitas pelo Internet Society, se observa que o Instagram não as cumpre de maneira integral e adequada, na medida em que, apesar de apontar conteúdos que não devem ser compartilhados na plataforma, não é transparente quanto às técnicas, tecnologias e critérios utilizados para a tomada de decisão sobre a remoção de um conteúdo.

Além disso, apesar de haver informações sobre a apelação ao Comitê de Supervisão, procedimentos anteriores a essa etapa são pouco claros, não sendo informado como são feitas as reavaliações das decisões, os critérios utilizados para fundamentá-las, e nem mesmo os procedimentos necessários para requerê-las estão disponíveis na Central de Ajuda. Assim, o usuário só terá acesso a esse tipo de informação depois que seu conteúdo já foi removido (as instruções são enviadas junto com a notificação de remoção de conteúdo (INSTAGRAM, [2021?d]), e não antes.

Dessa maneira, observa-se que, no que se refere à transparência em seus Termos de Uso, o Instagram ainda tem muito o que melhorar para que se adeque às recomendações, evite situações discriminatórias, tomadas de decisão parciais ou enviesadas e melhore o entendimento do usuário acerca de suas políticas de moderação de conteúdo. Apesar de ser claro quanto aos tipos de conteúdo que configurariam violações aos Termos ou que seriam considerados inadequados, e, portanto, suscetíveis à remoção e outras penalidades, as regras são pouco transparentes quanto aos procedimentos a serem tomados, os “direitos” dos usuários e o funcionamento interno da plataforma

Referências bibliográficas

DEAN, B. **Instagram Demographic Statistics**: how many people use instagram in 2021?. How Many People Use Instagram in 2021?. 2021. Updated Sep. 10, 2021. Disponível em: <https://backlinko.com/instagram-users#instagram-stats-top-picks>. Acesso em: 24 nov. 2021.

FACEBOOK. **Oversight Board Bylaws**. Janeiro 2022. Disponível em: https://about.fb.com/wp-content/uploads/2020/01/Bylaws_v6.pdf. Acesso em: 13 abr. 2022.

INSTAGRAM. **Termos de Uso**. [2021?a]. Disponível em: <https://www.facebook.com/help/instagram/581066165581870>. Acesso em: 25 nov. 2021.

_____. **Como o Instagram decide o que remover?** Central de Ajuda. [2021?b]. Disponível em: https://help.instagram.com/623395918520872/?helpref=search&query=removed%20post&search_session_id=475678d1287bf1738bd845620d4156ef&sr=2. Acesso em: 25 nov. 2021.

_____. **Acho que o Instagram não deveria ter removido minha publicação**. [2021?c] Central de Ajuda. Disponível em: <https://help.instagram.com/280908123309761>. Acesso em: 25 nov. 2021.

_____. **Um conteúdo que publiquei no Instagram foi removido por causa de uma denúncia de violação de propriedade intelectual**. Quais são os próximos

passos? [2021?d]. Disponível em: https://help.instagram.com/244286600038698/?helpref=search&query=apelação&search_session_id=13f7ff9dcb0359a04e9a02fc69150879&sr=3. Acesso em: 25 nov. 2021.

ISOC BRASIL – SOCIEDADE DA INTERNET NO BRASIL. **Decálogo de Recomendações sobre o Modelo Brasileiro de Responsabilidade de Intermediários**. São Paulo: ISOC Brasil, 2021. Disponível em: <https://isoc.org.br/files/Decálogo de Recomendações sobre o Modelo Brasileiro de Responsabilidade de Intermediários.pdf>

KURTZ, L. P.; DO CARMO, P. R. R.; VIEIRA, V. B. R. **Transparência na moderação de conteúdo: tendências regulatórias nacionais**. Belo Horizonte: Instituto de Referência em Internet e Sociedade, 2021. Disponível em: <https://bit.ly/3xjAUka>. Acesso em: 25 nov. 2021

CURSO GRATUITO

**ESTRUTURA E
FUNCIONAMENTO
DA INTERNET**

ASPECTOS TÉCNICOS,
POLÍTICOS E REGULATÓRIOS

ESTRUTURA E FUNCIONAMENTO DA INTERNET

ASPECTOS TÉCNICOS, POLÍTICOS E REGULATÓRIOS

DESINFORMAÇÃO

Realização



Apoio



O IMPACTO DA “PL DAS FAKE NEWS” NA LIBERDADE DE EXPRESSÃO

Amanda Cristina Alves Silva¹

RESUMO: Considerando as mudanças geradas pela revolução digital, o presente trabalho buscou abordar brevemente a questão da liberdade de expressão no ambiente virtual, analisando o impacto do Projeto de Lei popularmente conhecido como “PL das Fake News” – projeto dedicado ao combate à desinformação – na liberdade de expressão. Para realizar tal análise, abordamos primeiramente o conceito de liberdade de expressão sob a ótica de cultura democrática. Na sequência, abordamos as formas de regulação da liberdade de expressão. E, por fim, analisamos o projeto de lei em questão, e o impacto dos seus dispositivos para a liberdade de expressão no ambiente virtual.

PALAVRAS-CHAVE: Liberdade de Expressão; Regulação da Liberdade de Expressão; “PL das Fake News”.

Introdução

A revolução digital cria um ambiente em que frequentemente se observam conflitos entre interesses individuais e empresariais e, por conta disso, diferentes regulações surgem na tentativa de conciliar tais conflitos. Como exemplo, podemos citar os projetos de lei que buscam combater a desinformação no ambiente virtual, ao mesmo tempo em que se busca resguardar direitos fundamentais, como a liberdade de expressão. No Brasil, o projeto de lei mais avançado sobre o tema trata-se do PL 2630/2020, o chamado PL das *Fake News*.

¹ Advogada. Graduada em Direito pelo Instituto Brasiliense de Direito Público – IDP/EDB.

A partir deste contexto, então, o presente trabalho busca analisar a questão da liberdade de expressão no ambiente virtual, estudando em que medida o “PL das *Fake News*” impacta na liberdade de expressão no ambiente digital?

Para realizar tal análise, abordamos brevemente o conceito de liberdade de expressão sob a ótica de cultura democrática, bem como as formas de regulação da liberdade de expressão. Por fim, analisamos de forma pontual o projeto de lei em questão, e o impacto dos seus dispositivos para a liberdade de expressão no ambiente virtual.

Liberdade de expressão na era da informação digital

O desenvolvimento tecnológico ao longo dos anos vem gerando diferentes impactos na sociedade e nas formas de interação, Jack Balkin nomeia este marco de mudanças de “revolução digital”, que ele considera como sendo a criação e a vasta disponibilização de tecnologias que permitem a cópia, modificação, anotação, colagem, transmissão, e distribuição de conteúdo digital. Para o autor, essas mudanças impactam a organização social e as práticas de liberdade de expressão de quatro formas em especial (BALKIN, 2004).

A primeira delas se relaciona com o custo de copiar e distribuir informações, que caiu drasticamente, de modo que um grande número de pessoas agora tem acesso à possibilidade de compartilhar e distribuir suas opiniões de forma barata e com largo alcance. O segundo diz respeito ao alcance do conteúdo compartilhado, uma vez que agora as informações atravessam fronteiras culturais e geográficas de forma fácil e barata. Em terceiro lugar está a facilidade de inovar a partir de um conteúdo já existente no ambiente virtual. O quarto impacto trata-se da democratização do discurso, uma vez que as tecnologias de distribuição e transmissão são disponibilizadas para um grande número de pessoas, de diferentes segmentos da sociedade, e de diferentes partes do mundo (BALKIN, 2004).

Ao mesmo tempo em que as características da revolução digital empoderam pessoas comuns e democratizam o discurso, também empodera empreendimentos, possibilitando a inovação a partir das novas dinâmicas criadas. Dessa forma, Balkin aponta que a revolução digital é também uma revolução econômica e tecnológica, pois mais tipos de mídia e produtos de informação podem ser vendidos

para mais pessoas, em mais lugares. Consequentemente, a informação passa a ser fonte de riqueza e, da mesma forma, a estrutura necessária para distribuir a informação também se constitui fonte de riqueza (BALKIN, 2004).

Então, a partir da democratização dos conteúdos digitais e com o crescimento da importância desses conteúdos enquanto fonte de riqueza e poder econômico, surgem diferentes conflitos característicos da era digital, e que envolvem diferentes direitos e garantias como, por exemplo, a liberdade de expressão (BALKIN, 2004).

A partir desses conflitos surgem diferentes formas de interpretar a liberdade de expressão no contexto digital. Balkin apresenta, então, a sua perspectiva, a qual ele baseia na ideia de cultura democrática. Para o autor, a liberdade de expressão no ambiente digital não pode ser interpretada apenas sob a ótica da democracia, mas deve considerar toda a expressão cultural, desde os debates mais corriqueiros até os mais complexos (BALKIN, 2004).

Para Balkin, então, a liberdade de expressão possui uma natureza popular que envolve a criatividade, a interatividade, a importância para comunidade e autoformação, de modo que sua interpretação apenas do ponto de vista de deliberações políticas e de assuntos públicos seria extremamente limitante. Desse modo, a leitura da liberdade de expressão a partir da democracia deve ser feita considerando a democracia em seu sentido mais amplo, no sentido de cultura (BALKIN, 2004).

Assim, ao interpretar a liberdade de expressão a partir da ideia de cultura democrática, Balkin aponta que esse princípio apresenta quatro componentes, quais sejam: i) o direito de publicar, distribuir e alcançar uma audiência; ii) o direito de interagir com outros, trocar ideias, influenciar e ser influenciado, transmitir e absorver cultura; iii) o direito de se apropriar de material cultural disponível, modificar e criar algo novo, e compartilhar os resultados com os outros; e iv) o direito de participar e produzir cultura e, portanto, o direito de ter voz no desenvolvimento da cultura e das forças comunicacionais que moldam o ser (BALKIN, 2004).

A proteção prática dessas liberdades, no entanto, envolve diferentes aspectos. Balkin aponta que as evoluções mais importantes no tema de liberdade de expressão não ocorrem a partir de discussões constitucionais, mas sim através de decisões relacionadas a regulação da tecnologia (BALKIN, 2008). Nesse contexto, tem-se que a liberdade de expressão não está relacionada apenas com a ausência de censura, mas também com uma infraestrutura de livre expressão. Nessa

infraestrutura estão incluídos tanto os meios de comunicação e propagação de informação, quanto a possibilidade de criação de novos meios de comunicação, ou seja, a liberdade de inovação (BALKIN, 2008).

Regulando a liberdade de expressão no ambiente digital

Quando abordamos a temática de regulação da liberdade de expressão na era da informação digital, faz-se necessário analisar outras ideias também trabalhadas por Balkin, sobre o que é chamado de “triângulo de Balkin”. O autor leciona que existem dois sistemas que envolvem a regulação da liberdade de expressão. O primeiro sistema trata-se de uma dinâmica dualista, onde tem-se de um lado o Estado que restringe ou limita a manifestação do emissor, que está do outro lado (BALKIN, 2018).

No entanto, em um contexto marcado pela revolução digital, tal sistema não se configura mais suficiente para lidar com as questões relacionadas à liberdade de expressão. Assim, Balkin entende que o sistema que protege a liberdade de expressão hoje é pluralista, o qual envolve a participação não apenas do Estado e do emissor do discurso, mas também das empresas que disponibilizam a infraestrutura onde ocorre essa comunicação digital (BALKIN, 2018).

Para ilustrar tal relação, Balkin propõe um triângulo composto por i) os Estados e entidades supranacionais, ii) os emissores, dentre os quais se inclui a sociedade civil e também a mídia de massa; e iii) as empresas de infraestrutura de internet (BALKIN, 2018). A dinâmica de regulação entre eles ocorre da seguinte forma: os Estados regulam os emissores e os meios de comunicação em massa tradicionais através de uma abordagem “*oldschool*”; os Estados regulam, tentam cooperar e coagir a infraestrutura da internet através de uma abordagem “*newschool*”; e, por fim, a infraestrutura da internet regula os emissores utilizando-se da governança privada (BALKIN, 2018).

A abordagem de regulação da liberdade de expressão que o autor chama de “*oldschool*”, ou tradicional, trata-se da relação em que o Estado se utiliza de sanções e outras formas de punição ou de retribuição para controlar o discurso de pessoas,

associações ou empresas de comunicação. Já a abordagem “*newschool*”, por outro lado, trata-se da regulação direcionada às empresas responsáveis pela infraestrutura da internet (BALKIN, 2018).

A nova abordagem de regulação da liberdade de expressão, todavia, apresenta problemas. O principal deles a ser abordado aqui, trata-se da “censura colateral”. Tal problema ocorre quando o Estado faz com que A controle o discurso de B, sob ameaça de punição, dessa forma, o ente privado passa a regular o discurso online em nome do Estado (BALKIN, 2018).

Neste cenário, o desdobramento mais provável é que o ente privado opte pela precaução e, portanto, bloqueie ou filtre mais conteúdos na tentativa de evitar sanções, bem como retire conteúdos que possam ser questionados de alguma forma, com a mesma finalidade. Isto porque, o responsável pela infraestrutura tem o interesse de que a maioria de seus usuários se sinta confortável, de modo que restringir o discurso de poucos não vai ser prejudicial para o seu modelo de negócio, diferentemente de repetidas responsabilizações governamentais pelo discurso de terceiros, o que pode ser um risco à sua atividade (BALKIN, 2018).

Nesse contexto, o autor destaca a necessidade de que as empresas de infraestrutura sejam enxergadas não apenas como facilitadoras da comunicação digital, mas sim como governantes dos espaços sociais, uma vez que possuem cada vez mais capacidade de vigilância e controle, constituindo-se, portanto, como governança privada (BALKIN, 2018).

PL das fake news e a liberdade de expressão

Assim como outros países, nos últimos anos o Brasil passou a visualizar o crescimento da desinformação no ambiente virtual, em especial no que se refere aos

temas relacionados ao período eleitoral e possíveis candidatos.² As consequências desse crescimento têm sido vistas na sociedade, de modo que parece inegável a necessidade de legislar a respeito.

É sob essa motivação que é editado o Projeto de Lei nº 2630/2020, o qual se intitula como Lei Brasileira de Liberdade, Responsabilidade e Transparência na Internet, e pretende estabelecer as “normas, diretrizes e mecanismos de transparência para provedores de redes sociais e de serviços de mensageria privada a fim de garantir a segurança e ampla liberdade de expressão, comunicação e manifestação do pensamento”³.

A proposta elenca em seu art. 4º quatro objetivos: I) o fortalecimento do processo democrático por meio do combate ao comportamento inautêntico e às redes de distribuição artificial de conteúdo e do fomento ao acesso à diversidade de informações na internet no Brasil; II) a defesa da liberdade de expressão e o impedimento da censura no ambiente online; III) a busca por maior transparência das práticas de moderação de conteúdos postados por terceiros em redes sociais, com a garantia do contraditório e da ampla defesa; e IV) a adoção de mecanismos e ferramentas de informação sobre conteúdos impulsionados e publicitários disponibilizados para o usuário.

O projeto de lei em questão é alvo de diversas críticas, que se dirigem à diferentes aspectos da proposta, tanto do que está previsto quanto do que não foi abordado.⁴ Um dos pontos criticados diz respeito a imposição de que os provedores de redes sociais e de serviços de mensageria adotem medidas para vedar o funcionamento de contas inautênticas (art. 6º, I).

De igual modo, o art. 7º ao prever que “os provedores de redes sociais e de serviços de mensageria privada poderão requerer dos usuários e responsáveis pelas

² Ver mais em “**A desinformação é a protagonista da crise institucional brasileira**”. Disponível em: <https://www.agazeta.com.br/editorial/a-desinformacao-e-a-protagonista-da-crise-institucional-brasileira-0921>. Acesso em: novembro, 2021.

³ Projeto de Lei 2630/2020. Disponível em: https://www.camara.leg.br/proposicoesWeb/prop_mostrarintegra;jsessionid=node05gpyflzixn1k1pan2711izv0r16960225.node0?codteor=1909983&file-name=PL+2630/2020. Acesso em: novembro, 2021.

⁴ Algumas das críticas podem ser vistas em “**PL das Fake News: Recomendações**”. Disponível em: <https://www.jota.info/opiniao-e-analise/artigos/pl-das-fake-news-recomendacoes-04112021>. Acesso em: novembro, 2021.

contas, em caso de denúncias por desrespeito a esta Lei, no caso de indícios de contas automatizadas não identificadas como tal, de indícios de contas inautênticas ou ainda nos casos de ordem judicial, que confirmem sua identificação, inclusive por meio da apresentação de documento de identidade válido” gerou questionamentos a respeito da razoabilidade e necessidade da medida.⁵

Apesar dos pontos questionáveis, o projeto também estabelece previsões que buscam dar maior transparência na aplicação dos termos de uso das plataformas. Como exemplo tem-se o art. 12, que dispõe que “os provedores de aplicação de internet submetidos a esta Lei devem garantir o direito de acesso à informação e à liberdade de expressão de seus usuários nos processos de elaboração e aplicação de seus termos de uso, disponibilizando mecanismos de recurso e devido processo”.

Percebe-se, de forma clara, que o projeto utiliza uma abordagem que Balkin chamou de “*newschool*”, uma vez que estabelece diversas medidas a serem adotadas pelos provedores de redes sociais e pelos provedores de serviços de mensageria privada, a fim de regular o discurso de particulares. Como consequência, portanto, é possível vislumbrar a eventual concretização da “censura colateral”, também abordada por Balkin.

Analisando o projeto sob a luz dos conceitos apresentados nos tópicos anteriores, em busca de verificar em que medida o PL das Fake News restringe a liberdade de expressão no ambiente digital, é possível fazer alguns apontamentos.

Percebe-se, de forma clara, que o projeto utiliza uma abordagem que Balkin chamou de “*newschool*”, uma vez que estabelece diversas medidas a serem adotadas pelos provedores de redes sociais e pelos provedores de serviços de mensageria privada, a fim de regular o discurso de particulares. Como consequência, portanto, é possível vislumbrar a eventual concretização da “censura colateral”, também abordada por Balkin.

⁵ Ver mais em “**Comentários ao PL nº 2630/2020**”. Disponível em: https://itsrio.org/wp-content/uploads/2021/03/Adalthon-de-Paula-Souza_COMENTARIOS-AO-PL-N-2.630_2020-O-PL-DAS-FAKE-NEWS-%E2%80%93CONTAS-INAUTENTICAS-IDENTIFICACAO-DE-USUARIOS-E-RASTREABILIDADE-DE-MENSAGENS.pdf. Acesso em: novembro, 2021.

Note-se, como exemplo, que a exigência de que os provedores verifiquem a identidade dos usuários em determinados casos, pode acarretar no estabelecimento – por parte dos provedores – de critérios de verificação da identidade para todos os usuários, o que, a princípio, aparenta ser uma medida excessiva e desnecessária.

Tem-se que em diferentes dispositivos o projeto de lei ressalta a necessidade de proteção à liberdade de expressão, à livre manifestação do pensamento, à liberdade cultural e outros aspectos que se englobam na visão da democracia cultural de Balkin. Assim, a norma não limita de forma expressa a liberdade de expressão interpretada sob este contexto.

Por outro lado, conforme já exposto acima, as imposições colocadas sob os provedores de aplicações e de serviços de mensageria apresentam potencial de gerar a censura colateral, uma vez que estabelecem ônus relevantes aos entes privados e, ainda, em razão do texto ter incompletudes de conceituação, deixando à cargo dos entes privados interpretar os significados da norma.

Cabe destacar, por fim, que não há garantias de que o objetivo precípua do projeto de lei em questão será alcançado, ou seja, que a desinformação seja embarreada. Em razão da configuração e da dinâmica do ambiente virtual, o rápido desenvolvimento e crescimento das redes sociais, em conjunto com a dificuldade de se colocar freios em tais mecanismos, o projeto – se aprovado – pode não ser eficaz. Ao mesmo tempo, pode vir a estimular de forma indireta uma limitação às liberdades individuais na internet. Além disso, o problema da desinformação não se restringe ao Brasil, de modo que observar a repercussão de regulações em outros países se faz necessário, a fim de que se perceba a eficácia ou não de tais normas.

Conclusão

Ao tentar verificar em que medida o projeto de lei estudado impacta a liberdade de expressão na internet, percebemos que seus dispositivos não representam um risco direto à liberdade de expressão conforme definida ao longo do texto. No entanto, é possível notar que diferentes dispositivos abrem espaço para a possibilidade de ocorrer a censura colateral. Isso porque, o normativo como um todo

impõe diversas obrigações para os provedores, e, ao mesmo tempo, deixa em aberto diferentes conceitos, o que pode dar às empresas uma “zona cinzenta” o suficiente para adotar regulações privadas que melhor prestigiem os seus interesses, e não o interesse da sociedade. Todavia, tal projeto ainda está pendente de aprovação e pode sofrer diferentes alterações em seu texto, de modo que sane as fraquezas em sua redação ou, inclusive, amplie as problemáticas já encontradas.

Referências bibliográficas

BALKIN, J. M. Digital speech and democratic culture: A theory of freedom of expression for the information society. **New York University Law Review**, v. 79, 2004. Disponível em: http://digitalcommons.law.yale.edu/fss_papers/240/. Acesso em: nov. 2021.

_____. Free Speech is a Triangle. **SSRN**, 2018. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3186205. Acesso em: nov. 2021.

_____. The future of free expression in a digital age. **Pepperdine Law Rev.**, v. 36, 2008. Disponível em: http://digitalcommons.law.yale.edu/fss_papers/223/. Acesso em: nov.2021.

MACHADO, C.; DURIGAN, V. PL das Fake News: Recomendações. **JOTA**, 04 nov. 2021. Disponível em: <https://www.jota.info/opiniao-e-analise/artigos/pl-das-fake-news-recomendacoes-04112021>. Acesso em: nov. 2021.

Opinião Gazeta. **A desinformação é a protagonista da crise institucional brasileira**. Publicado em 16 de setembro de 2021. Disponível em: <https://www.gazeta.com.br/editorial/a-desinformacao-e-a-protagonista-da-crise-institucional-brasileira-0921>. Acesso em: nov. 2021.

Projeto de Lei 2630/2020. Congresso Nacional. Disponível em: https://www.camara.leg.br/proposicoesWeb/prop_mostrarintegra;jsessionid=node05gpyfl-zixn1k1pan2711izv0r16960225.node0?codteor=1909983&file-name=PL+2630/2020. Acesso em: nov. 2021.

SOUZA, A. de P. **Comentários ao PL nº 2630/2020**. ITS Rio, mar. 2021. Disponível em: https://itsrio.org/wp-content/uploads/2021/03/Adalthon-de-Paula-Souza_COMENTARIOS-AO-PL-N-2.630_2020-O-PL-DAS-FAKE-NEWS-%E2%80%93CONTAS-INAUTENTICAS-IDENTIFICACAO-DE-USUARIOS-E-RASTREABILIDADE-DE-MENSAGENS.pdf. Acesso em: nov. 2021.

FORMAS DE COMBATE À DESINFORMAÇÃO SOBRE A PANDEMIA DE COVID-19 ADOTADAS PELO TWITTER

Jéssica Guedes Santos¹

RESUMO: A desinformação é um grande problema atinente à relação entre sociedade, tecnologia e direito. A sociedade se comunica massivamente pelas redes sociais, o que faz as plataformas serem um locus atual de disseminação de campanhas desinformativas, conduta que não foi diferente durante a pandemia de covid-19, inclusive, organizações internacionais de saúde destacaram a importância de combater a desinformação para proteger a saúde pública. Assim, o presente artigo investigou como o Twitter - rede importante no debate público - tratou a desinformação sobre a pandemia na sua plataforma analisando os relatórios de transparência, a política de moderação de conteúdo e os comunicados sobre o tema. Com base no estudo, foi possível obter três conclusões principais, quais sejam, i) a plataforma deve ter a responsabilidade de apresentar dados atualizados para a sociedade, ii) a plataforma deve seguir a sua própria política de moderação e iii) o uso da IA foi intensificado e aprimorado durante a pandemia.

PALAVRAS-CHAVE: Desinformação; Twitter; IA; Covid-19.

Introdução

A desinformação consiste em “todas as formas de informações falsas, imprecisas ou enganadoras criadas, apresentadas e promovidas para causar prejuízo de maneira proposital ou para fins lucrativos” (RAIS, 2020). Assim, a desinformação é construída estrategicamente com objetivo de falsear a realidade e gerar algum

¹ Mestranda em Direito pela Universidade de Brasília (UnB). Advogada em privacidade e proteção de dados. Cofundadora do Portal “Bot Jurídico”.

prejuízo social. O tema passou a participar mais frequentemente do debate público por meio das eleições de Donald Trump para a Presidência dos Estados Unidos em 2016.

O uso da internet e das plataformas de redes sociais são fatores fundamentais para pensar no atual cenário da desinformação diante da importância que detém para a comunicação. Neste sentido, dados da pesquisa TIC Domicílios 2020, organizada pelo CGI.br divulgada em agosto de 2021, informam que o Brasil tem 152 (cento e cinquenta e dois) milhões de usuários da internet e que 72% (setenta e dois por cento) desses usuários acessa as redes sociais. Esses números demonstram o amplo alcance que as redes sociais possuem na sociedade.

Desde o aumento no uso das redes sociais e da crescente onda de desinformação, existe uma discussão patente sobre a moderação de conteúdo, tanto de forma direta pelas plataformas quanto de forma indireta pelo Judiciário, e recebe novos contornos por conta da pandemia de covid-19.

Neste sentido, o presente artigo parte da pergunta de pesquisa: “Como o Twitter enfrentou a desinformação em conteúdos que tratassem sobre a pandemia de covid-19?” com o objetivo geral de identificar o uso de mecanismos de IA pela plataforma para combater a desinformação disseminada na plataforma durante a pandemia.

Desinformação e plataformas de redes sociais

As redes sociais são um importante *locus* para a disseminação de desinformação diante da criação de bolhas que direcionam o conteúdo específico para grupos de interesse que já estão polarizados em certos assuntos (PARISER, 2011).

As redes sociais permitem que o conteúdo desinformativo seja disseminado rapidamente tanto na rede originária na qual é postada como em outras redes sociais, sempre direcionando o formato da mensagem para as bolhas de comunicação. Sendo que, o engajamento da desinformação pode ser feito organicamente ou artificialmente por meio de *bots* com o intuito de criar, por vias artificiais, aparência de espontaneidade por parte de movimentos sociais de base (DOURADO, 2021, p. 98).”

A ligação entre desinformação, redes sociais e efeitos políticos somente foi mais amplamente discutida após a exposição do chamado “escândalo da Cambridge Analytica”, empresa britânica de mineração e análise de dados que utilizou dados de aplicações da rede social Facebook para direcionar notícias falsas que possivelmente influenciaram em diversos processos eleitorais, entre eles, nas eleições americanas de 2016 (AMER, 2019).

Após a comprovação de que os dados e preferências pessoais estavam sendo utilizados sem qualquer ciência dos usuários, foi iniciado um amplo debate sobre a necessidade das redes sociais protegerem os dados dos usuários e desenvolverem formas de enfrentar a desinformação, tanto no aspecto estatal relacionado com a regulação do tema, como no aspecto interno com alterações na moderação de conteúdo.

É claro que o tema apresenta desafios, uma vez que a desinformação é multifacetada e vai ser espalhada em diversos tipos de conteúdos e plataformas buscando atingir o maior número de destinatários possíveis, uma vez que a repetição do conteúdo desinformativo é peça chave para permitir a maior amplitude e engajamento social desse tipo de conteúdo (BACHUR, 2021).

Por si só, essa estrutura gera dificuldades para pensar em uma estrutura única de combate à desinformação, tanto por conta das peculiaridades de cada plataforma, quanto pela autonomia no funcionamento. Esse cenário permite inferir que se houver posturas que restrinjam e/ou limitem a desinformação em alguma plataforma, ocorra a migração dos agentes e da mensagem desinformativa para outra, como, por exemplo, no Brasil, após a adoção de algumas medidas específicas pelo WhatsApp, foi constatada a migração de agentes desinformativos para o Telegram (SPAGNUOLO, 2021).

Assim, buscando estabelecer medidas centrais sobre o tema, o debate acerca da moderação de conteúdo é necessário para apresentar diretrizes que possam ser utilizadas pelas plataformas. Atualmente, existem termos de uso específicos que estabelecem as hipóteses pelas quais as plataformas podem excluir conteúdos desinformativos. Para além da exclusão, por exemplo, o Twitter e o Facebook também moderam conteúdo de outras formas, como a limitação do alcance e a sinalização das postagens.

Mas toda essa discussão tem um conteúdo interno muito forte, já que as plataformas definem as suas políticas e a executam da forma que acharem mais conveniente para a empresa. A ausência de transparência e equidade na moderação de conteúdo (RODRIGUES, KURTZ, 2020) levou a participação da sociedade civil nesse debate, tanto com construções acadêmicas sobre o assunto, quanto com iniciativas práticas.

O próprios usuários e organizações começaram a estudar e estruturar ações de combate à desinformação nas redes. Uma dessas ações é a criação e fortalecimento das agências de checagem de fatos que permitem a averiguação da veracidade de discursos públicos (SANTANA, SIMEÃO, 2019). Mas, existem ocasiões que as próprias plataformas apresentam algum tipo de dificuldade para a continuidade dessa medida, como no caso da Fátima, bot criado pela agência de checagem Aos Fatos, que tinha funções conversar com o usuário, enviar notícias chegadas e ensinar o usuário a checar notícias, mas foi banido pelo Twitter em agosto de 2021, já que a plataforma impediu o acesso da Fátima a sua application programming interface (API).

Medidas adotadas pelo twitter

Em março de 2020, o Twitter anunciou a sua estratégia de combate à desinformação sobre a pandemia informando que “continuaremos a remover conteúdo comprovadamente falso ou potencialmente enganoso que tem maior risco de causar danos às pessoas (GADDE, DERELLA, 2020)”. Interessante destacar que nesse comunicado o Twitter anunciou que aumentaria o uso de *machine learning* e automação quanto às medidas de moderação de conteúdo relacionada com a pandemia.

Em maio de 2020, o Twitter fez um aditivo ao comunicado informando que seriam adicionados avisos e alertas nas postagens sobre o covid-19 fazendo a análise sobre três categorias, quais sejam, i) informações enganosas - “declarações ou afirmação que tenham sido confirmadas como falsas ou enganosas por especialistas no tema em questão, como autoridades em saúde pública” com aviso e remoção, ii) informações enganosas - “declarações ou afirmações cuja precisão, veracidade ou credibilidade é genuinamente contestável ou desconhecida” com aviso e alerta e iii) afirmações não confirmadas - “informações (que podem ser

verdadeiras ou falsas) que ainda não foram verificadas e confirmadas no momento que são compartilhadas (ROTH, PICKLES, 2020).”

O Twitter também editou a “Política de informações enganosas sobre a covid-19” proibindo que a plataforma fosse usada para disseminar informações falsas ou enganosas sobre a covid-19. Para fundamentar a política, a empresa apontou que foi preciso adotar tal postura diante do surgimento de teorias da conspiração e disseminação de boatos não confirmados que poderiam colocar as pessoas em risco.

Em dezembro de 2020, o Twitter divulgou novo comunicado sobre o assunto informando abordagem com as informações enganosas sobre a vacina destacando que priorizaria a remoção e a adição de rótulos em conteúdo que alegassem que as vacinas eram usadas para causar dano ou controlar populações.

Assim, em um quadro geral, o Twitter fez uma gradação com base na gravidade, no tipo de violação e no histórico de violações anteriores por meio da contagem de transgressão e, assim, estabeleceu três punições nos termos indicados nas tabelas abaixo:

Transgressão	Consequência
Informações enganosas sobre a natureza ou tratamento da covid ou sobre a pandemia ou vacinas	Remoção do tweet após duas transgressões e impedimento temporário em acessar a conta
Conteúdo que viola a política, mas que o Twitter prefere não remover	Aplicação do aviso, redução da visibilidade do tweet, desativação de curtidas/respostas/retweets e fornecimento de link para mais explicações. Se os tweets marcados forem considerados como nocivos conta como uma transgressão
Bloqueio da conta e suspensão permanente	Identificação de conta que é dedicada a espalhar a desinformação sobre o covid-19

Tabela 1 – Elaboração própria

Punição	Números de transgressões
Não há	1
Bloqueio da conta por 12 horas	2
Bloqueio da conta por 12 horas	3
Bloqueio da conta por 7 dias	4
Bloqueio ou suspensão permanente	5 ou mais

Tabela 2 – Elaboração própria

Em janeiro de 2021, o Twitter divulgou uma página na qual concentra diversas informações acerca das iniciativas da plataforma sobre o combate ao covid-19, sendo que, até 12.01.2021, a plataforma teria removido 8493 (oito mil, quatrocentos e noventa e três) tweets e contestado 11,5 (onze milhões e quinhentos mil contas).

Mas, foi possível observar que a adoção formal dessas regras não foi capaz de gerar envolvimento direto equânime da plataforma. No Brasil, foram observados casos nos quais indivíduos e conteúdos cometiam seguidas transgressões sobre as políticas de moderação da covid-19 e não tinham nenhuma punição aplicada pela plataforma (ELY *et al*, 2021).

Neste sentido, considerando o debate global do tema e a inobservância dos critérios adotados pelo Twitter em outras partes do mundo no Brasil, o Ministério Público Federal (MPF), em sede de inquérito civil público que investiga as possíveis violações de direitos fundamentais causadas pela atuação nos provedores sobre a desinformação e violência no mundo digital, em 06 janeiro de 2022, requisitou que a plataforma prestasse informações detalhadas sobre (i) a plataforma de denúncia de conteúdos desinformativos sobre a covid-19 e os motivos pelos quais outros países possuíam essa plataforma e o Brasil não e (ii) os critérios adotados pela plataforma para conferir a verificação de usuários.

O questionamento do MPF foi impulsionado em grande medida diante da reação da sociedade civil da usuária Bárbara Destefani, que era popular na plataforma por espalhar desinformação sobre a covid-19 e instituições político-democráticas.

Diante da repercussão do caso, o Twitter divulgou nota em 17 de janeiro de 2022 informando a adoção do mecanismo de denúncia de informações potencialmente enganosas para o Brasil, Espanha e Filipinas. Antes, essa ferramenta somente funcionava nos Estados Unidos, Austrália e Coreia do Sul. Igualmente, em 25 de janeiro de 2022, o Twitter divulgou uma atualização da central de transparência atualizando pedidos de informação e remoção e, conseqüentemente, as aplicações das regras do Twitter, mas sem fazer um recorte específico para desinformação.

Conclusão

A propagação de desinformação passou a ser estratégia de sucesso para prejudicar indivíduo, causa ou organização, ultrapassando o âmbito eleitoral e encontrando lastro também na pandemia de covid-19.

Organizações internacionais de saúde destacaram às campanhas de desinformação que foram realizadas contra a conscientização sobre a gravidade do coronavírus, assim como contra as medidas de proteção da população, entre elas, a vacinação.

Os instrumentos estudados no presente artigo permitem dizer que o Twitter adotou medidas institucionais para impedir que a desinformação sobre a pandemia se espalhasse na sua plataforma criando a política específica de moderação de conteúdo nesses casos e informando medidas relacionadas com o aprimoramento da IA para remoção e marcação de conteúdos desinformativos.

Contudo, ainda é possível identificar dois problemas sérios que a plataforma tem que enfrentar sobre o tema. O primeiro é a ausência de dados atualizados sobre a conduta do Twitter. O segundo é a possível ausência de cumprimento efetivo das próprias políticas da plataforma.

Os dois problemas supracitados culminam na percepção que a plataforma não adota todas as medidas necessárias para combater a desinformação e que adota medidas mais específicas sobre o assunto quando provocado por contextos político-sociais ou por autoridades públicas, o que vai de encontro à imagem que a plataforma quer passar para os seus usuários e para a opinião pública.

Referências bibliográficas

AOS FATOS. **Após suspensão, Fátima, robô checadora do Aos Fatos, é bloqueada pelo Twitter**. Publicado em: 03 ago. 2021. Disponível em: <https://www.aosfatos.org/noticias/apos-suspensao-fatima-robo-checadora-do-aos-fatos-e-bloqueada-pelo-twitter/>. Acesso em: fevereiro de 2022.

BACHUR, J. P. Desinformação política, mídias digitais e democracia: Como e por que as fake news funcionam?. **Direito Público**, [S. l.], v. 18, n. 99, 2021. Disponível em: <https://www.portaldeperiodicos.idp.edu.br/direitopublico/article/view/5939> Acesso em: fevereiro de 2022.

ELY, D. *et al.* **Posts antivacina contrariam regra do Twitter alcançam mais de 147 mil interações**. Jun/2021 Disponível em: <https://www.aosfatos.org/noticias/posts-antivacina-contrariam-regra-do-twitter-e-alcancam-mais-de-147-mil-interacoes/>. Acesso em: fevereiro de 2022.

COMITÊ GESTOR DA INTERNET NO BRASIL (CGI.br). **Pesquisa sobre o uso das tecnologias da informação e comunicação no Brasil**: pesquisa TIC Domicílios, ano 2020: Apresentação dos principais resultados. São Paulo: CGI.br, 2021. Disponível em: https://cetic.br/media/analises/tic_domicilios_2020_coletiva_imprensa.pdf. Acesso em: dezembro de 2021.

DOURADO, T. **Fake News**: quando mentiras viram fatos políticos. Porto Alegre: Zouk, 2021. Op.cit. p. 98.

GADDE, V.; DERELLA, M. **Uma atualização sobre nossa estratégia contínua durante a covid-19**. Jul/2020. Disponível em: https://blog.twitter.com/pt_br/topics/company/2019/uma-atualizacao-sobre-nossa-estrategia-continua-durante-o-covid-19. Acesso em: dezembro de 2021.

PARISER, E. Cuidado com as “bolhas de filtro” online. **TED Talk**, TED 2011, março de 2011. Disponível em: https://www.ted.com/talks/eli_pariser_beware_online_filter_bubbles?language=pt#t-261802. Acesso em: dezembro de 2021.

PRIVACIDADE HACKEADA. Direção: Karim Amer, Jehane Noujaim. Netflix, 2019.

ROTH, Y.; PICKLES, N. **Atualizando nossa abordagem a informações enganosas**. Mai/2020. Disponível em: https://blog.twitter.com/pt_br/topics/company/2019/atualizando-nossa-abordagem-a-informacoes-enganosas. Acesso em: dezembro de 2021.

RAIS, D. Capítulo 11. Desinformação no contexto democrático. In: Abboud, G.; JR, N.; CAMPOS, R. **Fake News e Regulação** – Ed. 2020. São Paulo: Editora Revista dos Tribunais. Disponível em: <https://thomsonreuters.jusbrasil.com.br/doutrina/1197132515/fake-news-e-regulacao-ed-2020>. Acesso em: fevereiro de 2022.

RODRIGUES, G.; KURTZ, L. **Transparência sobre moderação de conteúdo em políticas de comunidade**. Belo Horizonte: Instituto de Referência em Internet e Sociedade, 2020. Disponível em: <https://irisbh.com.br/wp-content/uploads/2020/10/Transpar%C3%Aancia-sobre-moderac%C3%A7%C3%A3o-de-conte%C3%BAdo-em-pol%C3%ADticas-de-comunidade-IRIS.pdf>. Acesso em: fevereiro de 2022.

SPAGNUOLO, S. **Telegram, o novo refúgio da extrema direita no Brasil**. Fev/2021. Disponível em: <https://brasil.elpais.com/brasil/2021-02-21/telegram-o-novo-refugio-da-extrema-direita.html>. Acesso em: fevereiro de 2022.

SANTANA, G. P.; SIMEÃO, E. L. M. S. Notícias falsas: origens, meios de disseminação, contextos e enfrentamento. In: **Seminário Hispano-Brasileño de Investigación en Información, Documentación y sociedad**, 8, 2019, São Paulo.

TWITTER. **Coronavírus: fique seguro e informado no twitter**. 2021. Disponível em: https://blog.twitter.com/en_us/topics/company/2020/covid-19#metrics. Acesso em: dezembro de 2021.

_____. **Política de informações enganosas sobre a covid-19**. 2021. Disponível em: <https://help.twitter.com/pt/rules-and-policies/medical-misinformation-policy>. Acesso em: dezembro de 2021.

TWITTER SAFETY. **Covid-19: nossa abordagem para informações enganosas sobre vacinas**. 2020. Disponível em: https://blog.twitter.com/en_us/topics/company/2020/covid19-vaccine. Acesso em: dezembro de 2021.

ESTRUTURA E FUNCIONAMENTO DA INTERNET

ASPECTOS TÉCNICOS, POLÍTICOS E REGULATÓRIOS

ALGORITMOS E USO ÉTICO DE INTELIGÊNCIA ARTIFICIAL

UTILIZAÇÃO DE INTELIGÊNCIA ARTIFICIAL NA TOMADA DE DECISÕES EM DIFERENTES CENÁRIOS: ASPECTOS POSITIVOS E NEGATIVOS, O PAPEL DA ATUAÇÃO HUMANA NESSE CONTEXTO E O GRAU DE TRANSPARÊNCIA NAS DECISÕES

Beatriz Stefani Castro¹

RESUMO: Sabe-se que, atualmente, a sociedade se encaminha cada vez mais para um cenário em que a Inteligência Artificial e as inovações tecnológicas ganham um espaço relevante na vida das pessoas nos mais diversos âmbitos. Tendo isso em vista, o artigo procura estudar e analisar algumas formas de atuação da Inteligência Artificial na esfera das tomadas de decisão em alguns possíveis contextos. Com isso, buscou-se mostrar de que maneira isso é realizado na prática - a partir de três principais grandes exemplos: no contexto da saúde, em processos de recrutamento de empresas e no judiciário; os principais aspectos positivos; os principais pontos de atenção e quais seriam as ferramentas para que seja um processo benéfico às partes envolvidas. Sendo assim, para a análise das três situações mencionadas, ponderou-se tanto os aspectos incentivadores para a utilização de Inteligência Artificial, quanto os prováveis obstáculos que poderão ser enfrentados. Em conclusão, constatou-se a importância da transparência e accountability quando se trata da disposição dos algoritmos, a fim de que aqueles que não possuem conhecimento sobre a Inteligência Artificial estejam o mais cientes possível de como a máquina chegou a determinada conclusão. Além disso, explora-se a importância da atuação humana em conjunto com a tecnologia, com o objetivo de garantir que a resolução seja bastante precisa.

PALAVRAS-CHAVE: Inteligência Artificial; Transparência; Tomada de decisão; Atuação Humana

¹ Beatriz Stefani Castro, estudante da Graduação em Direito pela Escola de Direito de São Paulo da Fundação Getúlio Vargas.

No decorrer das últimas décadas, faz-se cada vez mais evidente o espaço que o avanço tecnológico ocupa no cotidiano, de maneira a estar inserido desde em coisas mais simples, como algum novo recurso nos *smartphones*, até em cenários como facilitação de diagnóstico médico. Se antigamente era difícil visualizar um futuro em que as máquinas ou robôs seriam responsáveis por auxiliar em tomadas de decisão, por exemplo, sabe-se que esse cenário se torna cada vez mais comum e próximo. Certamente, atrelado às modificações recorrentes na contemporaneidade, urge que o correto e acessível manuseio dessas tecnologias também seja prioridade em sua aplicação nas mais diversas situações enfrentadas.

No processo em que a Inteligência Artificial é utilizada em determinadas tecnologias, o que se tem é um processamento intenso de informações, dados pessoais e atuação de algoritmos justamente a fim de otimizar o serviço a ser oferecido. Isso porque a máquina precisa possuir alto grau de confiabilidade e precisão em processos de tomada de decisão, principalmente em casos nos quais a atuação humana em conjunto é minimizada. Contudo, por vezes não é explicado pelos desenvolvedores como funciona a tomada de decisão baseada em algoritmos, o que dificulta a devida *accountability* nesse processo. Ora, principalmente pelas novas tecnologias terem tamanha e ascendente influência na vida das pessoas é que se é preciso entender o procedimento por trás de algo que não possui fácil tradução.

Nesse sentido, a Recomendação do Parlamento Europeu nº 12, de fevereiro de 2017² dispõe que se faz necessário esclarecimento sobre o processo por trás da tomada de decisão que envolve Inteligência Artificial. Além disso, mais recentemente a Portaria GM nº 4.617 de 2021 aborda a aplicação do referido recurso tecnológico mediante à Estratégia Brasileira de Inteligência Artificial (EBIA), estabelecendo diversas diretrizes e objetivos para tal. Resta evidente, por conseguinte, a relevância da ética e maior transparência possível, a fim de que as tecnologias não acabem fomentando decisões estabelecidas a partir de algum viés sem explicação lógica para tanto. Isto considerado, vale reiterar os diversos contextos em que a tomada de decisão com interferência de Inteligência Artificial é

² Disponível em: https://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_PT.html

inserida atualmente, a exemplo da área da saúde, processos seletivos de empresas e na atuação do judiciário.

Atualmente, a utilização de tecnologia avançada no contexto da saúde cresce constantemente, com a finalidade de não somente facilitar procedimentos, mas também de aumentar a acurácia e precisão da atuação médica. Isso porque, uma vez combinadas a *expertise* do médico com a Inteligência Artificial, estima-se que os resultados possuam bastante acurácia. Voltando-se para alguns exemplos de aplicabilidade da referida tecnologia para otimização de procedimentos da área da saúde, tem-se a utilização de Inteligência Artificial na relação médico-paciente, de maneira que a máquina realiza uma avaliação do atendimento médico por meio de um estudo de elementos como fala e resposta do profissional, a fim de que a consulta seja sempre aprimorada. Ademais, há a utilização de Inteligência Artificial para potencializar de maneira benéfica o diagnóstico médico e leitura de exames.

Percebe-se que os referidos sistemas, e outras várias exemplificações de como as tecnologias vêm auxiliando nesse setor, necessitam da observância de alguns elementos para seu efetivo funcionamento da maneira mais positiva possível tanto para o médico, quanto para o paciente. Primeiramente, é imprescindível que os algoritmos envolvidos no processo contem com um grau de confiabilidade, no sentido de operar com poucas falhas, e com um sistema que seja capaz de constatar esses possíveis erros por parte da máquina. Para além disso, como já mencionado anteriormente, outro princípio fundamental que deve ser orientador dessa relação entre a tecnologia e a saúde é a transparência, no sentido de que os algoritmos possam ser explicados e auditáveis quando necessário.

Passando para a análise da tomada de decisão com interferência da máquina em um contexto diverso, os *softwares* desenvolvidos para processos seletivos de empresas, por exemplo, também apresentam pontos controversos. Certamente, não se pode negar os benefícios desse método para a empresa contratante, uma vez que se faz possível um acompanhamento mais detalhado de todas as etapas, bem como alta acurácia de análise do comportamento dos candidatos durante o recrutamento. Entretanto, um ponto a se prestar atenção é se há algum viés presente na tomada de decisão da máquina, como já observado em alguns casos anteriores.

Fica claro que a referida tecnologia lida com diversos dados dos candidatos, inclusive com dados pessoais sensíveis³ em alguns casos, o que apresenta, por si só, um potencial discriminatório de certa forma. Ao passo que a Inteligência Artificial consegue, nesse contexto, criar processos de recrutamento mais personalizados de acordo com o que a empresa preza em seus funcionários, também pode atuar como um fomento à discriminação nesse mesmo cenário. Diante de uma lacuna jurídica no que tange à regulação específica para a Inteligência Artificial, alguns outros mecanismos de regulação podem ser acionados, haja vista a governança. Reitera-se, portanto, a importância ainda da presença humana por trás do procedimento, a fim de operar como supervisora das decisões que são tomadas pela máquina. Ressalta-se, aqui, que anteriormente a LGPD previa a obrigatoriedade de supervisão humana para as decisões tomadas com base na máquina, contudo, a Mensagem nº 288, de 8 de julho de 2019, vetou essa disposição. Ainda assim, reitera-se a importância ainda da presença humana por trás do procedimento, a fim de operar como inspetora dessas decisões.

Ademais, a governança de algoritmos também pode exercer um papel fundamental na prevenção contra práticas discriminatórias no âmbito da utilização de Inteligência Artificial. De acordo com os professores Danilo Doneda e Virgílio Almeida, a partir do estudo da natureza do algoritmo, de análises de risco ou até mesmo agindo em cima dos dados que o algoritmo utiliza para seu funcionamento, a governança poderá ser realizada de maneira a buscar a minimização dos riscos trazidos pela tecnologia e incrementar sua eficiência. (DONEDA, ALMEIDA, 2016)

Um outro cenário possível em que a Inteligência Artificial pode desempenhar auxílio na tomada de decisão é no âmbito jurídico. Sabe-se que há grande quantidade de processos judiciais em trâmite atualmente e o incontestável congestionamento da justiça nesse sentido. Dessa forma, pode-se pensar o uso da tecnologia como auxiliadora em alguns contextos, tanto no momento de classificação das ações judiciais, indicando similaridade das causas, por exemplo, mas também em decisões de casos complexos, algo que apresenta maiores entraves como será explorado posteriormente. Ainda que não seja uma prática comum ao redor do

³ A Lei Geral de Proteção de Dados define "dados pessoais sensíveis", em seu artigo 5º, inciso II, como "dado pessoal sobre origem racial ou étnica, convicção religiosa, opinião política, filiação a sindicato ou a organização de caráter religioso, filosófico ou político, dado referente à saúde ou à vida sexual, dado genético ou biométrico, quando vinculado a uma pessoa natural".

mundo, alguns países, como a Estônia, já vêm implementando um sistema semelhante (RODAS, 2019).

Novamente, a problemática do viés da decisão por máquinas permanece. Ao mesmo tempo em que há a possibilidade de a Inteligência Artificial elaborar uma decisão que possa ser considerada válida, também não se pode negar uma possível formulação de decisão parcial, por exemplo. Dessa forma, parece ser difícil a elaboração de um sistema que não precise contar com a atuação humana em conjunto. A tomada de decisão, afinal, ainda estaria extremamente vinculada ao pensamento humano e, possivelmente ainda por um período longo, não há que se falar em substituição da atuação do ser humano pela máquina na tomada de decisão, mas sim de uma execução em conjunto a fim de otimizar esse processo nos mais diversos contextos.

O processo de tomada de decisão baseado em Inteligência Artificial pode ser definido, basicamente, como lógico-matemático. Percebe-se que não conta, portanto, com uma capacidade de interpretação que envolve as particularidades de cada caso no que tange às sensibilidades que o ser humano possui. De certa forma, quando se trata de decisões judiciais, há uma crítica ao envolvimento da Inteligência Artificial nesse processo, uma vez que há o risco de gerar decisões automáticas e “mecânicas”, a partir da estrutura lógico-matemática supramencionada. (ARAÚJO, SIMIONI, 2019)

De maneira geral, mediante a necessária supervisão, a atuação da Inteligência Artificial no contexto de tomada de decisão pode ser benéfica. Há, no entanto, que se pensar nas dificuldades de tornar esse processo auditável, considerando que a capacidade de propagação de informações é intensa e o desenvolvimento e funcionamento por trás das máquinas se torna cada vez mais complexo. Como exemplo disso, por vezes, é um obstáculo para os humanos entenderem especificamente o que a máquina quis colocar com determinado resultado, considerando que podem ser utilizados processamentos lógicos que são incompreendidos pelo ser humano. (GUTIERREZ, 2019).

Em suma, resta indubitável: a Inteligência Artificial certamente ocupa e ocupará cada vez mais espaço nos processos de tomada de decisão. O ponto de atenção principal reside no monitoramento humano de decisões enviesadas e parciais que, por vezes, podem resultar em uma perpetuação de discriminações e resultados prejudiciais. Não se quer dizer com isso que a tomada de decisão baseada no

pensamento do ser humano seja privada de inclinações e parcialidade. Muito pelo contrário, sabe-se que essas são características inerentes ao humano e que também influenciarão quando o processo de tomada de decisão contar exclusivamente com a sua atuação. A inserção da Inteligência Artificial nesse contexto chega, portanto, para auxiliar esse processo, considerando que a tecnologia possui uma capacidade de processamento de dados que transpassa a capacidade humana.

Referência bibliográficas

ARAÚJO, E. da S. e; SIMIONI, R. L. Decisão jurídica e inteligência artificial: um retorno ao positivismo. **Revista de Direito**, Viçosa, v. 12, n. 2, p. 1-20, jun. 2019.

BRASIL. **Lei nº 13.709, de 14 de agosto de 2018**. Dispõe sobre a proteção de dados pessoais e altera a Lei nº 12.965, de 23 de abril de 2014 (Marco Civil da Internet). Brasília, DF: Presidência da República; 2018. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709.htm. Acesso em 20. fev. 2022.

DONEDA, D.; ALMEIDA, V. A. F. O que é a governança de algoritmos? **Politics**. Uma publicação do Instituto Nupef. Outubro de 2016. Disponível em: <https://www.politics.org.br/edicoes/o-que-%C3%A9-governan%C3%A7a-de-algoritmos>. Acesso 11 dez. 2021.

GUTIERREZ, A. É possível confiar em um sistema de inteligência artificial?: práticas em torno da melhoria da sua confiança, segurança e evidências de accountability. In: FRAZÃO, A.; MULHOLLAND, C. (coord.). **Inteligência artificial e direito: ética, regulação e responsabilidade**. São Paulo: Revista dos Tribunais, 2019. p. 83-97.

MONTICELLI, J. M.; DOS SANTOS, J. G. **Utilização e Aceitabilidade da Inteligência Artificial na Tomada de Decisão Médica**, p. 1-18, 2021. Disponível em: <https://www2.faccat.br/portal/sites/default/files/Utiliza%C3%A7%C3%A3o%20e%20Aceitabilidade%20da%20Intelig%C3%Aancia%20Artificial%20na%20Tomada%20de%20Decis%C3%A3o%20M%C3%A9dica.pdf>. Acesso em: 11 dez. 2021.

NOGAROLI, R.; SILVA, R. da G. Inteligência artificial na análise diagnóstica da Covid-19: possíveis repercussões sobre a responsabilidade civil do médico. **Migalhas Patrimoniais**, [s. l.], 30 mar. 2020. Disponível em: <https://www.migalhas.com.br/coluna/migalhas-patrimoniais/322941/inteligencia-artificial-na-analise-diagnostica-da-covid-19--possiveis-repercussoes-sobre-a-responsabilidade-civil-do-medico>. Acesso em: 20 fev. 2022.

PINTO, H. A. A utilização da inteligência artificial no processo de tomada de decisões: por uma necessária accountability. **Ril Brasília**, Brasília, v. 225, n. 57, p. 43-60, nov. 2019. Disponível em: https://www12.senado.leg.br/ril/edicoes/57/225/ril_v57_n225_p43.pdf. Acesso em: 11 dez. 2021.

RODAS, S. Algoritmos e IA são usados para que robôs decidam pequenas causas. **Revista Consultor Jurídico**, 27 de outubro de 2019. Disponível em: <https://www.conjur.com.br/2019-out-27/algoritmos-ia-sao-usados-robos-decidam-pequenas-causas>. Acesso em: 12 dez. 2021.

ROQUE, A. V. Inteligência Artificial na tomada de decisões: três premissas básicas. **Revista Eletrônica de Direito Processual – Redp**, Rio de Janeiro, v. 22, n. 1, p. 58-78, abr. 2021. Quadrimestral.

SILVEIRA, S. A. da; SILVA, T. R. da. Controvérsias sobre danos algorítmicos: discursos corporativos sobre discriminação codificada. **Revista Observatório**, Palmas, v. 6, n. 4, p. 1-17, jul. 2020.

VALENTINI, R. S. **Julgamento por computadores?** As novas possibilidades da juscibernética no século XXI e suas implicações para o futuro do Direito e do trabalho dos juristas. 2017. 152 f. Monografia (Especialização) - Curso de Direito, Universidade Federal de Minas Gerais, Belo Horizonte, 2017.

WELCHEN, V. *et al.* **Inteligência Artificial e a Tomada de Decisão na Saúde: Estudos a Nível de Stricto Sensu no Brasil.** In: XIX Mostra de iniciação científica, pós-graduação, pesquisa e extensão programa de pós-graduação em Administração, 19., 2019, Caxias do Sul, 2019. p. 1-12.

TRANSPARÊNCIA E EXPLICABILIDADE DE ALGORITMOS: DA CLUSTERIZAÇÃO AO INDIVÍDUO

Fabiana Faraco Cebrian¹

RESUMO: A computação ubíqua e pervasiva tem transformado a vida em sociedade. Nesse processo, relevantes aspectos da vida cotidiana foram alterados como a representação do indivíduo em espaços digitais, o uso rotineiro de algoritmos e a produção contínua de dados. Características que antes eram individualizadas, como a figura física do indivíduo e seus dados pessoais, passaram a ser categorizadas, classificadas, agrupadas e correlacionadas por algoritmos. O indivíduo adquire novos atributos que extrapolam ou suprimem os traços de sua identidade física e dados pessoais convencionais. Surge a era da produção de sugestões e decisões automatizadas por meio de algoritmos de Inteligência Artificial que refletem diretamente no tecido social. Novos direitos emergem como o direito fundamental à proteção de dados pessoais (CF, art. 5º, LXXIX) e a Lei Geral de Proteção de Dados (LGPD). Para a proteção do titular de dados, aspectos técnicos e jurídicos passaram a fazer parte dos regulamentos e legislações sobre a proteção de dados pessoais. O princípio da transparência e o dever de explicabilidade das decisões automatizadas e perfis digitais ganham relevância na sociedade que vive em vigilância e exposição. O artigo por meio do método dedutivo e pesquisa bibliográfica busca analisar brevemente as sugestões e decisões automatizadas que orientam as decisões humanas. Estas são definidas a partir da construção de perfis digitais que podem ser formados por diversas bases de dados. Tal fator pode conduzir a uma sociedade de decisões e sugestões incompletas e parcialmente informadas sem a devida transparência e explicabilidade.

PALAVRAS-CHAVE: Transparência; Explicabilidade; Decisões Automatizadas; Perfil Digital; Sociedade.

¹ Advogada. Mestranda em Direito pela Pontifícia Universidade Católica do Paraná - PUCPR. E-mail: fabiana.cebrian@gmail.com.

Introdução

A revolução tecnológica proporcionada pelas Tecnologias da Informação e Comunicação (TICs) têm alterado aspectos essenciais da vida em sociedade. Novos debates surgem sobre a utilização de algoritmos, o aumento exponencial da geração de dados e seus reflexos sociais. A sociedade por meio de sua exposição digital e vigilância, por vezes involuntária, passou a gerar uma quantidade significativa de dados. Tal fator impulsionou o desenvolvimento de algoritmos para as mais diversas finalidades e intensificou o modo de vida *onlife*. No mundo atual torna-se inútil questionar se o indivíduo está online ou offline, pois a convergência do analógico com o digital permitiu a convivência de modo híbrido. Este novo modo de interação social abriu espaços para o desenvolvimento de soluções e serviços que envolvem a produção de decisões automatizadas e sugestões por meio de Inteligência Artificial. Para tanto, passou a ser necessária a definição do perfil digital originado do cruzamento das mais variadas bases de dados. Não existem garantias sobre a origem ou qualidade de tais bases de dados. Tais questões desafiam as noções tradicionais de direito e abrem caminho para o desenvolvimento de legislações tecnológicas. A fim de proteger o indivíduo titular de dados, o direito fundamental à proteção de dados pessoais foi recentemente positivado na Constituição Federal em seu art. 5º, LXXIX. No mesmo escopo, a Lei nº. 13.709/2018, também conhecida como Lei Geral de Proteção de Dados Pessoais (LGPD) regula de maneira ampla o tratamento de dados pessoais. Esta apresenta a transparência como um dos princípios basilares e abarca de forma direta a explicabilidade em algoritmos. Surge a defesa de um direito à explicação clara e transparente sobre o caminho que levou o algoritmo à decisão que afete interesses do titular e do seu perfil pessoal. Neste sentido, o artigo por meio do método dedutivo e pesquisa bibliográfica busca analisar brevemente os reflexos de sugestões e decisões automatizadas, bem como de perfis digitais gerados por meio de dados que podem apresentar incompletudes e erros. Tais construções não fidedignas, por vezes impostas pelas plataformas, geram uma sociedade digital de perfis parciais que tendem a conduzir a uma sociedade de decisões incompletas e parcialmente informadas. Ao mesmo tempo, sugestões e decisões automatizadas atingem o seu objetivo final de orientar as decisões humanas na sociedade sem a devida transparência e explicabilidade.

O desenvolvimento de algoritmos na sociedade de exposição e vigilância

Os algoritmos estão presentes em várias fases do desenvolvimento social. Antes do advento dos computadores, os algoritmos foram utilizados pela sociedade com a função de externalizar sua capacidade cognitiva no formato de instruções. Com o advento dos computadores, os algoritmos passaram para uma nova fase em que auxiliam o homem a ampliar sua capacidade cognitiva em busca de eficiência e redução do tempo na execução de tarefas. Com o passar dos anos, o desenvolvimento de algoritmos sofreu modificações incentivadas pela transposição de barreiras científicas, tecnológicas e econômicas. Deu-se início a uma nova geração de algoritmos baseada em aplicativos computacionais que se renovam a cada dia. Essa geração abrange a popularização de dispositivos móveis, por vezes vestíveis e outros periféricos, como Internet das Coisas que permitiram à computação alcançar o patamar de ubíqua². Os algoritmos, que antes eram restritos a equipamentos fixos e instruções, atingem a mobilidade. Estes passam a caminhar junto dos indivíduos, permeiam os espaços físicos da sociedade e inevitavelmente mediam as relações humanas e de seus perfis digitais conquistando, assim, posição de relevância na sociedade. Observa-se o início de uma sociedade de exposição e vigilância. Cada ato de rotina em notebooks e smartphones, transações com cartão de crédito ou débito, pesquisas na web e mídia social, entrada em prédios dentre outras movimentações, deixam um rastro que pode reconstituir as ações do cotidiano. Cada rastro digital pode ser armazenado, minerado e agrupado e passam a constituir uma nova identidade virtual: o perfil digital que pode ser mais demonstrável e tangível do que o eu analógico (HARCOURT, 2015, p. 01 - 02). No mesmo sentido, Han (2018, p. 122), afirma que “deixamos rastros digitais em todo lugar. Nossa vida digital se forma de modo exato na rede”. O mesmo autor esclarece que a sociedade atual vive um panóptico digital, ou seja, exposição e vigilância digital em que a confiança cede lugar ao controle (HAN, 2018, p. 122). Em relação ao perfil digital, Freitas (2017, p. 27) leciona que “a caracterização de perfil refere-se aos métodos e técnicas computacionais aplicados aos dados pessoais ou não dos usuários. E, em tempos de *Big Data*, dados não faltam para serem processados”. A mesma autora salienta que “estas técnicas auxiliam na representatividade estatística, ou seja, na determinação da qualidade de uma

² Vide WEISER, M. The Computer for the 21st Century. **Scientific American**, v. 265, n. 03, p. 78-89, set. 1991.

amostra constituída de modo a corresponder à população no seio da qual ela é escolhida (FREITAS, 2017, p. 27-28). Portanto, este perfil digital, mediante a combinação de diversas bases de dados, busca reconstituir o eu analógico e representar características de grupos e seus relacionamentos. Sem o desenvolvimento de algoritmos embarcados em dispositivos móveis, provavelmente a vigilância e exposição de perfis digitais seriam temas de difícil alcance.

A inteligência artificial e o processo de sugestões e tomada de decisões automatizadas

No transcurso do desenvolvimento tecnológico, a sociedade passou a interagir com sua inteligência natural, algoritmos não computacionais e algoritmos de Inteligência Artificial (IA). De acordo com Doneda (2018, p. 04), “não há dúvidas quanto à relevância econômica e social das decisões tomadas com base em algoritmos, e seria mesmo difícil se pensar o funcionamento da sociedade contemporânea sem a utilização de algoritmos”. Embora o homem ainda detenha, em grande parte, o seu poder de decisão, a ubiquidade permitiu que os algoritmos iniciassem seu processo de organização da vida coletiva rotineiramente. Neste momento, as decisões e sugestões originárias de algoritmos de IA baseadas em perfis digitais passaram a orientar as decisões humanas. “Neste mundo, os sistemas de computador podem, em grande parte, libertar as pessoas da necessidade de tomar decisões, ou seja, substituir as decisões humanas” (HOFFMANN-RIEM, 2020, p. 47). Ao refletir sobre a interferência humana no desenvolvimento de algoritmos de IA e a influência das decisões automatizadas na sociedade, pode-se delinear um paralelo com a pirâmide do conhecimento. De acordo com a pirâmide do conhecimento (*Data-Information-Knowledge-Wisdom*), para o conhecimento ser alcançado, são necessários dados, que geram informações e seguidamente o conhecimento. A partir do conhecimento, é possível alcançar a sabedoria que se refere à capacidade de tomar decisões com base no conhecimento (ACKOFF, 1989). Logo, com a interpretação dessa pirâmide, pode-se traçar um comparativo com a IA baseada em conhecimento. Nela a sabedoria é deixada para os homens e estes decidem a forma como os dados de entrada e os resultados obtidos podem ser utilizados e aplicados ao mundo real ou virtual. Nesse ponto existe uma via dupla: o homem delega à IA a possibilidade de simular o pensamento humano e a IA devolve o resultado para que o homem defina qual será a melhor decisão a ser tomada. Na IA baseada em aprendizado estatístico, não se delega mais ao

homem a sabedoria, ou seja, o poder de decisão a partir do conhecimento adquirido. Com o uso do conhecimento, a IA toma decisões sobre qual caminho o algoritmo deve percorrer, ou seja, cria parcialmente os próprios procedimentos, pois sempre haverá a fase de ajustes de hiperparâmetros de treinamento – que é realizada pelo homem –, existindo, assim, intervenção humana. Isso remete a um ambiente híbrido, formado pela inteligência natural do programador e a inteligência artificial do algoritmo (CORTIZ, 2021, p. 48-55). Em ambos os casos, decisões de IA baseadas em perfis digitais são dependentes de bases de dados. Tais bases podem abarcar erros e incompletudes que interferem na geração do perfil digital e conseqüentemente na decisão automatizada. Observa-se que mesmo com os avanços da Internet das Coisas e dispositivos vestíveis, ainda não é possível coletar, vigiar ou expor todos os aspectos da vida humana. Outro ponto relevante é que decisões automatizadas diferem de sugestões, a primeira refere-se a algo que deve ser cumprido como, por exemplo, a análise de crédito. Esta ocorre de maneira automatizada e o resultado é imediatamente informado ao indivíduo que obterá ou não o crédito. Já a sugestão se relaciona aos sistemas de recomendações. Estes se apresentam como sistemas que visam auxiliar os indivíduos a lidar com a infinidade de opções disponíveis (ZHANG; LU; JIN, 2021 p. 01). As sugestões não são obrigatoriamente acatadas pelo indivíduo, embora voluntariamente podem ser aceitas. Por exemplo, em aplicativos de trânsito e entrega, estes sugerem o caminho mais rápido a ser percorrido. Em ambos os casos, o indivíduo foi avaliado por meio de seu perfil digital que gerou a sugestão ou decisão automatizada, baseada em grupos que realizam o mesmo trajeto. Ou seja, para alguns grupos o caminho apresentado pode ser diferente de outros. O problema central não está nas sugestões apresentadas, mas na possibilidade de o aplicativo ocultar outras opções existentes e viáveis para determinados grupos e isto influenciar na decisão do indivíduo.

A explicabilidade e a transparência de sugestões e decisões automatizadas

No ambiente de rede, formado por uma variedade de aplicativos, os indivíduos podem exibir e demonstrar comportamentos que não refletem a sua realidade pública no mundo real. O mundo *onlife*³ permite outras possibilidades, como a criação de características que não existem no mundo real. O perfil digital pode se desprender da figura real do indivíduo em múltiplas pessoas. Logo, este perfil pode ser composto por características distintas descritas pelo próprio homem, bem como demais características impostas pelas plataformas e outras oriundas de técnicas de clusterização efetuada por algoritmos de agrupamento. De acordo com Freitas (2017, p. 29), “podem ser aplicados diferentes algoritmos tanto para descobrir padrões quanto para determinar a correlação entre conjuntos de dados, de modo a estabelecer um perfil, visto que tais padrões e correlações permitem identificar ou representar pessoas ou grupos de pessoas”. Portanto, é um conjunto de partes que constroem o perfil digital. Ou seja, perfis digitais parciais podem definir sugestões e decisões baseadas em incompletudes capazes de serem acatadas pelos indivíduos. Salienta-se que de acordo com a LGPD são considerados dados pessoais aqueles utilizados na composição de perfil pessoal. Como distinguir e encontrar o perfil digital que mais se assemelha ao indivíduo para a tomada de decisão? Este aspecto merece destaque, pois será necessário um maior debate sobre a transparência na formação de perfis e decisões. De acordo com Rodotà (2014, p. 34-35), “o próprio corpo, então, apresenta um problema de limites e mostra que é impossível conceber direitos e garantias tomando como referência os espaços do passado, justamente aqueles espaços dessa dinâmica social, cultural e tecnológico que mudaram tão radicalmente”. Para tanto, legislações e regulamentos sobre proteção de dados, abordam aspectos técnicos e jurídicos para a proteção do titular de dados. Neste aspecto, a transparência e explicabilidade não se confundem. A primeira é um princípio jurídico presente na LGPD (Art. 6º, VI) que visa fornecer informações claras e precisas sobre quais e como os dados foram utilizados para determinada finalidade e quem são os agentes de tratamento. Este princípio apresenta como finalidade tutelar direitos fundamentais por meio do tratamento ético dos dados. Já a explicabilidade presente na LGPD (Art.

³ Vide FLORIDI, L. *The Onlife Manifesto: Being Human in a Hyperconnected Era*. [S.l.]: Springer Open, 2015 (edição digital).

20, § 1º), refere-se à descrição do percurso realizado pelo algoritmo para a tomada de decisões ou sugestões, portanto, uma forma de fomentar e reforçar a transparência em algoritmos. Resta claro que, quanto mais complexos são os modelos de IA utilizados para a representação de perfil digital e decisões automatizadas, mais custosa será a transparência e explicabilidade da influência do algoritmo em sua construção. É notório que obter apenas a explicabilidade da decisão automatizada, não implicará diretamente na transparência e explicabilidade do perfil digital. Outro ponto relevante trata-se da forma de apresentação da explicabilidade algorítmica de modo claro e interpretável pelo indivíduo. São necessários estudos sobre formas de explicar, por exemplo, modelos complexos de IA para usuários não especializados. Tais respostas, sob os critérios de clareza e interpretáveis, podem tornar a explicabilidade ineficaz ou até mesmo criar a expectativa de um direito exercido. Por meio da explicabilidade, pode-se eventualmente convencer o titular de dados que o resultado está correto, ao invés de explicar como o algoritmo alcançou a decisão.

Considerações finais

Isso posto, resta evidente a correlação entre homens, computadores, algoritmos, inteligência artificial. As atuais demandas da sociedade tornam o homem incapaz de conviver fora deste novo ecossistema híbrido, de inteligência natural e artificial. Este ambiente fomenta a amplificação da sociedade de exposição e vigilância. O desenvolvimento de todos os envolvidos, inclusive do homem, passa a introduzir novos paradigmas sobre o uso de algoritmos. Estes incluem as decisões automatizadas e sugestões tomadas por inteligência artificial com base na definição do perfil digital. Este perfil, composto por diversas bases de dados podem levar a uma representação não fidedigna do indivíduo. Porém, como é sabidamente inviável vigiar todos os aspectos da vida, restam questionamentos sobre se a sociedade de exposição e vigilância seria uma sociedade baseada em sugestões e decisões incompletas e parcialmente informadas que refletem no tecido social. Neste sentido, a Lei Geral de Proteção de Dados, aborda o princípio da transparência e o dever de explicabilidade, a fim de proteger o titular de dados. Este poderá contestar inclusive características de seu perfil digital que considere pejorativas em relação aos aspectos de sua personalidade, visto que essa caracterização também é originária de uma decisão automatizada. A efetividade jurídica dos

dispositivos da LGPD que versam sobre transparência e explicabilidade são desafios contemporâneos que podem revelar novos aspectos sociais, ainda não descobertos, mas que merecem proteção.

Referências bibliográficas

ACKOFF, R. L. **From Data to Wisdom**. New York: John Wiley & Sons, p. 170-172, 1989. Disponível em: <http://faculty.ung.edu/kmelton/Documents/DataWisdom.pdf>. Acesso em: 06. dez. 2021.

CORTIZ, D. Inteligência Artificial: Conceitos fundamentais. In: **Inteligência Artificial: Sociedade, Economia e Estado**. São Paulo: Thomson Reuters. 2021, p. 45 – 49.

DONEDA, et al. Considerações iniciais sobre inteligência artificial, ética e autonomia pessoal. **Revista Pensar**, Fortaleza, v. 23, n. 4, p. 01-17, out./dez. 2018. Disponível em: <https://periodicos.unifor.br/rpen/article/view/8257/pdf>. Acesso em: 04. dez. 2021.

FLORIDI, L. **The Onlife Manifesto: Being Human in a Hyperconnected Era**. [S.l.]: Springer Open, 2015 (edição digital).

FREITAS, C. O. A. A Obscuridade dos Algoritmos e a LGPD. São Paulo: **O Estado de São Paulo**. ago. 2020. Disponível em: <https://politica.estadao.com.br/blogs/fausto-macedo/a-obscuridade-dos-algoritmos-e-a-lgpd/>. Acesso em: 05. dez. 2021.

FREITAS, C. O. A. Tratamento de dados pessoais e a legislação brasileira frente ao profiling e à discriminação a partir das novas tecnologias. **Rev. de Direito, Governança e Novas Tecnologias**, Maranhão, v. 3, n. 2, p. 18–38, jul/dez. 2017.

HAN, Byung-Chul. **No exame: perspectivas do digital**. Trad. Lucas Machado. Petropolis-RJ: Vozes, 2018, 134 p.

HARCOURT, B. E. The expository society. In: **Exposed: Desire and Disobedience in the Digital Age**. Cambridge: Harvard University Press, 2015. p. 1-28.

HOFFMANN-RIEM, W. **Teoria Geral do Direito Digital**: Transformação Digital: Desafios para o Direito. 1. ed. Rio de Janeiro: Forense, 2020, 240 p.

RODOTÀ, S. **El Derecho a Tener Derechos**. 1. ed. Madri: EditoriaTrotta, 2014. 392 p.

WEISER, M. The Computer for the 21st Century. **Scientific American**, v. 265, n. 03, p. 78-89, set. 1991. Disponível em: <http://web.stanford.edu/class/cs240e/papers/weiser.pdf>. Acesso em: 07. dez. 2021.

ZHANG, Q.; LU, J.; JIN, Y. Artificial intelligence in recommender systems. **Complex & Intelligence Systems**, n. 7, p. 439 – 457, 2021.

TRANSPARÊNCIA DE ALGORITMOS E GIG ECONOMY: DISCUSSÕES ACERCA DA VIABILIDADE DE REALIZAÇÃO DE PERÍCIA ALGORÍTMICA

Gabriela Marcassa Thomaz de Aquino¹

RESUMO: A expansão das plataformas digitais de intermediação de trabalho (gig economy) tem propiciado diversas discussões no âmbito legislativo, judiciário e acadêmico. Uma das mais recentes discussões pauta-se na viabilidade de realização de perícia algorítmica. Considerando esse cenário, o presente ensaio tem o objetivo de contribuir com essas discussões, a partir da abordagem da transparência algorítmica, discutindo sobre a definição e função dos algoritmos, como esse sistema se insere dentro da gig economy, quais as principais discussões acerca da perícia algorítmica no Poder Judiciário e quais as possibilidades para efetivação da transparência nesse ambiente. Para verificação dos principais pontos de discussão acerca da perícia algorítmica no judiciário foi realizado um levantamento, no site do Tribunal Superior do Trabalho (TST), e os resultados foram tabulados em uma planilha do Excel. Dessa forma, considerando as discussões acerca da transparência, as justificativas para a realização de perícia algorítmica nos processos analisados e o modelo de negócios das plataformas digitais de intermediação de trabalho, o ensaio propõe que outras estratégias de auditoria podem ser implementadas, tanto pela academia quanto pelo próprio judiciário, a fim de garantir a efetivação da transparência e a garantia de manutenção de segredo de negócios das empresas.

PALAVRAS-CHAVE: Gig economy; Perícia algorítmica; Transparência algorítmica; Auditoria algorítmica; Plataformas digitais.

¹ Doutoranda e Mestra em Direito do Trabalho pela Faculdade de Direito da Universidade de São Paulo. Pesquisadora do Centro de Ensino e Pesquisa em Inovação (CEPI/FGV-SP). E-mail: gabrielaquino01@hotmail.com.

Os algoritmos podem ser definidos como um conjunto de funções lógicas, inseridas no código de um sistema, que preveem uma sequência finita de ações para a resolução de um problema ou a execução de uma tarefa, ou seja, os algoritmos não são o programa, mas a sequência de ações realizadas para que o problema seja resolvido ou a tarefa seja realizada (GOGONI, 2019).

Os algoritmos agem considerando dados de entrada, *feedback*, efeitos de suas decisões e outros componentes do sistema, por isso também podem ser chamados de sistemas algorítmicos. Tais sistemas podem criar suas operações com base nos dados que recebem e não em regras fixadas (SILVEIRA; SILVA, 2020, pp. 1-2), o que faz com que suas ações sejam complexas.

Além disso, com a expansão dos processos de inteligência artificial, decisões que antes eram tomadas por seres humanos (reflexão humana) agora são feitas automaticamente. O *software* codifica uma infinidade de regras e instruções e as computa em poucos segundos (PASQUALE, 2015, p. 8).

Essa expansão tem trazido diversas discussões, uma delas está relacionada com a falta de transparência desses sistemas algorítmicos e eventuais processos discriminatórios. Por conta dessa discussão, há um movimento que busca a transparência dos códigos e o reconhecimento de que esses sistemas possuem vieses e definições prévias embutidas em seus modelos (DIAKOPOULOS, 2016).

Para Sérgio Rodrigo de Pádua, cabe ao direito buscar regular o desenvolvimento de inteligências artificiais, de modo a preservar a possibilidade de pesquisa e desenvolvimento dessas tecnologias sem, no entanto, deixar de fomentar os debates de explicações acerca das decisões tomadas por sistemas baseados em inteligência artificial, decisões estas que impactam a vida das pessoas. (NISHIYAMA et al., 2020, p. 912).

Carlos Araya Paz enfatiza ainda que, embora haja um número grande de aplicações de inteligência artificial bem sucedidas, em alguns casos as decisões tomadas pelo algoritmo são opacas, de modo que só são compreendidas pelos cientistas ou projetistas do modelo. (PAZ, 2021). Nesses modelos, a transparência é comprometida.

Essa discussão sobre a transparência nos algoritmos alcançou, recentemente, as atividades de *gig economy*, resultando em pedidos de realização de perícia algorítmica.

Gig economy é um conceito ainda em construção e é entendido neste ensaio como o ambiente de negócios em que há intermediação de trabalho humano (geograficamente localizado ou baseado na web) por meio de plataformas digitais, no qual prevalecem contratos flexíveis, de curta duração e o pagamento dos trabalhadores se dá por tarefas realizadas (CEPI, 2021a).

Nesse modelo de negócios, a plataformas estabelece, de forma unilateral, os termos e condições aplicáveis às partes e distribui tarefas, fixa o valor da remuneração, indica o tempo para a realização da tarefa distribuída, estabelece padrões mínimos de qualidade e eventuais sanções etc. (KALIL, 2020, pp. 121-124/CEPI, 2021b, p.3).

Para que a intermediação de trabalho seja possível, as plataformas se utilizam de recursos tecnológicos de geolocalização, gerenciamento de tempo e de execução do trabalho, distribuição/alocação de trabalhadores, pagamento online e perfilização² (KALIL, 2020, p. 92). O emprego de algoritmos pelas plataformas digitais levou a literatura a desenvolver o conceito de gerenciamento algorítmico, que é caracterizado por “práticas de supervisão, governança e controle conduzidas por algoritmos de *software*” sobre trabalhadores remotos (MÖHLMAN; ZALMANSON, 2017, p.4).

Esse gerenciamento algorítmico se caracteriza, segundo os autores, sobretudo, por cinco características: rastreamento constante do comportamento do trabalhador, avaliação constante do desempenho dos trabalhadores, decisões automatizadas, ausência de interação humana e pouca transparência. (CEPI, 2021b, pp. 3-4).

Com o aumento de ações trabalhistas visando o reconhecimento de vínculo empregatício entre as plataformas digitais (principalmente plataformas de transporte e de delivery) e os trabalhadores, um dos questionamentos surgidos foi sobre a

² Desenvolvimento de perfis baseados nos padrões comportamentais dos usuários.

configuração da subordinação³ nessa relação, subordinação essa que poderia ser verificada mediante perícia algorítmica.

Visando identificar quais são os pedidos relacionados a perícia algorítmica das plataformas de intermediação de trabalho e quais os argumentos para a realização ou não realização da perícia, foi feito um levantamento de jurisprudência no site do Tribunal Superior do Trabalho (TST)⁴. Os termos utilizados para a busca foram “perícia algorítmica” e “algoritmo” que resultaram em: 5 acórdãos, 16 decisões monocráticas e 9 decisões da Corregedoria Geral. Após a análise de pertinência, os documentos que se relacionavam com a temática estavam distribuídos da seguinte forma: 1 acórdão⁵, 9 decisões monocráticas⁶ e 5 decisões da Corregedoria Geral⁷.

Todas as decisões analisadas envolviam ou a plataforma Uber ou a plataforma 99, sendo que em todos os processos houve o pedido de reconhecimento de vínculo empregatício e realização de perícia algorítmica para provar a existência de subordinação entre trabalhador e plataforma. Todas as decisões suspenderam a realização da perícia algorítmica, aguardam ainda o julgamento do órgão competente para que se possa definir se haverá ou não o procedimento.

Entre os argumentos favoráveis à realização da perícia estavam a necessidade de comprovação da existência de subordinação, sendo a perícia o único procedimento apto a comprovar tal tese e o interesse social público em conhecer como se dá a tomada de decisão do algoritmo da plataforma digital.

As plataformas, no entanto, alegam que a realização de perícia algorítmica é uma medida desproporcional, que viola regras concorrenciais e afeta os segredos de

³ O requisito da subordinação é analisado em conjunto com os outros requisitos aptos a configurarem o vínculo empregatício: pessoa física, personalidade, não eventualidade e onerosidade.

⁴ Levantamento realizado em 08 de dezembro de 2021.

⁵ CorPar - 1001652-15.2020.5.00.0000.

⁶ ROT - 0103519-41.2020.5.01.0000, CorPar - 1001223-14.2021.5.00.0000, CorPar - 1001147-87.2021.5.00.0000, TutCautAnt - 1000825-67.2021.5.00.0000, CorPar - 1000822-15.2021.5.00.0000, CorPar - 1000215-02.2021.5.00.0000, CorPar - 1002165-80.2020.5.00.0000, CorPar - 1002164-95.2020.5.00.0000 e CorPar - 1001652-15.2020.5.00.0000.

⁷ CorPar - 1001223-14.2021.5.00.0000, CorPar - 1001147-87.2021.5.00.0000, CorPar - 1000822-15.2021.5.00.0000, CorPar - 1000215-02.2021.5.00.0000 e CorPar - 1001652-15.2020.5.00.0000.

negócios das plataformas digitais. Segundo as plataformas, os segredos de negócio seriam violados ainda que a perícia fosse realizada sob sigilo judicial, tendo em vista que o perito teria acesso às informações.

Acerca dessa vantagem competitiva que as empresas buscam proteger (segredo de negócio), Nicholas Diakopoulos (2016) argumenta que a transparência do código-fonte dos algoritmos não precisa ser total, sendo um exagero na maioria dos casos. Para o autor, a divulgação de determinadas informações seria suficiente para garantir a transparência algorítmica.

Além disso, Sandvig, Hamilton, Karahalios e Langbort propuseram um conjunto de cinco abordagens possíveis para a auditoria de sistemas algorítmicos: auditoria de usuário não-invasiva; auditoria colaborativa, *sock-puppet audit*, auditoria de raspagem e auditoria de código (SANDVIG et al., 2014, p.16).

A primeira abordagem - auditoria de usuário não-invasiva - consiste no uso de entrevistas em profundidade e observação não-participante com o objetivo de investigar os modos, as dinâmicas e as percepções dos usuários quanto ao sistema estudado.

A segunda - auditoria colaborativa - relaciona-se com a construção de sistemas colaborativos com o objetivo de avaliar determinados pontos do sistema por meio do uso, relato e codificação distribuída. Já a terceira abordagem - *sock-puppet audit* - envolve a simulação de usuários com variáveis controladas pelo projeto de pesquisa.

Por fim, as duas últimas abordagens relacionam-se com aspectos estritamente técnicos, sendo que a auditoria de raspagem consiste na coleta de dados nos sistemas (acesso por meio de APIs, captura de tela e afins), enquanto a auditoria de código se propõe a analisar a cadeia de decisões, escolhas metodológicas, *datasets*, pacotes e módulos de programação. A abordagem de auditoria de código esbarra nas discussões de segredo de negócio, conforme identificado nos processos analisados no Tribunal Superior do Trabalho (TST).

Percebe-se, portanto, que a discussão acerca da transparência algorítmica não é uma questão meramente teórica, sendo, inclusive, judicializada nos casos das plataformas de intermediação de trabalho.

Considerando o crescimento das discussões acerca da necessidade e viabilidade de realização de perícias algorítmicas em plataformas digitais de intermediação de trabalho para identificação de caracterização da subordinação e, consequentemente, do vínculo empregatício, e levando em conta as estratégias de abordagem para a auditoria de sistemas algorítmicos, tem-se que este instrumento (perícia algorítmica) não é a única estratégia que pode ser adotada para maior compreensão dos sistemas algorítmicos.

Pesquisas acadêmicas que adotem métodos de abordagem como a auditoria de usuário não-invasiva, auditoria colaborativa ou *sock-puppet audit* podem auxiliar na discussão, identificando pontos essenciais para o debate público acerca dos limites da transparência algorítmica. Além disso, essas eventuais pesquisas podem identificar pontos importantes a serem observados em eventual abordagem de auditoria estritamente técnica, propondo quesitos a serem analisados em uma auditoria de raspagem, de modo a garantir a transparência dos algoritmos sem corromper segredos de negócio.

Além disso, para se pensar em um caminho regulatório de questões envolvendo desenvolvimento de inteligências artificiais e transparência é importante considerar a participação das empresas no debate, de modo a possibilitar uma maior segurança e transparência dos algoritmos, sem impactar, no entanto, nas pesquisas que buscam o desenvolvimento tecnológico.

Referências bibliográficas

CEPI. Centro de Ensino e Pesquisa em Inovação da FGV Direito SP. **Gig economy e trabalho em plataformas no Brasil: do conceito às plataformas**. São Paulo: FGV Direito SP, 2021a. Disponível em: <https://bibliotecadigital.fgv.br/dspace/handle/10438/31523>. Acesso em: dez. 2021.

CEPI. Centro de Ensino e Pesquisa em Inovação da FGV Direito SP. **Briefing temático #9: Dados, Algoritmos e Trabalhos em Plataforma**, versão 01. São Paulo: FGV Direito SP, 29 de setembro de 2021b. Disponível em: https://bibliotecadigital.fgv.br/dspace/bitstream/handle/10438/31435/BT9_Briefing_tematico_9_Dados_Algoritmos.pdf?sequence=1&isAllowed=y. Acesso em: mar. 2022.

DIAKOPOULOS, N. **Accountability in Algorithmic Decision Making**. Communications of the ACM, February 2016, Vol.59, pg. 56-62. Disponível em: <https://cacm.acm.org/magazines/2016/2/197421-accountability-in-algorithmic-decision-making/fulltext>. Acesso em: dez. 2021.

GOGONI, R. **O que é algoritmo?**. Tecnoblog. Americana, dezembro 2019. Disponível em: <https://tecnoblog.net/311612/o-que-e-algoritmo/>. Acesso em: dez. 2021.

KALIL, R. B. **A regulação do trabalho via plataformas digitais**. São Paulo: Blucher, 2020. Disponível em: <https://openaccess.blucher.com.br/article-list/9786555500295-476/list#undefined>. Acesso em: dez. 2021.

MÖHLMANN, M.; ZALMANSON, L. Hands on the wheel: navigating algorithmic management and Uber drivers' autonomy. In: **International Conference on Information Systems**. Seoul: December 2017, 10-13. Disponível em: https://www.researchgate.net/publication/319965259_Hands_on_the_wheel_Navigating_algorithmic_management_and_Uber_drivers%27_autonomy. Acesso em: dez. 2021.

NISHIYAMA, A. M. et al. **Interdisciplinaridade e direitos humanos**. Rio de Janeiro: Pembroke Collins, 2020. v. 2; 952 p. Disponível em: <https://www.cae-djus.com/wp-content/uploads/2020/11/interdisciplinaridade-e-direitos-humanos-vol2.pdf#page=911>. Acesso em: mar. 2022.

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

PAZ, C. A. Transparencia algorítmica ¿un problema normativo o tecnológico?. **CUHSO**, vol. 31, Temugo, dezembro, 2021. Disponível em: https://scielo.conicyt.cl/scielo.php?pid=S2452-610X2021005000002&script=sci_arttext. Acesso em: mar. 2022.

SANDVIG, C. et al. Auditing algorithms: Research methods for detecting discrimination on internet platforms. **Data and discrimination**: converting critical concerns into productive inquiry, v. 22, 2014. Disponível em: <http://www-personal.umich.edu/~csandvig/research/Auditing%20Algorithms%20--%20Sandvig%20--%20ICA%202014%20Data%20and%20Discrimination%20Preconference.pdf>. Acesso em: dez. 2021.

SILVEIRA, S. A. DA; SILVA, T. R. DA. Controvérsias sobre danos algorítmicos: discursos corporativos sobre discriminação codificada. **Revista Observatório**, v. 6, n. 4, p. a1pt, 1 jul. 2020. Disponível em: <https://sistemas.uft.edu.br/periodicos/index.php/observatorio/article/view/11069/17889>. Acesso em: dez. 2021.

ESTRUTURA E FUNCIONAMENTO DA INTERNET

ASPECTOS TÉCNICOS, POLÍTICOS E REGULATÓRIOS

PRIVACIDADE E PROTEÇÃO DE DADOS PESSOAIS

OS DADOS SÃO DE QUEM?

Rodrigo Duarte Guedes¹

RESUMO: O fluxo internacional de dados é uma questão que vem se colocando para além da discussão das características da Internet, pois os dados não são o novo petróleo, na medida em que não são recursos naturais, mas dentro da conjuntura atual da sociedade possibilitam a criação de fluxos de poder e dinheiro. O presente artigo destacará como os dados se tornaram centrais e a necessidade dos Estados intervirem para a construção de soluções locais para armazenamento dos seus dados.

PALAVRAS-CHAVE: Dados; Fluxo Internacional de Dados; Regulação da Internet; Internet; Economia Política da Internet.

Rotina de produção de dados

Ao se imaginar uma rotina diária, é possível vislumbrar o quanto estamos compartilhando o dia com diferentes ambientes digitais que intermediam distintas relações por meio de tecnologias: as plataformas. Por exemplo, ao acordar e buscar as informações sobre o que ocorre no mundo, abre-se o Twitter. Para saber das novidades do círculo de amigos, utiliza-se o Facebook e o Instagram. Para diminuir o tédio no caminho para o trabalho ou a fim de dar aquela força nos exercícios matinais, utiliza-se o Spotify. O próprio trabalho, em tempos atuais, pode se dar dentro de plataformas como Uber, Ifood e/ou ser permeado pelas outras citadas.

Assim, sem prestar muita atenção, o dia foi compartilhado com o uso de distintos ambientes digitais que funcionam como espaços que oferecem acesso a outras pessoas e agentes em vários lados, conectando vendedores e compradores,

¹ Graduado em Ciências Sociais pela Universidade Federal do Rio de Janeiro (UFRJ), mestre em Políticas Públicas, Estratégia e Desenvolvimento pelo Instituto de Economia da UFRJ. É doutorando em Ciência da Informação pelo Instituto Brasileiro de Informação em Ciência e Tecnologia em parceria com a Escola de Comunicação da UFRJ.

anunciantes e usuários, produtores audiovisuais e públicos. Sendo necessário destacar que a intermediação realizada pelas plataformas possibilita também a transformação da rotina em dados.

Segundo a pesquisa TIC Domicílios 2019², no Brasil estima-se que 133,8 milhões de pessoas, com dez anos ou mais, estão conectadas à rede. Destas, quase a totalidade (99% da amostra), tem no telefone celular o dispositivo mais utilizado para uso da Internet. Esses dados gerais, quando transpostos para rotina descrita anteriormente, fornecem uma noção da quantidade de dados gerados e apropriados pelas diferentes plataformas utilizadas pelas pessoas (NICBR, 2020).

O presente ensaio investigará a questão dos dados, mais especificamente a discussão sobre o fluxo internacional de dados. Inicialmente vislumbrou-se a quantidade de pessoas que em 2019 estavam conectadas à Internet, assim, atrelada a esta estimativa temos o artigo de Viola (2019, p. 18), onde o autor pontua que atualmente “um mero armazenamento de dados na nuvem – desde que envolva um servidor localizado no exterior – pode ser suficiente para configurar uma transmissão transfronteiriça, a exigir o cumprimento de diversas obrigações”. Assim, ao longo do ensaio defenderemos a necessidade da intervenção dos Estados na construção de soluções locais para o armazenamento dos dados de seus cidadãos.

Onde os dados da nuvem ficam armazenados?

A quantidade de dados possíveis de serem gerados pelo montante de 133,8 milhões de pessoas conectadas e a utilização de serviços de plataformas digitais que podem colocar seus servidores em países do exterior compõem o cenário destacado na introdução e que será aprofundado neste tópico para a discussão sobre a localização dos dados e o conseqüente fluxo internacional dos mesmos.

Inicialmente é importante destacar o que são esses dados, a Lei Geral de Proteção de Dados (LGPD) destaca que dados pessoais são as informações relacionadas à pessoa natural identificada ou identificável. Assim, os dados

² TIC – Tecnologias da Informação e Comunicação.

identificados são, por exemplo, o CPF que possibilita através de seus números identificar uma pessoa específica. Os dados identificáveis são aqueles que utilizados de forma isolada não possibilitam a identificação de uma pessoa, porém, sua utilização em conjunto com outros dados pode levar a identificação. Um exemplo desta situação são os dados de CNPJ, que apesar de serem dados sobre uma esfera de pessoa jurídica podem levar até o CPF e assim aos dados pessoais. Por outro lado, temos os dados não pessoais que são aqueles possíveis de serem anonimizados.

Outro ponto interessante que emerge da utilização da Internet é que as conexões feitas para utilização da mesma implicarão trânsito de dados. Isto ocorre, pois a lógica da rede das redes é ser interoperável e de possibilitar o fluxo de informações. Tendo em vista tal característica a questão sobre a localização dos dados, em inglês data localization, se coloca em discussão, pois como destacado por Brandão (2020), os dados para além da classificação anterior também têm uma dimensão física. Logo, quando se aborda a localização dos dados o que está em debate é em qual território os dados estarão retidos.

Entrando um pouco mais no debate sobre a territorialidade Brandão (2020), salienta que existem diversos critérios para a localização dos dados, em geral, a autora aponta que as escolhas se centram em aspectos lógicos, como a quantidade de tráfego e o clima. Contudo, existem casos em que ocorre uma localização mandatória, isto é, determinado país através das suas previsões legais e regulatórias determinam que os dados fiquem armazenados em seu território.

Ao abordarmos os dados e sua localização é nevrálgico que não nos desconectemos do cenário que vivemos, assim, conforme Silveira (2021, p. 35) destaca estamos em uma conjuntura de um capitalismo cada vez mais informacional. Os dados e sua organização não escapam deste contexto, porém algumas questões passam a ficar encobertas e ofuscadas, tornando-se assim em não-questões, na medida em que não figuram mais nas discussões. Silveira (2021, p. 35) aponta quatro pontos, contudo para o presente ensaio, destacarei em específico o segundo. Neste o autor salienta que não se tem um questionamento maior sobre as “consequências negativas tanto locais quanto nacionais na utilização das estruturas tecnológicas das plataformas, uma vez que elas respeitariam os contratos”.

A falta deste tipo de questionamento nos leva a situações como a exposta por Lucena (2021) que destaca a situação onde empresas nacionais menores nem chegam a pensar sobre os aspectos que envolvem a transferência internacional de dados, pois pelo seu porte ficam à mercê de empresas dominantes do segmento como Amazon e Google, logo seus dados são transferidos internacionalmente sem que as empresas tenham prestado atenção para tal fato.

Outro ponto importante é a observação de Lucena (2021), destacando que as grandes empresas são muito afeitas à globalização quando se trata de expandir o modelo de negócios e conteúdo, porém em contrapartida defendem com maior vigor sua operação segundo uma jurisdição específica. Este fato fica bem visível com a defesa das plataformas sediadas nos Estados Unidos da América (EUA) da sua jurisdição quando se trata de colaboração com entidades policiais.

Cabe lembrar que após o final da Segunda Guerra Mundial, dentro do plano diplomático, os EUA, através do seu Departamento de Estado, se dedicam a legitimar a doutrina de livre fluxo da informação. Para além da proximidade desta doutrina com a de livre troca, também defendida pelos EUA, há uma relação com os insumos necessários para as máquinas de comunicação que entre outras coisas demandam uma livre circulação da informação. Neste contexto o acesso livre a informação é defendido justamente sob os termos exclusivos de venda e compra, logo a busca não é atingir o potencial de libertação da informação, mas sim a construção de oligopólios.

Essa recapitulação histórica e o exemplo das empresas nacionais menores são importantes na medida em que desvelam a existência de mais do que aspectos lógicos na definição dos critérios de localização dos dados. Como veremos a seguir com a plataformização e a dataficação, existem aspectos econômicos que influenciam diretamente nesta territorialidade dos dados.

A utilização dos dados em negócios é anterior as plataformas digitais, porém com estas seu uso ganha escala de tal forma que o conceito de plataformização surge e segundo Helmond (2015, p. 1) representa “a extensão das plataformas de mídia social para o resto da web e seu esforço para tornar os dados da web externos ‘prontos para a plataforma’”. Assim, para compreendermos esta mudança é interessante destacar que desde 1991 a web se tornou o principal meio para acesso e utilização da Internet, contudo desde 2010 as plataformas digitais assumem este lugar. Em suma, por meio da plataformização se vislumbra como

ocorreu a mudança dos sites de redes sociais para plataformas de redes sociais, a transformação está centrada na ascensão da web 2.0³ e na utilização de API's⁴ que possibilitam a comunicação entre os serviços das plataformas e desenvolvedores terceiros.

Essa modificação com a concentração nas plataformas como meio para utilização da web desemboca na dataficação definida por José Van Dijck (2014, p. 198), se apoiando em Mayer-Schoenberger e Cukier (2013), como “a transformação da ação social em dados quantificados online, permitindo o rastreamento em tempo real e a análise preditiva”⁵. Poell, Nieborg, e Van Dijck (2020, p. 5) reforçam que a noção de dataficação se refere “às maneiras pelas quais as plataformas digitais se transformam em dados, práticas e processos que historicamente escaparam à quantificação”. Porém, ampliam o escopo ao salientarem que os dados capturados pelas plataformas não são apenas aqueles relacionados a “dados demográficos ou de dados de perfil oferecidos pelos usuários ou solicitados por meio de pesquisa online, mas principalmente a metadados comportamentais”.

Vislumbramos que os dados podem ser classificados em pessoais, estes sendo identificados e identificáveis, e em não pessoais, assim como não ficam voando em nuvens, mas possuem uma dimensão física, logo demandam seu armazenamento em alguma localidade. Abre-se assim a indagação sobre onde os mesmos deveriam ser armazenados, se tal escolha fica a cargo das empresas e da sua lógica capitalista, ou se caberia aos Estados a regulamentação do armazenamento em seus territórios. Frisamos que atualmente as empresas definem onde os dados coletados por elas serão armazenados.

³ “Tim O’Reilly cunhou o termo Web 2.0 que popularizou-se rapidamente a partir da publicação do artigo intitulado ‘What is Web 2.0: Design Patterns and Business Models for the Next Generation of Software’ (O’Reilly, 2005)”. (BOMFIM; SAMPAIO, 2008, p. 1)

⁴ Application Programming Interfaces (APIs), em português pode ser traduzida como Interface de Programação de Aplicativos

⁵ Texto original: “[...] the transformation of social action into online quantified data, thus allowing for real-time tracking and predictive analysis”. (VAN DICJK, 2014, p. 198)

A escolha de onde armazenar os dados

Apesar da resolução desta questão não ser simples, pois envolve diversos aspectos e pode, dependendo do ponto de análise, impactar na lógica da Internet de ser interoperável e permitir o fluxo de informações, defendemos que cabe uma maior regulação do Estado. Tal posição se assenta na leitura de que neste novo cenário composto pelo crescimento no volume de dados e no desenvolvimento de novas tecnologias que possibilitam a contínua “geração” de dados há um rearranjo na lógica de negócios e as empresas, ao perceberem o potencial da exploração dos dados para extração de valor e controle passam a se reestruturar buscando acumular e reter o máximo de dados possível.

Assim, a partir do exposto, pontua-se que o caminho para resolução da questão sobre o fluxo internacional de dados é que os países, em especial do Sul, centrem sua atenção na utilização dos dados pelo prisma econômico, ao oposto do que se verifica na atualidade onde países como o Brasil, segundo Silveira (2021, p. 38), compram “os melhores produtos e serviços pelo menor preço. O uso é subentendido como o passaporte para o avanço econômico”. É central a criação de soluções próprias que abranjam desde o armazenamento até o desenvolvimento de aprendizagem de máquinas utilizando os dados produzidos no seu país.

Sadowski (2019, p. 2), em seu artigo, reforça essa compreensão de que os dados não são um recurso natural e são transformados em capital. Dentro desta lógica, os dados assumem tamanha centralidade que permeiam o funcionamento de casas inteligentes até cidades inteligentes, das finanças até a governança, da produção até a distribuição e de dispositivos de consumo até os sistemas empresariais. Esses são alguns exemplos trazidos pelo autor para retratar como “os dados são uma forma fundamental de capital para tudo”⁶. Logo, “Sem dados, muitas destas tecnologias e organizações não seriam capazes de funcionar, e muito menos de gerar valor”⁷.

Vimos que os dados se tornaram em insumos centrais dentro da atual fase do

⁶ Texto original: “data is a foundational form of capital for everything”. (SADOWSKI, 2019, p. 2)

⁷ Texto original: “Without data, many of these technologies and organisations would not be able to operate, let alone be able to generate value”. (SADOWSKI, 2019, p. 2)

capitalismo e a disputa pela sua extração e armazenamento se configurou em embates entre empresas e Estados. Até o momento com exceção de poucos países a definição sobre o local de armazenamento dos dados se faz de acordo com o interesse das empresas.

Portanto, partindo-se da premissa que os dados são insumos transformados em fluxos de poder e dinheiro defende-se que a resolução da questão sobre o fluxo internacional de dados perpassa uma compreensão de que à Internet no atual cenário privilegia determinadas empresas e localidades. Logo, a criação de soluções próprias que possibilitem tanto o armazenamento quanto a utilização dos dados dentro do território dos Estados se avizinha como uma possibilidade para busca de um novo desenvolvimento e avanço das forças produtivas.

Referências bibliográficas

BOMFIM, M. N. C.; SAMPAIO, F. F. **A Web 2.0**, suas tecnologias e aplicações educacionais. [S. l.: s. n.], 2008.

[Entrevista] **Data Localization**: entre o nacional e o global, com Luiza Brandão (IRIS). Centro de Ensino e Pesquisa em Inovação (CEPI FGV Direito SP). 07 out. 2020. Disponível em: PARTE 1 <https://youtu.be/no0c4HiyZ-c> (19min31s) PARTE 2 <https://youtu.be/6EcXC1nOXoA> (13min45s)

[Entrevista] **Fragmentação da Internet e localização de dados**: impasses e externalidades, com Cláudio Lucena (UEPB / FCT / INPD). Centro de Ensino e Pesquisa em Inovação (CEPI FGV Direito SP). 13 set. 2021(37min48s). Disponível em: <https://youtu.be/c0qBfr-cqBY>.

HELMOND, A. The platformization of the web: Making web data platform ready. **Social Media+ Society**, p. 1–11, 2015.

MAYER-SCHOENBERGER, V.; CUKIER, K.. **Big Data**: A Revolution that will transform how we live, work, and think. London: John Murray Publishers. 2013

NIC.BR. Núcleo de Informação e Coordenação do Ponto BR. **Pesquisa sobre o Uso das Tecnologias de Informação e Comunicação nos Domicílios Brasileiros** – TIC Domicílios 2019. São Paulo: Conselho Gestor da Internet no

Brasil, 2020. Disponível em: <https://cetic.br/pt/tics/domicilios/2019/individuos/>. Acesso em: 17/08/2020.

POELL, T.; NIEBORG, D.; VAN DIJCK, J. Plataformização. **Revista Fronteiras – estudos midiáticos**, vol. 22, no. 1, 2020.

SADOWSKI, J. When data is capital: Datafication, accumulation, and extraction. **Big Data & Society**, vol. 6, no. 1. 2019.

SILVEIRA, S. A; SOUZA, J. (Org.); Cassino (Org.). **Colonialismo de dados**: como opera a trincheira algorítmica na guerra neoliberal. 1. ed. São Paulo: Autonomia Literária, 2021. v. 1. 212p.

VAN DIJCK, J. Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. **Surveillance & Society**, vol. 12, no. 2, p. 197–208, 2014.

VIOLA, M. Transferência de dados entre Europa e Brasil: Análise da Adequação da Legislação Brasileira. **Instituto Tecnologia e Sociedade**, [s.l.], 2019. Disponível em: https://itsrio.org/wp-content/uploads/2019/12/Relatorio_UK_Azul_INTERACTIVE_Justificado.pdf.

ESTRUTURA E FUNCIONAMENTO DA INTERNET

ASPECTOS TÉCNICOS, POLÍTICOS E REGULATÓRIOS

SOBERANIA DIGITAL

FLUXOS TRANSNACIONAIS E INTERNET ABERTA NO SÉCULO XX

Jaqueline Trevisan Pigatto¹

RESUMO: Este artigo busca analisar brevemente algumas das demandas e das respostas normativas da governança da Internet, encontradas entre as décadas de 2010 e 2020 para lidar com problemas transnacionais de fluxo de dados e a manutenção de uma Internet aberta. Com um olhar internacional e institucional, vê-se a necessidade de mecanismos de coordenação e cooperação entre os atores, preservando o espírito multissetorial que está embutido nessa governança desde o princípio. Ainda que haja uma presença mais forte nessas décadas da figura estatal, práticas de correção mostram que são cada vez mais necessárias respostas em conjunto, a fim de preservar os benefícios da rede mundial de computadores e as adequações demandadas pelos atores que a compõem.

PALAVRAS-CHAVE: Governança da Internet; Multissetorialismo; Soberania; Dados pessoais; Instituições internacionais

Introdução

Internet aberta x Estados soberanos

Compreender a Governança da Internet é compreender a Internet em si, seu funcionamento e o determinismo de sua construção que levou à criação e manutenção de práticas, leis e atores que afetam as sociedades e a dependência digital criada nelas. As características centrais da Internet que a acompanham

¹ Doutoranda em Ciências Sociais pela Faculdade de Ciências e Letras (UNESP-Araraquara); mestre em Relações Internacionais pelo Programa de Pós-graduação San Tiago Dantas (UNESP - UNICAMP - PUC/SP); e pesquisadora e líder de projeto na Associação Data Privacy Brasil de Pesquisa. Contato: jaqueline.t.pigatto@unesp.br

desde sua criação na década de 1970, até sua definitiva abertura para qualquer usuário na década de 1990, como descentralização e composição multissetorial (ou *multistakeholder*), são categorias também aplicáveis à sua governança: uma pluralidade de atores (governos, setor privado, acadêmicos e sociedade civil) descentralizados que compõem diversos arranjos, regimes e práticas transnacionais, ao mesmo tempo em que evoluem em seus benefícios e consequências multinível (desde o local, passando pelos níveis nacional, regional e global).

Ainda que muitos Estados entrem em conflito com esse caráter aberto e universal da Internet por conflitar com leis e princípios de soberania territorial, como coloca Doneda (2020): “A existência de uma internet aberta e em escala global é o elemento fundamental que torna possível a existência dos aplicativos e ferramentas que são recursos ainda mais importantes em uma situação de isolamento social como a atual”, referindo-se ao contexto recente da pandemia de COVID-19.

O grande desafio imposto pelas atuais configurações da governança da Internet, que parece impor a todo momento um conflito entre preferências dos atores estatais e não-estatais, e um fluxo transnacional de dados, criou uma corrida acirrada que envolve regulações, inovação e componentes socioeconômicos e até mesmo políticos. Por um lado, o setor privado sai na frente com a inovação, os conhecimentos e agilidades técnicas para lidar com problemas do dia a dia, em dispositivos que estão nas mãos das pessoas a todo momento realizando trocas de dados e informações de preferências pessoais, transações financeiras, etc. Por outro, Estados tendem a ver esse fluxo concentrado em grandes corporações estrangeiras como uma ameaça e, por vezes, como violação de sua soberania estatal - preocupação que aumentou após interferências em campanhas políticas e que também trazem cada vez mais desigualdades de trabalho e acesso, por conta de algoritmos e processos pouco transparentes de muitos serviços digitais.

O caráter multissetorial se mostra, assim, cada vez mais necessário junto a mecanismos de cooperação e correção - ou seja, o Estado não é mais o único agente regulador de modo unilateral. Respostas em conjunto são cada vez mais demandadas para que se possa aproximar um consenso não apenas técnico, mas político, em torno de diversos temas da governança da Internet, em especial o fluxo transnacional de dados e seus valores econômico e social.

Constitucionalismo digital

As respostas normativas da década de 2010

O Constitucionalismo Digital é uma corrente recente de estudos que advogam por uma nova fase da governança da Internet, na qual não se considera mais o ciberlibertarianismo de John Perry Barlow (BARLOW, 1996), nem mesmo é defendido um caráter autoritário do Estado que poderia fragmentar a Internet. Trata-se de uma ideologia, como visto por Celeste (2019), que enfatiza a necessidade de se adaptar valores constitucionais para o ambiente digital, que ao mesmo tempo que possibilita ao indivíduo o exercício de seus direitos fundamentais, também aumenta os riscos de ameaça aos mesmos.

Nesse sentido, se faz necessário que noções de não-interferência - princípio básico da soberania nas relações internacionais² - seja revisto. Isto em certa medida já vem sendo respondido com legislações nacionais e regionais que apresentam algum grau de extraterritorialidade. O exemplo mais comum aqui é da Regulação Geral de Proteção de Dados (GDPR) da União Europeia, que se impõe sobre plataformas estrangeiras que tratam dados pessoais de cidadãos do bloco.

Outro exemplo até mesmo anterior e que se destaca pelo caráter multissetorial é do Marco Civil da Internet (Lei nº 12.965/2014), a “*Internet Bill of Rights*” brasileira, aprovada em 2014 e baseada nos pilares da proteção de dados, neutralidade da rede e responsabilidade de intermediários. Foi a primeira legislação desse tipo no mundo e marcou um curto, porém bem sucedido período de liderança do Brasil na governança global da Internet. Foi também em 2014 que o país sediou o encontro NETMundial, por iniciativas do governo brasileiro e da Corporação da Internet para Atribuição de Nomes e Números (ICANN), entidade global e multissetorial que gerencia os nomes de domínio e protocolos de websites da rede. Entretanto, a instabilidade interna que se seguiu fez com que essa pauta fosse gradativamente perdendo importância no país, papel que ao fim da década

² Aqui referente a soberania nacional, o conceito implica um exercício de autoridade estatal exclusiva em seu território, pautado no princípio de que nenhum Estado interfira em assuntos internos de outro, como foi estabelecido pela Paz de Westfália em 1648.

parece ter se transferido para a União Europeia - especialmente França e Alemanha.

Um exemplo é a nuvem Gaia-X³, anunciada conjuntamente pelos dois países e que visa ser uma alternativa aberta e segura aos provedores de nuvem estadunidenses (de empresas como Google e Amazon), seguindo os valores europeus de proteção de dados. Tal iniciativa se encaixa na estratégia de autonomia digital e de soberania de dados europeia, dentre outras propostas como o *Digital Services Act* (com foco em conteúdo das plataformas), o *Digital Markets Act* (focado em ações antitruste) e a própria GDPR (proteção de dados).

Com diferentes legislações e práticas surgindo, a década de 2020 parece trazer um olhar atento para mecanismos de cooperação e coordenação, como posto pelo relatório do Secretário Geral das Nações Unidas “*The Age of Digital Interdependence*” (DIGITAL COOPERATION, 2019). Além disso, ao pensarmos nos múltiplos níveis de interação entre indivíduos e serviços transnacionais, não basta pensar em uma governança global, mas sim uma governança global e multinível. Os indivíduos e as sociedades em geral interagem não diretamente com o Estado, mas sim com ramos específicos do governo. Assim, é de se pensar menos na replicação de instituições domésticas no nível global, e mais em conectar as instituições já existentes com as redes globais, pois essas instituições governamentais exercem poder coercitivo e possuem legitimidade pública (SLAUGHTER, 2003).

Além disso, falar em cooperação pode significar uma troca de experiências, o que valoriza a continuidade e intensificação de *policy spaces* como o Fórum de Governança da Internet (IGF), tanto o global quanto os nacionais e regionais. Outro fenômeno observado é a influência que certas medidas possuem no cenário internacional, como foi o caso da própria GDPR. A China - conhecida por seu caráter mais autoritário (censura de determinados conteúdos e rígido controle doméstico) - aprovou em 2021 uma lei de proteção de dados que se assemelha à GDPR, avançando também em outras legislações ainda em debate no Ocidente.

Um outro espaço de debate sobre inovações tecnológicas e ações de governança

³ Para mais informações, ver: <https://www.gtai.de/gtai-en/invest/industries/information-technologies/gaia-x-germany-and-france-create-european-data-ecoysystem-528520>.

nasce na China durante a década de 2010, a Conferência Mundial da Internet (*World Internet Conference*), que já procura estabelecer princípios e olha principalmente para a cibersegurança, em uma visão diferente da ocidental, mas também voltada à construção de uma comunidade global para a governança da rede. Por esse evento fica claro a estratégia chinesa de soberania digital, onde a cooperação internacional não impede que cada Estado exerça seu próprio controle doméstico. Assim, é bastante comum ver representantes de países vizinhos à China no evento, enquanto a participação de estadunidenses e europeus é mínima.

Coordenação e *accountability*

As possíveis respostas desenhadas no início da década de 2020

A partir de casos de conjuntura como as revelações de Edward Snowden em 2013, a eleição de Donald Trump em 2016 e a votação pelo *Brexit* na Europa, junto às revelações do uso indevido de dados do Facebook pela consultoria Cambridge Analytica, a necessidade de regulação ficou clara tanto para os Estados quanto para o setor privado. O discurso do presidente francês Emmanuel Macron no IGF de 2018 é icônico neste sentido, de convidar o setor privado e a sociedade civil para regular junto ao Estado, ou o Estado regularia por conta própria. Desde então, observa-se um esforço em várias frentes por parte do setor privado - como elaborar mecanismos de prestação de contas e maiores controles de privacidade aos usuários. Destaca-se a empreitada do *Oversight Board* - um comitê independente criado pelo Facebook que emite decisões globais sobre as políticas da rede social, institucionalizando assim uma espécie de autorregulação privada global e multissetorial, mas cujo *enforcement* frente ao Estado ainda é questionável.

Alguns Estados têm tomado medidas protecionistas que são muito debatidas pela possibilidade de infringir o ideal da Internet aberta. As políticas de localização de dados são o maior exemplo, onde as empresas devem armazenar os dados de seus usuários em seus países de origem, como acontece na China⁴. Até agora,

⁴ NICAS, J. et al. *Censorship, Surveillance and Profits: A Hard Bargain for Apple in China*. **The**

não existem evidências de que onde se é aplicada essa política houve aumento de segurança digital e privacidade. Outro ponto a ser pensado no futuro próximo é o maior uso de novas tecnologias e possibilidades com o 5G, que permitirá uma coleta maior ainda de dados por usos de soluções baseadas na Internet das Coisas e na Inteligência Artificial, enfatizando a necessidade de conversas regulatórias e bases principiológicas. As principais perguntas permanecem: quem está coletando, processando e armazenando os dados, com qual propósito e para quem? Reforça-se, desse modo, a necessidade de se investir em políticas de transparência e coordenação entre os atores.

No que podemos considerar um processo de *forum shifting*⁵, a governança da Internet está cada vez mais com seus assuntos presentes em espaços tradicionais de políticas como a Organização Mundial do Comércio (OMC), o G7 e a Organização para a Cooperação e Desenvolvimento Econômico (OCDE), do que apenas restrita aos espaços específicos de sua comunidade epistêmica, como o próprio IGF. Esse fenômeno reforça as intenções dos atores de manter uma Internet aberta, que permita os benefícios dos fluxos transnacionais de dados e trocas de conteúdo, ao mesmo tempo que preserva e possibilita o exercício de direitos fundamentais pelos seus usuários.

Referências bibliográficas

BARLOW, J. A Declaration of the Independence of Cyberspace. **EFF**, 1996. Disponível em: <http://bit.ly/2OJU00Z>. Acesso em: 24 mar. 2018.

CELESTE, E. Digital constitutionalism: a new systematic theorization. **International Review of Law, Computers & Technology**, v. 33, n. 1, p. 76–99,

New York Times, May 17, 2021. Disponível em: <https://www.nytimes.com/2021/05/17/technology/apple-china-censorship-data.html>. Acesso em 29 nov. 2021.

⁵ O termo *forum shifting* foi utilizado por Marília Maciel (DiploFoundation) para se referir aos temas que “vazam” para outros espaços não focados exclusivamente em governança da Internet. Ver: [Sem. Internacional] **Os desafios da governança global da Internet** (Áudio em Português). 30 mar. 2021. Disponível em: https://www.youtube.com/watch?v=x-3sERWir_o. Acesso em: 7 abr. 2021.

2019.

DIGITAL COOPERATION. **The Age of Digital Interdependence**. Report of the UN Secretary-General's High-level Panel on Digital Cooperation. Disponível em: <http://bit.ly/2TbxVqU>. Acesso em: 17 jun. 2019.

DONEDA, D. Criptografia, segurança e confiança em tempos de pandemia. **Jota**, 06/05/2020. Disponível em: <https://www.jota.info/opiniao-e-analise/artigos/criptografia-seguranca-e-confianca-em-tempos-de-pandemia-06052020>. Acesso em: 13 abr. 2021.

SLAUGHTER, A. Everyday Global Governance. **Daedalus**, v.132, n.1, p.83–90, 2003.

